

# **BUILDING EXTRACTION AND ROOF TYPE CLASSIFICATION OF AERIAL IMAGES FOR MAXIMAL PV PANEL INSTALLATION**

**A PROJECT REPORT**

*Submitted by*

**SHRUTHI M                      2018103592**

**GAYATHRI M                      2018103535**

**JAYAPRIYA M                      2018103029**

*in partial fulfillment for the award of the degree*

*of*

**BACHELOR OF ENGINEERING**

**in**

**COMPUTER SCIENCE AND ENGINEERING**



**COLLEGE OF ENGINEERING, GUINDY**

**ANNA UNIVERSITY:: CHENNAI 600 025**

**JUNE 2022**

# **ANNA UNIVERSITY : CHENNAI 600 025**

## **BONAFIDE CERTIFICATE**

Certified that this project report “**BUILDING EXTRACTION AND ROOF TYPE CLASSIFICATION OF AERIAL IMAGES FOR MAXIMAL PV PANEL INSTALLATION**” is the bonafide work of “**SHRUTHI M, GAYATHRI M, and JAYAPRIYA M**” who carried out the project work under my supervision.

### **SIGNATURE**

**DR.S. VALLI**  
**HEAD OF THE DEPARTMENT**

Department of Computer  
Science and Engineering

College of Engineering  
Guindy, Anna University  
Chennai - 600025

### **SIGNATURE**

**DR.P. UMA MAHESWARI**  
**SUPERVISOR**

Associate Professor  
Department of Computer  
Science and Engineering

College of Engineering  
Guindy, Anna University  
Chennai - 600025

## ACKNOWLEDGEMENT

Foremost, we would like to express our sincere gratitude to the below mentioned people who have been instrumental in the completion of this project.

We extend our utmost gratitude to our guide **Dr. P. Uma Maheswari**, Associate Professor for guiding us with patience throughout the project. We thank her for leading us in the right direction through insightful discussions and ideas and encouragement at every stage of the project. We value and appreciate her knowledge, principles and ability to bring out the best in us. We could never have accomplished so much without this unwavering support and guidance.

We are deeply thankful to our beloved Head of the Department, **Dr. S. Valli**, Department of Computer Science and Engineering, for providing us with the facilities of the department and for molding us both technically and morally for achieving greater success in life.

We are grateful to the panel of reviewers **Dr. S. Chitrakala, Professor, Dr. V. Vetrisevi, Professor and Dr. G. S. Mahalakshmi, Associate Professor** for their insightful suggestions and critical reviews throughout the course of our project.

We also thank our parents, family and friends for bearing with us throughout the course of our project and for providing the opportunity for undergoing this course in such a prestigious institution.

**SHRUTHI M**

**GAYATHRI M**

**JAYAPRIYA M**

## ABSTRACT

Environmental issues are a major source of concern these days, as they threaten the existence of humankind. The transition to renewable energy sources such as wind and solar has begun as mankind faces the issue of meeting its energy demands in a more sustainable manner. Several studies are being conducted on optimizing the use of solar energy through the use of highly efficient solar panels. However, this is associated with a lot of concerns on its own. Traditional approaches, such as online assessment, however pose significant challenges as they are time-consuming and costly. It also demands a large amount of human labor, with competent individuals visiting the site to assess each building and determine the type of roof to determine the PV panel installation arrangement.

In accordance, this project focuses on creating a comprehensive and detailed pipeline to address the problem of classifying and manually assessing rooftops for solar panel placement. For this purpose, understanding urban dynamics, and accurate mapping of buildings from satellite images is crucial. Thus, given a satellite image with wide coverage and spatial resolution of 7.5cm, we employ a deep-learning model called MultiRes-UNet, which is an improved version of the original UNet network for building segmentation. The experimental results show an average IoU of 95.25% and average dice coefficient value of 97.56% for building segmentation using MultiRes UNet. The rooftops of segmented buildings are then extracted by

applying image processing techniques like finding contours, color filling and background subtraction.

The extracted rooftops are manually classified into three categories: flat, gable, and hip, and then sent into the second stage for roof type classification. Various deep learning models including a customized CNN model and three transfer learning approaches (ResNet50, EfficientNetB4, and VGG16) are used for classification and a comparative analysis is made between the models. Although the shallow CNN model has a slightly lower accuracy of 78%, it is still acceptable given the very limited data. With the AIRS dataset, we conclude that the transfer learning models have an average accuracy of 91.93%, indicating that the models have been fine-tuned to work with the sparse data we have. Because different DL models performed well on different portions, majority voting was employed as the assembling method. The results show that using majority voting improved classification accuracy by 5.67%.

This is finally followed by a simulation of modular layout of PV panels on rooftops based on customized inputs on the dimensions and tilt angle of the panel. The arrangement of PV panels also takes into consideration the obstructions (tree shadows, obstacles) to ensure maximal consumption of energy. This pipeline, thus, resolves the time-consuming, laborious task of manual assessment of rooftops.

## ABSTRACT

சுற்றுச்சூழல் பிரச்சினைகள் மனிதகுலத்தின் இருப்புக்கு அச்சுறுத்தலாக இருக்கிறது. காற்று மற்றும் சூரிய ஆற்றல் போன்ற புதுப்பிக்கத்தக்க எரிசக்தி ஆதாரங்களுக்கான மாற்றம் மனிதகுலம் தனது ஆற்றல் தேவைகளை மிகவும் நிலையான முறையில் பூர்த்தி செய்வதில் சிக்கலை எதிர்கொள்கிறது. அதிக திறன் கொண்ட சோலார் பேனல்களைப் பயன்படுத்துவதன் மூலம் சூரிய சக்தியின் பயன்பாட்டை மேம்படுத்துவது குறித்து பல ஆய்வுகள் நடத்தப்படுகின்றன. இருப்பினும், இந்த அணுகுமுறைகள் குறிப்பிடத்தக்க சவால்களை முன்வைக்கின்றன, ஏனெனில் அவை நேரத்தைச் செலவழிக்கும் மற்றும் விலை உயர்ந்தவை. ஒவ்வொரு கட்டிடத்தையும் மதிப்பிடுவதற்கும், PV பேனல் நிறுவல் ஏற்பாட்டைத் தீர்மானிக்க மற்றும் கூரையின் வகையைத் தீர்மானிப்பதற்கும் தகுதிவாய்ந்த நபர்கள் தளத்தைப் பார்வையிடுவதன் மூலம், அதிக அளவிலான மனித உழைப்பையும் இது கோருகிறது.

இதற்கிணங்க, இந்த திட்டம் சோலார் பேனல் வைப்பதற்கான கூரைகளை வகைப்படுத்துதல் மற்றும் கைமுறையாக மதிப்பிடுவதில் உள்ள சிக்கலைத் தீர்க்க ஒரு விரிவான பைப்லைனை உருவாக்குவதில் கவனம் செலுத்துகிறது. இந்த நோக்கத்திற்காக, நகர்ப்புற இயக்கவியலைப் புரிந்துகொள்வது முக்கியமானது. எனவே, பரந்த கவரேஜ் கொண்ட செயற்கைக்கோள் படம் கொடுக்கப்பட்டால், அசல் UNet நெட்வொர்க்கின்

மேம்படுத்தப்பட்ட பதிப்பான MultiRes UNet பயன்படுத்தப்படுகிறது. MultiRes UNet- ஐப் பயன்படுத்தி சராசரி IoU 95.25% மற்றும் சராசரி Dice coefficient 97.56% என சோதனை முடிவுகள் காட்டுகின்றன.

பிரித்தெடுக்கப்பட்ட கூரைகள் கைமுறையாக flat, gable மற்றும் hip என மூன்று வகைகளாக வகைப்படுத்தப்படுகின்றன. CNN மற்றும் மூன்று பரிமாற்ற கற்றல் அணுகுமுறைகள் (ResNet50, EfficientNetB4, VGG16) உள்ளிட்ட பல்வேறு நெட்வொர்க்கள் வகைப்படுத்தலுக்குப் பயன்படுத்தப்படுகின்றன. CNN நெட்வொர்க் 78% என்ற சற்றே குறைவான துல்லியத்தைக் கொண்டிருந்தாலும், மிகக் குறைந்த தரவைக் கொண்டுள்ளதால் இது ஏற்றுக்கொள்ளத்தக்கது. AIRS தரவுத்தொகுப்பைக் கொண்டு, பிற 3 நெட்வொர்க்கள் சராசரியாக 91.93% துல்லியத்தைக் கொடுக்கின்றன. இதற்கிணங்க ensembling (பெரும்பான்மையான வாக்களிப்பு) முறை பயன்படுத்தப்பட்டது. பெரும்பான்மை வாக்குகளைப் பயன்படுத்தி வகைப்படுத்தல் துல்லியம் 5.67% மேம்பட்டதாக முடிவுகள் காட்டுகின்றன.

இறுதியாக, பேனலின் பரிமாணங்கள் மற்றும் சாய்வு கோணத்தின் அடிப்படையில் கூரைகளில் PV பேனல்களின் அமைப்பு உருவகப்படுத்தப்படுகிறது. இத்தகைய அமைப்பின் மூலம் நாம் சூரிய ஆற்றலை அதிகப்படுத்த உள்ளோம்.

## **TABLE OF CONTENTS**

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
	<b>ABSTRACT - ENGLISH</b>	<b>iii</b>
	<b>ABSTRACT - TAMIL</b>	<b>v</b>
	<b>LIST OF FIGURES</b>	<b>x</b>
	<b>LIST OF TABLES</b>	<b>xii</b>
	<b>LIST OF ABBREVIATIONS</b>	<b>xiii</b>
<b>1.</b>	<b>INTRODUCTION</b>	<b>1</b>
	1.1 PROBLEM STATEMENT	1
	1.2 SUSTAINABLE ENVIRONMENT	2
	1.3 BUILDING SEGMENTATION	2
	1.4 ROOF TYPE CLASSIFICATION	4
	1.5 PV MODULE FITTING	5
	1.6 OVERALL OBJECTIVES	5
	1.7 SCOPE	6
	1.8 ORGANIZATION OF THESIS	6
<b>2.</b>	<b>RELATED WORK</b>	<b>7</b>



<b>3.</b>	<b>SYSTEM DESIGN</b>	<b>12</b>
	3.1 SYSTEM ARCHITECTURE	12
	3.2 DETAILED MODULE DESIGN	14
	3.2.1 BUILDING DETECTION	14
	3.2.2 ROOF TYPE CLASSIFICATION & BOUNDARY DETECTION	18
	3.2.3 PV MODULE FITTING	21
<b>4.</b>	<b>IMPLEMENTATION</b>	<b>23</b>
	4.1 DATASET	23
	4.2 BUILDING DETECTION	24
	4.2.1 CLIPPING ORIGINAL AERIAL IMAGES	24
	4.2.2 CLIPPING CORRESPONDING BINARY MASK IMAGES	25
	4.2.3 RESIZING & NORMALIZATION	26
	4.2.4 ADAPTING THE MULTI-RES UNET ARCHITECTURE	26
	4.2.5 APPLYING THRESHOLD	37
	4.2.6 ROOFTOP EXTRACTION	39
	4.3 ROOF TYPE CLASSIFICATION AND BOUNDARY DETECTION	42

4.3.1	PREPARATION OF DATASET & LABELING ROOF TYPE	42
4.3.2	DATA AUGMENTATION	44
4.3.3	ROOF TYPE CLASSIFICATION	45
4.3.4	MAJORITY VOTING	48
4.3.5	BOUNDARY DETECTION	50
4.4	PV MODULE FITTING	53
4.4.1	ROTATION OF PANELS	53
4.4.2	PV PANEL PLACEMENT	54
4.5	PERFORMANCE METRICS	57
4.6	TEST CASES	69
4.7	COMPARATIVE ANALYSIS	75
<b>5.</b>	<b>CONCLUSION AND FUTURE WORK</b>	<b>77</b>
	<b>REFERENCES</b>	<b>79</b>

## LIST OF FIGURES

3.1	Overall system architecture of building extraction and roof type classification for maximal PV panel installation	12
3.2	Module Design of Building Detection	16
3.3	Module Design for Roof type classification & Boundary detection	19
3.4	Module Design for PV Module Fitting	21
4.1	Code snapshot for image clipping	24
4.2	Original Aerial Image (left), Patches of image after performing clipping on the original image (right)	25
4.3	Ground Truth segmented image (left), Corresponding patched segmented image after clipping (right)	25
4.4	Aerial & ground truth mask after resizing & employing min max scaler	26
4.5	Architecture of MultiRes Block	28
4.6	Architecture of Res Path	29
4.7	Architecture of MultiRes UNet	30
4.8	MultiRes Block	30
4.9	Res Path	31
4.10	Defining the MultiRes model	32
4.11	Brief Model Summary	33
4.12	Training the model for 100 epochs	34
4.13	Results of building segmentation with MultiRes UNet model	36
4.14	Comparison of results before and after applying threshold	37
4.15	Comparison of ground truth images with our predictions before and after applying threshold	38
4.16	Code implementation of drawing contours	39
4.17	Bounding boxes drawn around the segmented mask images	40

4.18	Original satellite image (left), masked image after training on MultiRes UNet model (center), extracted rooftops (right)	41
4.19	Results of image after performing data augmentation	44
4.20	Model summary of CNN model	45
4.21	Architecture of VGG16 model	46
4.22	Architecture of ResNet50 model	47
4.23	Architecture of EfficientNetB4 model	48
4.24	Results on Potsdam dataset after applying majority voting	49
4.25	Images after performing white balancing to remove haze	50
4.26	Boundaries detected on rooftops by applying different techniques	52
4.27	Code implementation of rotation of points	54
4.28	Code implementation for region availability of solar panels	55
4.29	PV panels with dimensions length- 20mm; tilt angle- 40°; width- 10 mm (left) and length - 20mm; tilt angle - 40°; width -20mm (right)	56
4.30	Placement of PV panels when there are obstacles	56
4.31	Graphs of various performance metrics for building segmentation	59
4.32	Graphs showing training & validation accuracy, loss for CNN model	62
4.33	Confusion matrix and classification report on test data for CNN model	62
4.34	Graphs showing training & validation accuracy, loss for ResNet50	63
4.35	AUC-ROC metrics for ResNet50 model	63
4.36	Confusion matrix and classification report on test data for ResNet50	64
4.37	Graphs showing the training & validation accuracy, loss for EfficientNet	64
4.38	AUC-ROC metrics for EfficientNetB4 model	65
4.39	Confusion matrix & classification report on test data for EfficientNet	65
4.40	Graphs showing training & validation accuracy, loss for VGG-16 model	66
4.41	AUC-ROC metrics for VGG-16 model	66
4.42	Confusion matrix and classification report on test data for VGG-16	66

## LIST OF TABLES

4.1	Hyperparameter values used in training our MultiRes UNet model	35
4.2	Distribution of roof type dataset	42
4.3	Samples of images of each class from Christchurch, New Zealand	43
4.4	Performance results of the model on training and validation set	60
4.5	Results of different models on our manually labeled dataset	67
4.6	Results of testing our trained models on Potsdam dataset	68
4.7	Comparative analysis for building segmentation using MultiRes UNet	75
4.8	Comparative analysis of different DL models on AIRS dataset and dataset from Ankara, Turkey	76

## **LIST OF ABBREVIATIONS**

<b>AIRS</b>	Aerial Imagery for Roof Segmentation
<b>AUC</b>	Area under Curve
<b>FCN</b>	Fully Convolutional Network
<b>IoU</b>	Intersection over Union
<b>LBP</b>	Local Binary Pattern
<b>LIDAR</b>	Light Detection and Ranging
<b>MCC</b>	Matthews Correlation Coefficient
<b>PV</b>	Photovoltaic
<b>SVM</b>	Support Vector Machine

# **CHAPTER 1**

## **INTRODUCTION**

In this chapter, we will be discussing the need for a sustainable environment and potential methods undertaken to resolve climate change issues. This is followed by understanding the importance of remote-sensing in getting to know better about urban dynamics, and terrestrial mapping and building types. An overall synopsis of the project is briefly mentioned out.

### **1.1 PROBLEM STATEMENT**

Sustainable environment is required for the advancement of economic development, to improve energy security, access to energy, and mitigate climate change. Energy production sources such as coal, oil, and natural gas are responsible for one-third of global greenhouse gas emissions, and there is a growing need for everyone to switch to solar power for electricity generation. Estimating a roof's solar potential, on the other hand, is a time-consuming process that requires manual labor and site inspection. The current algorithms only work with LIDAR data and do not predict the number of solar PV panels based on the type of rooftops. Furthermore, mapping urban buildings for rooftop segmentation presents its own set of issues, since aerial satellite photos are typically of low resolution.

In this project, we use the AIRS dataset that provides a wide coverage of aerial imagery with 7.5 cm resolution and propose a mechanism to address the above problem: State-of-the-art MultiRes U-Net architecture is used for building detection as a first step to identify and segment buildings from aerial image. This is followed by classifying rooftops from the extracted buildings

with different deep learning and transfer learning algorithms and detecting the boundaries of rooftop. Finally, a fitting algorithm is used to provide a simulation of PV panels fitted on top of roofs. This proposed solution uses a single drone/satellite image to recognize and classify rooftops for buildings dispersed throughout an area, thus automating the entire process and reducing human labor.

## **1.2 SUSTAINABLE ENVIRONMENT**

Over the last few decades, climate change has become a global problem, affecting a variety of life forms. Sustainable development is thus the need of the hour for the survival of mankind. There is a rising awareness that harnessing renewable resources is the way to go in order to construct a sustainable environment. Potential tapping of solar power for generating electricity has gained enormous popularity and attention in recent years, and people are increasingly gravitating toward the PV revolution.

However, traditional approaches, such as online assessment are time-consuming and expensive, and they require a significant amount of human effort, as concerned experts must visit the site to inspect the building in order to determine whether PV panels should be installed and the modular layout of PV panel installation. However, by automating the process of building roof extraction for PV panel placement, a lot of money and time can be saved. In this way, PV sales can be made more efficient by using an automatic PV system design.

## **1.3 BUILDING SEGMENTATION**

We propose a 3-step mechanism as a solution to address this. AIRS dataset, which provides a wide coverage of aerial imagery with 7.5 cm



resolution is used here. The first stage is building detection from aerial satellite images. Image segmentation has gained huge traction with applications in sectors as diverse as medicine, pathology, and geo-sensing. Building segmentation, which is widely utilized in urban planning, topography mapping, disaster assessment, analyzing geographical land occupation, and other applications, has become a recent subject of interest due to the abundance of high-resolution remote sensing data. Despite this, low detection precision on aerial pictures limits automatic mapping of buildings.

A large portion of the effort entails manually demarcating buildings from satellite imagery. Feature extraction is the traditional method that is used to identify and segment different classes in remote sensing data. Buildings created for various purposes have diverse patterns of roof surface and border, which necessitates the use of state of the art deep learning models to create an exceptionally high-dimensional feature space.

The most famous UNet architecture, a CNN based model, that was originally developed for segmenting biomedical images has been used in the arena of geo-sensing for aerial image segmentation with few modifications. Over the years, numerous other models have been tried, with U-Net model being the baseline one. This includes, but is not limited to, FCN, DeepLabv3, and ICTNet. A novel framework called ICTNet, leverages border localization for classification and reconstruction of buildings. Here, the model combines the localization accuracy and use of context of the UNet network architecture, with Dense and Squeeze-Excitation (SE) layers.

Similarly, customized CNN models are currently used for image segmentation, specifically building segmentation. However, most of the models weren't accurate in delineating the edges/boundaries of buildings accurately. In this project, we propose a deep learning framework called

MultiRes UNet that has demonstrated efficacy in multimodal bio-medical image segmentation. Instead of using the direct connection as in UNet, an additional path called the ResPath with convolution operations is used to connect the encoder and decoder.

## **1.4 ROOF TYPE CLASSIFICATION**

Understanding roofs is very essential as different buildings have different layouts and types of roofs. The most general classes of roofs include, but are not limited to, gable, hip, half-hip, pyramid, flat, industrial roofs. Roofs are no longer solely for the purpose of providing shelter and drainage in the ever-changing world we live in. Their material compositions, construction methods, and design have all changed throughout time, thus determining the type of roof is critical. Following building segmentation, we undertake background subtraction to extract the rooftops from aerial images and classify them. This process yields us the rooftops and we intend to manually annotate the rooftops into different classes.

In recent years, several deep learning models including pre-trained models like VGG16, ResNet50, and EfficientNetB4 are used for rooftop classification for different types and results show that these models have been very efficient in classifying with an accuracy of 80%. Similarly, machine learning models like Support Vector Machine (SVM) have been used for classifying rooftops from LiDAR data in Geneva, Switzerland to estimate solar energy potential with an accuracy of 66%. The study helped us understand how solar roof-shape classification may be used in new building design, retrofitting existing roofs, and efficient solar integration on building rooftops. In a similar way, we plan to apply customized CNN models with

pre-trained models such as VGG16, ResNet50 and provide a comparison on the performances of these models.

## **1.5 PV MODULE FITTING**

Solar roofs are a cheaper and more sustainable energy source. Installation of PV panels on rooftops make effective use of roof cover space while also generating electricity for the building's usage. Finally, based on the type of roofs, a simulation of modular layout of PV panels fitted on top of roofs is provided. In the case of flat roofs, we obtain the solar tilt angle, as well as the length and width of PV panels, and execute PV panel simulation. The same technique is carried out on tilted roofs, with the tilt angle being very small, and the successful positioning of PV panels is eventually demonstrated.

## **1.6 OVERALL OBJECTIVES**

To simulate the placement of PV panels on rooftop of buildings from aerial images based on roof type for maximum energy consumption, the following steps are to be performed:

- To segment and detect buildings in a given satellite image using MultiRes UNet model and perform background subtraction to extract the rooftops.
- To label the extracted building rooftops into different classes - flat, gable, and hip and create a dataset and train different DL models for roof type classification.
- To perform edge detection on the extracted rooftops to mark boundaries to find the necessary area for PV panel simulation.

## **1.7 SCOPE**

The primary goal of the project is to classify roof tops and provide a simulation of modular layout of PV panels fitted on top of roofs. Even though, manual assessment is more reliable, it is often time consuming and labor intensive when it comes to covering numerous buildings in a particular area. That is why automating the entire process is necessary. We must note that so far no automated approaches have been used to detect buildings from a wide coverage satellite image and classify roof types. This is where our one-shot pipeline comes into play, which can segment and detect buildings from aerial/satellite images, extract each of the rooftops and classify them based on their type in order to provide a simulation of PV panels fitted on them. This can easily automate the entire process of assessing hundreds of roof tops.

## **1.8 ORGANIZATION OF THESIS**

The details of the system architecture, implementation, and evaluation of the proposed three-staged pipeline will be discussed in the following chapters.

In Chapter 2, the relevant architectural models employed in previous works for building segmentation and deep learning models for roof type classification are discussed. A summary of issues along with proposed solutions is discussed. Chapter 3 gives the system design of the pipeline along with the detailed analysis of each of the modules. The implementation details, evaluation and performance measures are presented in Chapter 4. In addition, test cases and comparative analysis between the proposed system and existing models are analyzed. Finally in chapter 5, we conclude the proposed work and also list the future work that can be done on it.

## **CHAPTER 2**

### **RELATED WORK**

Over the years, image databases and computer vision have gotten a lot of attention. Many breakthroughs in image segmentation have been made as a result of this. Bio-medical and remote sensing are the two industries where segmentation is critical in understanding the images.

With respect to this, a multi-scale convolutional neural network that adopts an encoder-decoder U-Net architecture is utilized by X. Li, Y. Jiang, H. Peng and S. Yin (2019) for building segmentation. Here, a U-Net is constructed as the main network, and the bottom convolution layer of U-Net is replaced by a set of cascaded dilated convolution with different dilation rates and an auxiliary loss function added after the cascaded dilated convolution. The proposed method has achieved an IOU of 74.24 % on the whole dataset that covers regions like Austin, Chicago, and Vienna. The proposed model has aided in network convergence, and a major advantage is that it does not involve manual features and does not involve preprocessing or post-processing steps. However, the segmentation of middle parts in buildings are misaligned and the bulges on the boundaries are lost. Yet another major issue is that the algorithm performs well only in one subset (countryside and forest) but not in another.

By scraping data from Google Maps, V. Golovko, S. Bezobrazov, A. Kroshchanka, (2017) developed a CNN-based solar PV panel recognition system. Pre-processing techniques such as image scaling and sharpening are used, followed by the training of a 6-layer CNN model. Instead of using high resolution color satellite orthoimagery, the authors employed low-quality

satellite imagery (Google Maps satellite photos), which allows them to reduce the approach's requirements. Because some solar panels resemble roof tops, poor quality satellite pictures taken from Google Maps have led to erroneous classification and there is no validation on the dataset.

A mask R-CNN with three steps is proposed to extract buildings in the city of Christchurch from aerial images post-earthquake by Chen, Mengge and Jonathan Li (2019) to recognize small detached residences to understand the havoc caused by it. Feature extraction with ResNet is the first step. The RPN (Regional Proposal Network) is then utilized to locate RoI and filter out the irrelevant bounding boxes using object and background classification, as well as bounding box regression. The background and buildings are then identified by object classification. The RoIAlign method used instead of the RoI Pool gives better feature extraction. However due to a small training dataset, the model was unable to successfully demarcate building edges, resulting in low accuracy and precision when compared to other SOTA models.

A combination of image processing techniques, including Adaptive Edge Detection and contours are used by Kumar, Akash & Sreedevi, Indu. (2018) to segment out rooftop boundaries and obstacles present inside them along with polygon shape approximation. It provides a comparative analysis of the solar potential of buildings. Several types of the rooftop are considered to learn the intra-class variations. Because Google Maps India's satellite resolution is so low, the edges aren't fully identified, and there are outliers plotting solar panels outside of the building's rooftop area.

The problem of semantic segmentation of buildings from remote sensor imagery is addressed by a novel framework called the ICTNet proposed by B. Chatterjee and C. Poullis (2019). ICTNet: a novel network with the underlying

architecture of a fully convolutional network, infused with feature re-calibrated Dense blocks at each layer is combined with dense blocks, and Squeeze-and-Excitation (SE) blocks. Dense blocks connect every layer to every other layer in a feed-forward fashion. Along with good gradient propagation they also encourage feature reuse and reduce the number of parameters substantially as there is no need to relearn the redundant feature maps which allows the processing of large patch sizes.

Reconstruction is done by extruding the extracted boundaries of the buildings and comparative analysis is made between the two. With no 3D information on the buildings, the authors have used the building boundaries as a proxy for the reconstruction process and have got better overall IoU compared to other methods. The main limitation here is that there is no loss function for the reconstruction accuracy. Furthermore, due to the fact that ground truth photos used for training contain mistakes and are manually generated, there is a large variance in per-building IoU. In addition, the reconstruction accuracy is consistently lower than classification accuracy by an average of  $4.43\% \pm 1.65\%$ .

Satellite images from Google Earth for the city of Heilbron are collected and manually labeled following which object proportion distribution is used in image-level by Peiran Li, Haoran Zhang, Zhiling Guo, Suxing Lyu (2021) and object occurrence possibility at pixel level is statistically analyzed. SOTA PV segmentation model (DeepSolar) is used to extract visual features. Local Binary Pattern (LBP) is used for texture feature extraction & color histograms for color feature extraction. The authors have addressed the issue of class imbalance of PV and non-PV panels on rooftops by hard and soft sampling. The major drawback is that lighting conditions caused distinct color clustering groups in PV/Non-PV color clustering, resulting in

misclassification along with IOU being less than the acceptable range (0.5) for 1.2m resolution images.

Rooftop classification is the important step in identifying the type of PV panels that can be fitted. M. Buyukdemircioglu, R. Can, S. Kocaman (2021) have undertaken research to generate a roof type dataset from very high-resolution (10 cm) orthophotos of Cesme, Turkey, and to classify the roof types using a shallow CNN architecture. UltraCam Falcon large-format digital camera is used to capture orthophotos with 10cm spatial resolution and roofs are manually classified into 6 different labels. The prediction is investigated by comparing with three different pre-trained CNN models, i.e. VGG-16, EfficientNetB4, and ResNet-50.

Simple CNN models are hence easier to implement and require nominal hardware specifications. The shallow CNN model has achieved 80% accuracy. As the roof images were clipped automatically from orthophotos, there are few buildings with overlap. Half-hip roofs are not classified properly and the F1 score obtained for them is very low. The authors haven't experimented with alternate hyperparameter tweaking for the shallow CNN architecture, which is a serious flaw.

SVM based machine learning approach is used to classify building roofs in relation to their solar potential. The SVM classifier used by Nahid Mohajeri, Dan Assouline, Berenice Guiboud (2018) on an average produces 66% accuracy and is able to classify rooftop types into 6 major classes. The authors have calculated the ratio of useful roof area for each type of roof shape to that of the corresponding building footprint area and the results are close to a value of 1. This indicates that better the segmentation of building, maximum is the solar potential of each of the rooftop areas.

Although the above models address the issue of building segmentation



to a great extent, there are certain major flaws. In situations where shadows, vegetation and parking lots and other obstacles enclose buildings, the above frameworks showed poor results in building detection. In addition, there was no clear distinction to demarcate adjacent buildings and boundaries of buildings were not properly detected. Furthermore, when a satellite or aerial image spanning a large area with several buildings was provided, only manual cropping was done to extract roofs for classification, and there were no automated mechanisms in the approaches outlined above.

Some of the studies mentioned calculate the discrete number of PV modules that should be placed on the available roof space, but none of them computed the actual modular layout that would fit on each roof segment. Additionally, shadows and other obstacles were often neglected while considering the placement of PV panels which resulted in bad prediction of energy consumption.

To address the above problems, we propose a pipeline for building segmentation and roof type classification with a simulation of PV panels fitted on it. In this work, we adapt MultiRes UNet, with additional convolution layers to extract spatial features at various scales. The skip connections in UNet are replaced with a chain of convolution operations to reduce the semantic gap between encoder and decoder features and hence building segmentation results can be improved drastically. This is then followed by a technique called background subtraction to extract the rooftops automatically which resolves the problem of manual cropping and selection of roofs. Roof type classification is performed with deep learning models, following which we use edge detection algorithms to simulate a modular layout of PV panels fitted on top of roofs which considers several attributes including obstructions to maximize the solar energy consumption.

## CHAPTER 3

### SYSTEM DESIGN

This chapter provides a high level architecture diagram of the proposed three-staged pipeline along with the detailed module design.

#### 3.1 SYSTEM ARCHITECTURE

The proposed model's overall system architecture is depicted in fig. 3.1. The pipeline consists of 3 modules: (i) Building segmentation and detection, (ii) Roof type classification and boundary detection, and (iii) PV module fitting.

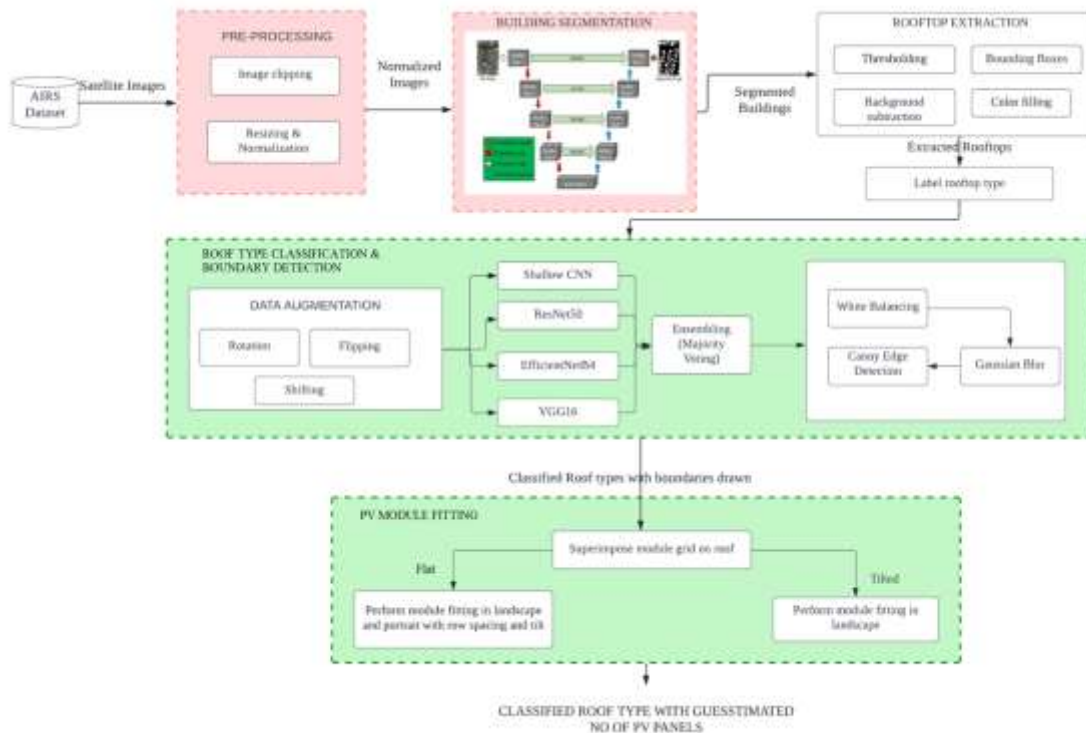
The AIRS dataset contains satellite images of Christchurch in New Zealand with 7.5cm resolution. The first module mainly emphasizes building detection and rooftop extraction from satellite images. The first stage of building segmentation includes two pre-processing steps: clipping satellite images and the accompanying ground truth masks, followed by scaling and normalization. The normalized images are then trained using the MultiRes UNet architecture for segmenting the buildings. MultiRes UNet is chosen here as it has provided great results with image segmentation in previous works. Subsequently, image processing techniques like finding contours, drawing bounding boxes and background subtraction are performed to extract the rooftops of buildings.

The extracted rooftops from the previous module are sent to the second stage of the pipeline. The rooftops are manually labeled into 3 different classes: namely flat, gable, and hip which are fed to four deep learning models for classification and a comparative analysis is made. An ensemble method

(majority voting) is used to further improve the accuracy of classification.

Following that, white hazing and auto canny edge detection is performed to identify and mark boundaries on rooftops for fitting PV module grids.

The final module resorts to providing a simulation of modular layout of PV panels fitted on top of roofs. This is accomplished by applying a panel fitting algorithm which considers the dimensions (length, width, tilt angle) of the PV panel, type of roof and obstructions on roof top to find the placement of panels.



## 3.2 DETAILED MODULE DESIGN

This chapter gives a description on the different modules involved in this project and an overview on what each module intends to do along with module wise block-diagrams and intermediate deliverables of each module.

Deep learning based methods, especially UNet models, exhibit great prediction performances on image segmentation tasks. The first module of this project involves building detection and rooftop extraction using MultiRes UNet segmentation. The dataset used here is the AIRS (Aerial Imagery Roof Segmentation) dataset with 857 aerial images in training set and 90 each in testing and validation set with original spatial dimensions of  $10000 \times 10000$  and spatial resolution of 7.5 cm. Given the large dimensions of satellite images, we clip the original images into size of  $1536 \times 1536$ . As a result, we use 1548 images for training and 144 images each in the validation and testing set.

The extracted rooftop images from the above segmentation are manually labeled into four different classes - namely, Gable, Flat, Complex and Hip. Since this contains around 1000 images, we could further increase the dataset size by augmentation techniques such as rotation, flipping, shifting. The detailed module design for each of the modules is presented below.

### 3.2.1 BUILDING DETECTION

#### 3.2.1.1 *Pre-Processing*

Before building segmentation, a set of pre-processing steps are to be followed. Aerial satellite images from AIRS dataset are of very large dimensions ( $10000 \times 10000$ ). Owing to the hardware constraints for training the model, as a first step in pre-processing we clip the images into smaller

images called ‘patches’ or ‘tiles’ to  $(1536 \times 1536)$  dimensions. By applying this approach, we generate 1548 images for the training set and 144 images each in the testing and validation set. This is followed by resizing the images to  $256 \times 256$  dimensions and further applying MinMax scaler as the normalization technique. Small patches of normalized images are sent to the next layer for model training.

### **3.2.1.2 Building Segmentation**

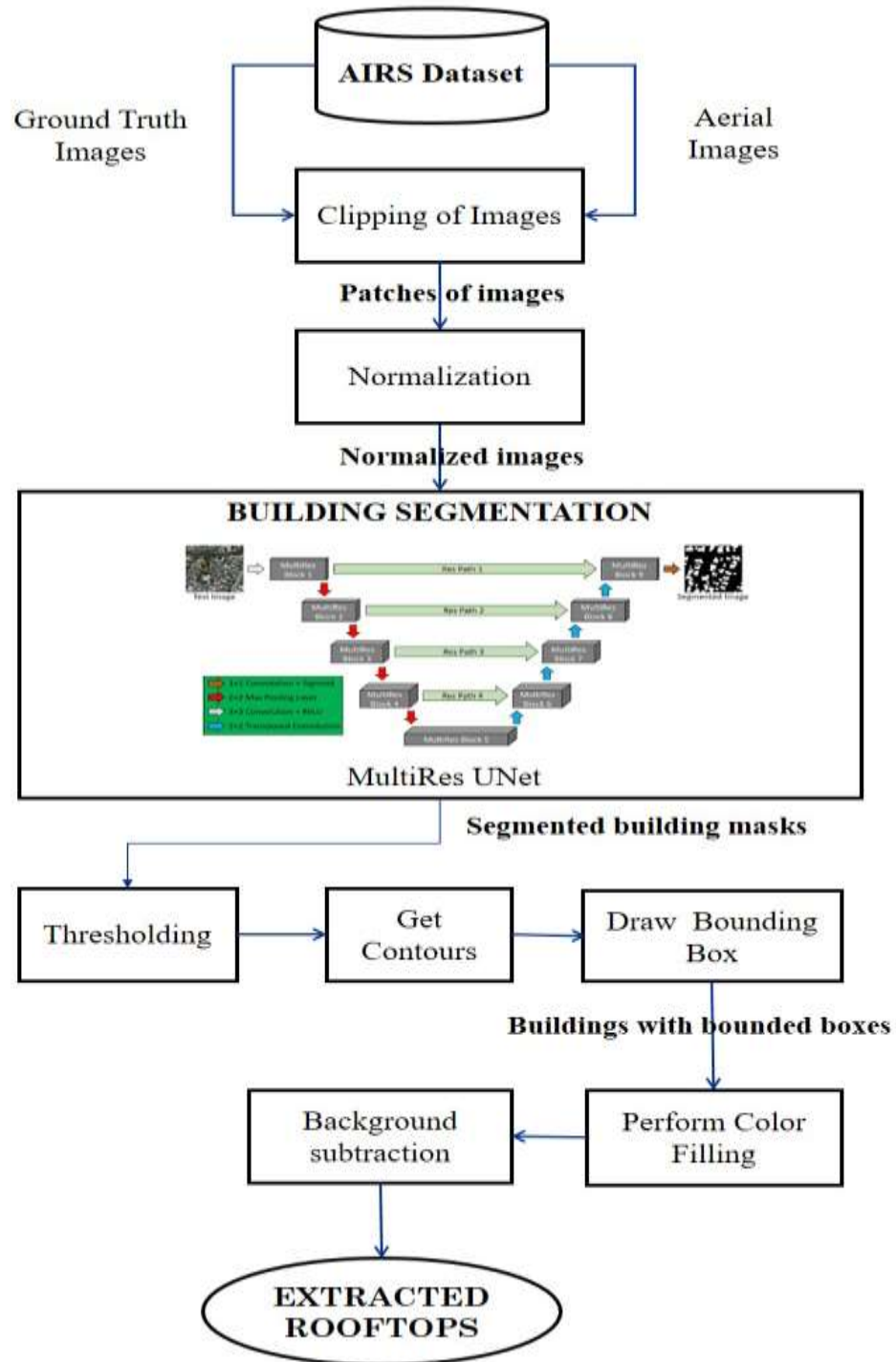
Once preprocessing is done, the images are fed into the MultiRes UNet model. The architecture of the model is discussed in the Implementation section below. The model is trained for 100 epochs and Adam optimizer is used for stochastic gradient descent. After tweaking the hyperparameters, we found that a batch size of 8 with learning rate = 0.0001 works well for our model. The loss function used here is binary cross entropy as we have two classes here - building and the background. ReLu is used as the activation function in the top layers and sigmoid is used as the activation function at the last layer. The segmented masks of buildings are compared with original ground truth masks and performance analysis is carried out.

### **3.2.1.3 Rooftop Extraction**

The segmented masks of buildings from previous layers undergo thresholding here. Thresholding is done to detect the edges more accurately. After this, contours around the buildings are identified and a bounding box is drawn to the buildings. Following this, background subtraction takes place to extract rooftops from widely dispersed satellite images and they are saved and stored in a new database.

**INPUT:** AIRS Dataset

**OUTPUT:** Rooftops of different buildings



**Fig. 3.2 Module Design of Building Detection**

## **PSEUDO-CODE**

*Begin*

### **Pre-Processing**

- a. *Get the original dimensions and the dimensions, stride of smaller patches.*
- b. *Find overlapping areas.  $\text{overlapping} := (\text{size} / \text{stride}) - 1$*
- c. *for  $i$  in range  $(\text{orig\_shape} / \text{stride} - \text{overlapping})$ :*
- d. *Crop from  $i*\text{stride}:i*\text{stride}+\text{size}$ ,  $j*\text{stride}:j*\text{stride}+\text{size}$*
- e. *Perform resizing and normalization.*

### **Training the MultiRes UNet Architecture**

- f. *Split dataset into train - 1548 images, validation - 72 & testing - 144*
- g. *Define the MultiRes model.*
- h. *do:*
  - i. *Tweak hyperparameters: (lr, batch size, epochs, optimizer, loss).*
  - ii. *Train the model.*
  - iii. *Plot graphs and check on validation data.*
- i. *while (find the best model)*
- j. *Save weights for the best model.*

### **Background Subtraction**

- k. *Convert RGB to grayscale.*
- l. *Find contours.*
- m. *Find the bounding box coordinates  $:= (x, y), (x+w, y), (x, y+h), (x+w, y+h)$*
- n. *Perform color filling.*
- o. *Remove background from foreground buildings.*

*End*

## **3.2.2 ROOFTYPE CLASSIFICATION AND BOUNDARY DETECTION**

### ***3.2.2.1 Data Augmentation***

The extracted rooftops from previous steps are manually labeled into four classes - Flat, Gable, Complex, and Hip. Data augmentation is done here to increase the size of the dataset for the roof type classification to increase the accuracy as well as to prevent over-fitting. The techniques used for data augmentation include rotation, shifting, flipping.

### ***3.2.2.2 Classification of roof type images with ensembling approach***

Four deep learning architectures are used here for roof type classification. Shallow CNN is used as a baseline model. Three pre-trained models ResNet50, EfficientNetB4 and VGG16 are also used and a comparative analysis is made. Further, majority voting is used as an ensembling approach to predict the roof type of unseen images.

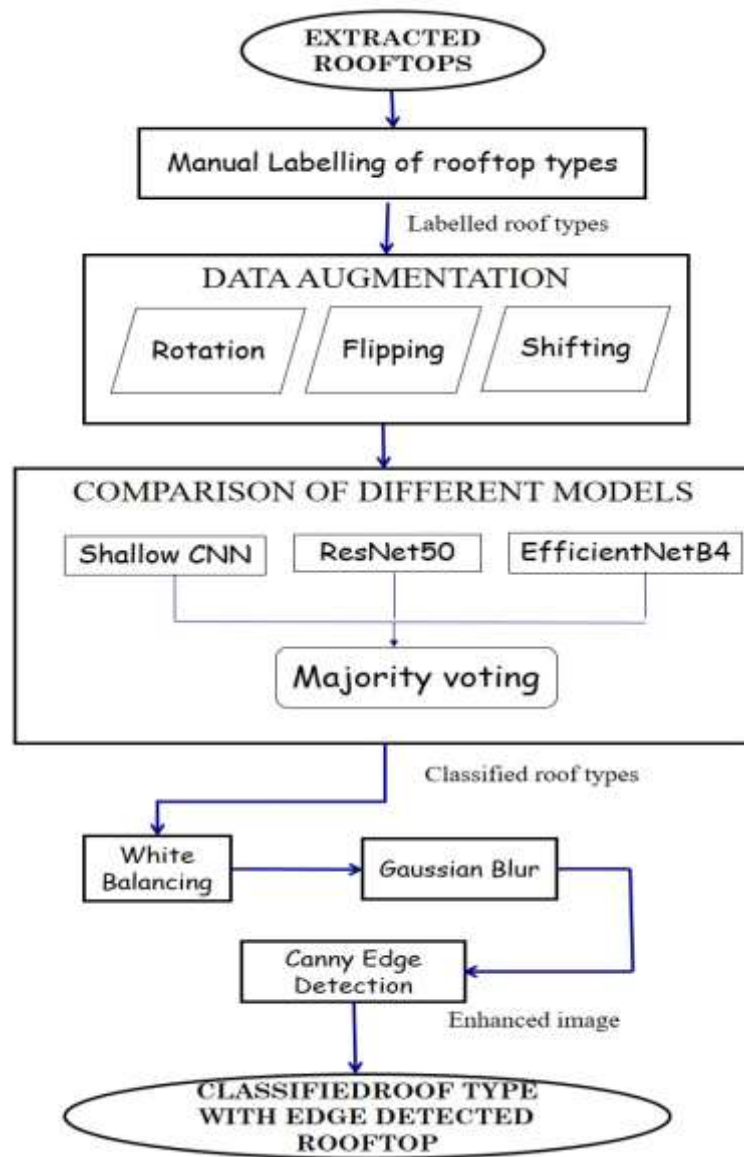
### ***3.2.2.3 Boundary Detection***

The extracted roof tops are now further enhanced by applying white balancing to remove haze and then fuzzy histogram hyperbolization takes place. This is followed by applying Gaussian blur and using Auto Canny Edge Detection for finding the rooftop boundaries.



**INPUT:** Extracted rooftops.

**OUTPUT:** Classified roof type with edge detected rooftop.



**Fig. 3.3 Module Design for Roof type classification & Boundary detection**

## **PSEUDO-CODE**

*Begin*

### **Data Augmentation**

- a. Rotation - Randomly generate an angle ( $\theta$ ) and rotate the image by ( $\theta$ ) degrees clockwise and anticlockwise.*
- b. Shifting - ImageDataGenerator (width\_shift\_range = 0.10)*
- c. Flipping - ImageDataGenerator (horizontal\_flip = True)*

### **Training different deep learning models**

- d. Split the dataset into train - 80%, validation - 10% and testing- 10%.*
- e. Define a customized CNN model.*
- f. Train:*
  - i. Tweak hyperparameters (Learning rate, batch size, and optimizer).*
- h. Use pre-trained ResNet50, EfficientNetB4 and VGG16.*
- i. Fine tune the pre-trained models.*
- j. Repeat above steps for the other 2 transfer learning models.*
- k. For an unseen image:*
  - i. Tweak hyperparameters (Learning rate, batch size, and optimizer).*
  - ii. Perform majority voting to identify the class that gets the maximum votes.*

### **Boundary Detection**

- l. White patching - Set a percentile score and remove haze from the image.*
- m. Gaussian Blur - Try for different kernel size and  $\sigma$ .*
- n. Find the threshold for adaptive canny edge technique.*

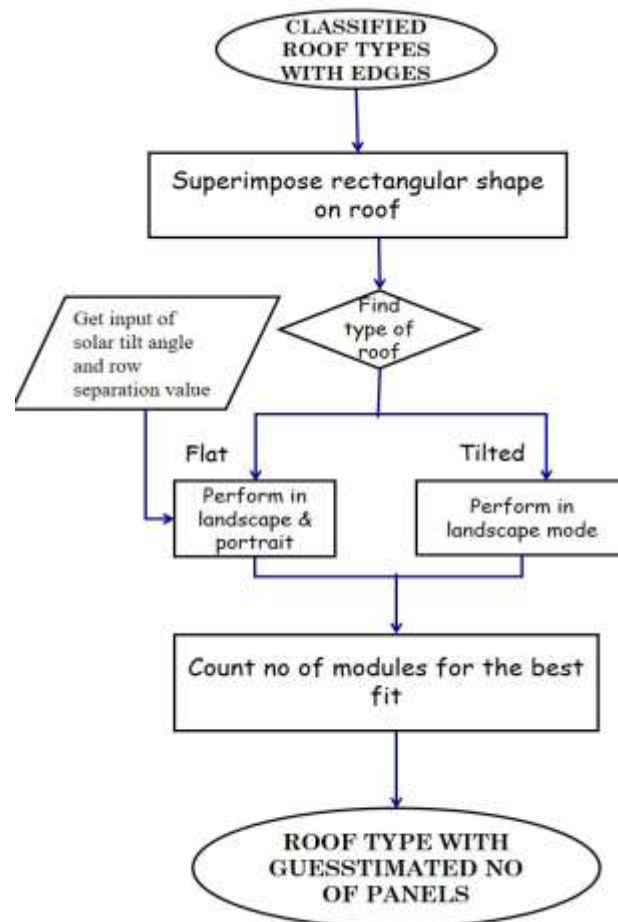
*End*

### 3.2.3 PV MODULE FITTING

The boundary detected rooftops are superimposed with a rectangular PV module shaped grid. Based on the type of roof, module fitting happens. PV module dimensions (length, width, tilt angle) along with the type of roof is taken into consideration and a simulation of PV modules fitted on top of roofs is provided as the final end output.

**INPUT:** Classified roof type with edge detected rooftop.

**OUTPUT:** Classified roof type with simulation of PV panel on rooftops.



**Fig. 3.4 Module Design for PV Module Fitting**

## PSEUDO-CODE

*Begin*

- a. Get dimensions of PV panel from user (length, width, and spacing).*
- b. Get tilt angle based on the type of roof.*
- c. if roof type is flat:*  
     $\text{tilt angle} := (20^\circ - 50^\circ)$   
    *elif roof type is tilted:*  
         $\text{tilt angle} := (0^\circ - 10^\circ)$
- d. Create a rectangular module grid according to the settings specified.*
- e. do:*

*Identify new points after projecting the tilted PV panels.*

$$(x', y') := (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta)$$

*Find contour area.*

*if area < 5:*

*Move to the next point as this is obstacle or smaller patch  
unsuitable for panel placement*

*else:*

*Fix a PV panel on the identified points.*

*if flat:*

*Shift the module grid with specified row spacing.*

*else:*

*Shift the module grid with 0 spacing.*

*Move to the next point on the line segment and fix panel there.*

*while (entire roof top is analyzed)*

*End*

## **CHAPTER 4**

### **IMPLEMENTATION DETAILS**

This chapter discusses the implementation of each of the three modules of the pipeline in detail along with the methodologies used, code and result snapshots, evaluation of performance and comparative analysis.

The whole process of training and testing the presented network for building detection and rooftop classification is executed under TensorFlow backend and Keras framework in Colab Pro with a memory of 16 GB, T4 GPU. The three modules are integrated together and a simple web application is developed with HTML, CSS for front end and Flask for the backend.

#### **4.1 DATASET**

The dataset used in this study is AIRS (Aerial Imagery for Roof Segmentation) dataset, which covers the full area of Christchurch in New Zealand and provides a wide coverage of aerial imagery with a spatial resolution of 7.5 cm and spatial dimensions of 10000 \* 10000. The training set has 857 images and the validation and testing set each contains 90 images. The dataset has aerial satellite images along with the corresponding ground truth mask images. The roof type classification dataset has 1115 images divided into three subcategories: Flat, Gable, and Hip, with 1020 images in the training set, 50 images in the validation set, and 45 images in the testing set.

## 4.2 BUILDING DETECTION

### 4.2.1 Clipping original satellite images

Satellite and aerial images are typically too large to be segmented directly and hence are usually cut into smaller patches or tiles. The satellite photographs are  $10000 \times 10000$  pixels, and the first step is to reduce the original satellite images to  $1536 \times 1536$  pixels. Fig. 4.1 gives the implementation of sliding window technique used where the size of smaller patches are of dimensions  $1536 \times 1536$  and stride length is of  $1536 \times 1536$ . Thus, we obtain 36 smaller patches for a single satellite image. This is done for training, validation, testing images and resulting in a dataset with 1548 images in the training set, 72 images in validation and 144 in the testing set.



**Fig. 4.1** Code snapshot for image clipping

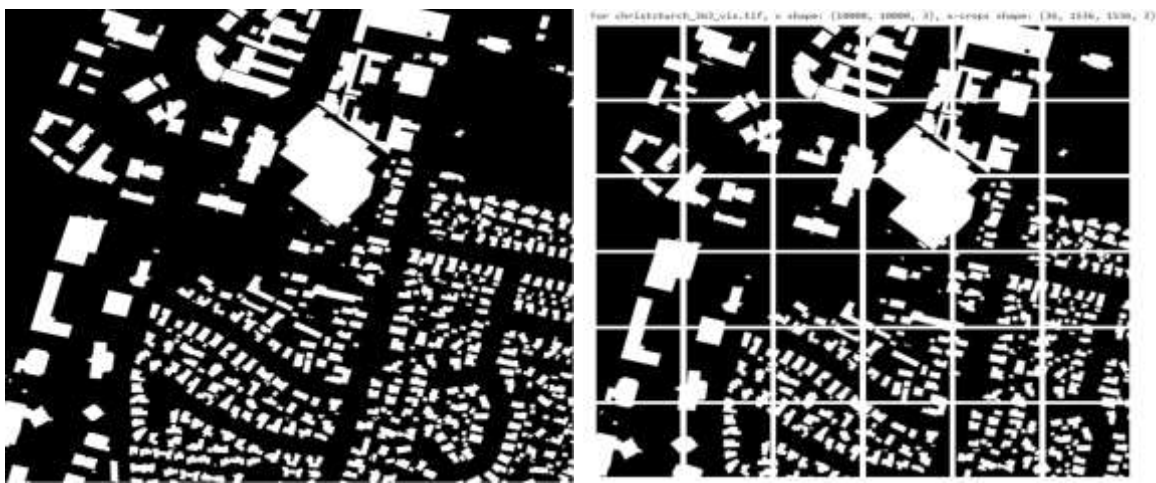


**Fig. 4.2 Original satellite image (left), Patches of image after performing clipping on the original image (right)**

Fig. 4.2 provides a sample of original satellite images which is then cut into 36 smaller patches by the above mentioned technique.

#### ***4.2.2 Clipping corresponding binary mask images***

The same technique is applied to the ground truth masks and fig. 4.3 shows the corresponding mask images and its clipped satellite images.



**Fig. 4.3 Ground truth segmented image (left), Corresponding patched segmented image after clipping (right)**

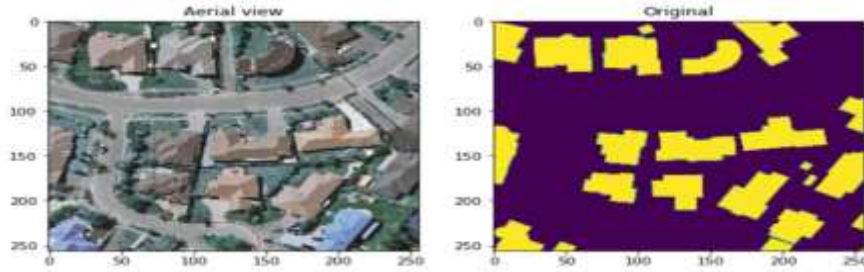


### 4.2.3 Resizing and Normalization

Images of size  $1536 \times 1536$  are still huge to process and train on complex neural nets as we do not have the required hardware specifications. As a result, the images are resized to  $256 \times 256$  dimensions with INTER\_CUBIC interpolation method as it leads to better resolution of images. MinMax scaler is employed for normalization as the upper and lower boundaries are well known (e.g. for images where pixel intensities go from 0 to 255 in the RGB color range). Fig. 4.4 indicates the image after performing resizing and normalization.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3.1)$$

**Min-max scaler**



**Fig. 4.4 Aerial view and ground truth mask after resizing and employing min-max scaler**

### 4.2.4 Adapting the MultiRes UNet architecture

The architecture of the MultiRes UNet network consists of 2 important blocks: the MultiRes Block and the Res Path.

#### 4.2.4.1 MultiRes Block

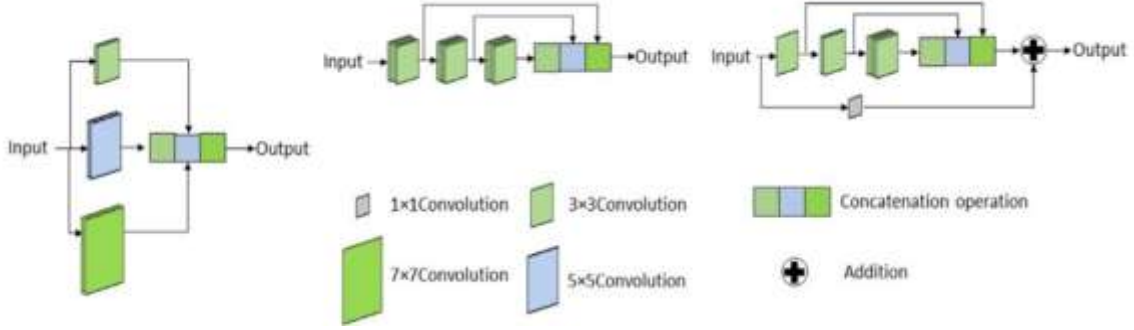
A sequence of two  $3 \times 3$  convolutional layers is utilized after each pooling layer and transposed convolutional layer in the U-Net design. A single



$5 \times 5$  convolutional operation is equal to a series of two  $3 \times 3$  convolutional processes. As a result, incorporating  $3 \times 3$ ,  $5 \times 5$ , and  $7 \times 7$  convolution operations in parallel is the simplest way to augment U-Net with multi-resolution analysis, following the strategy of Inception network. By replacing the convolutional layers with Inception-like blocks, the U-Net architecture should be able to reconcile the features learned from the image at different sizes.

The addition of more convolutional layers in parallel, on the other hand, dramatically increases the memory required. As a result, we use a sequence of smaller and lighter  $3 \times 3$  convolutional blocks to factorize the larger, more demanding  $5 \times 5$  and  $7 \times 7$  convolutional layers. The outputs of the second and third  $3 \times 3$  convolutional blocks, respectively, effectively mimic the  $5 \times 5$  and  $7 \times 7$  convolution processes. We hence concatenate the outputs from the three convolutional blocks to extract spatial information from different sizes. Fig. 4.5 depicts the connections between different convolutional layers in the MultiRes block.

Despite the fact that this tweak reduces the RAM requirement significantly, it is still fairly demanding. This is related to the fact that when two convolutional layers are present in a deep network in succession, the number of filters in the first one has a quadratic effect on the memory. Thus rather than keeping all three subsequent convolutional layers with the same number of filters, we gradually increase the number of filters in those layers (from 1 to 3), to prevent the memory requirements of the earlier layers from propagating too far into the network. In order to interpret some more spatial information, we also add a residual link and a  $1 \times 1$  convolutional layer.



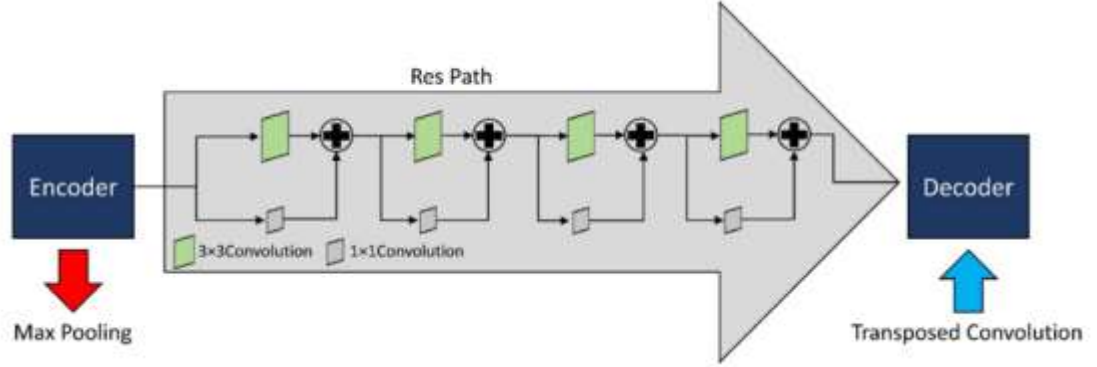
**Fig. 4.5 Architecture of MultiRes Block**

#### **4.2.4.2 Res Path**

Skip connections allow the network to transfer spatial information from encoder to decoder that is lost during the pooling procedure. Even if the dispersed spatial elements are preserved, there may be problems. In fig. 4.6, the first shortcut link connects the encoder to the decoder following the last deconvolution step before the first pooling. The encoder's features are expected to be lower level features because they are computed in the network's earlier layers. The decoder characteristics, on the other hand, are expected to be of a considerably higher level, as they are computed at the network's deepest layers. Therefore, they go through more processing.

As a result, there may be a semantic gap between the two sets of features that are being merged. This is due to the fact that not only are encoder features going through extra processing, but we're also fusing them with decoder features from much younger layers. Convolutional layers are used along the shortcut links to reduce the discrepancy between the encoder and decoder characteristics. The additional non-linear transformations applied to the features propagating from the encoder stage will account for the additional processing performed by the decoder step. In addition, instead of employing

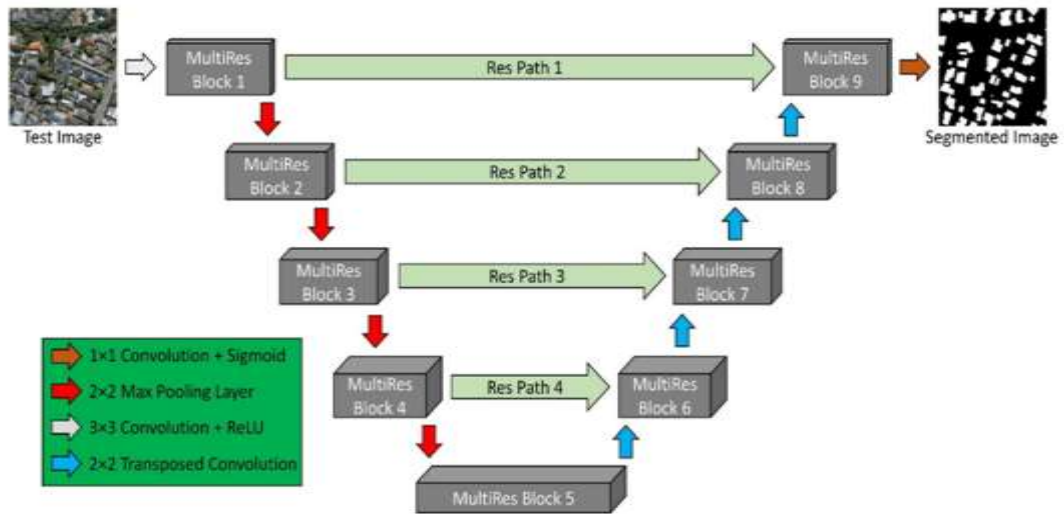
the standard convolutional layers, residual connections are inserted and concatenated with the decoder features.



**Fig. 4.6 Architecture of Res Path**

#### 4.2.4.3 Architecture

Four MultiRes blocks are each used in the encoder and decoder stage as presented in fig. 4.7. The number of filters in each of the MultiRes blocks is based on the formula:  $W = \alpha \times U$  where  $\alpha$  is the scalar coefficient whose value is set to 1.67 and  $U$  refers to the no of filters. The parameter  $W$  preserves an analogous connection between the suggested MultiRes-UNet network and the main UNet network. After every pooling or transposing of layers, the value of  $W$  became double, similar to the original UNet network. The values of  $U = [32, 64, 128, 256, 512]$  are set as follows. We also allocated filters of  $[W/6]$ ,  $[W/3]$ , and  $[W/2]$  to the three succeeding convolutions respectively. The filters used in the Res Path are as follows: 64, 128, 256, and 512. Batch Normalization is performed in each of the blocks to avoid overfitting. ReLU is used as the activation function in all the convolution layers and the last layer employs the Sigmoid activation function as we have only binary classes here (0 indicating background, 1 indicating the building).



**Fig. 4.7 Architecture of MultiRes UNet**

```
def MultiResBlock(U, inp, alpha = 1.67):

    W = alpha * U

    shortcut = inp

    shortcut = conv2d_bn(shortcut, int(W*0.167) + int(W*0.333) +
        int(W*0.5), 1, 1, activation=None, padding='same')

    conv3x3 = conv2d_bn(inp, int(W*0.167), 3, 3,
        activation='relu', padding='same')

    conv5x5 = conv2d_bn(conv3x3, int(W*0.333), 3, 3,
        activation='relu', padding='same')

    conv7x7 = conv2d_bn(conv5x5, int(W*0.5), 3, 3,
        activation='relu', padding='same')

    out = concatenate([conv3x3, conv5x5, conv7x7], axis=3)
    out = BatchNormalization(axis=3)(out)

    out = add([shortcut, out])
    out = Activation('relu')(out)
    out = BatchNormalization(axis=3)(out)

    return out
```

**Fig. 4.8 MultiRes Block**

```

def ResPath(filters, length, inp):

    shortcut = inp
    shortcut = conv2d_bn(shortcut, filters, 1, 1,
                        activation=None, padding='same')

    out = conv2d_bn(inp, filters, 3, 3, activation='relu', padding='same')

    out = add([shortcut, out])
    out = Activation('relu')(out)
    out = BatchNormalization(axis=3)(out)

    for i in range(length-1):

        shortcut = out
        shortcut = conv2d_bn(shortcut, filters, 1, 1,
                            activation=None, padding='same')

        out = conv2d_bn(out, filters, 3, 3, activation='relu', padding='same')

        out = add([shortcut, out])
        out = Activation('relu')(out)
        out = BatchNormalization(axis=3)(out)

    return out

```

**Fig. 4.9 Res Path**

Figures 4.8 & 4.9 depict the code implementation for MultiRes block and ResPath respectively as described previously. ReLU is used as the activation function here in the top layers except the final output layer as ReLU handles the problem of vanishing gradients and trains the model quicker.

The structure of MultiRes UNet is defined in fig. 4.10. We see there are four levels each in the encoder and decoder part. The number of filters in MultiRes block starts from 32 and increases gradually by multiples of two in the following layers. At each level of the multires block, three convolution operations of strides  $3 \times 3$ ,  $5 \times 5$ , and  $7 \times 7$  are performed followed by max pooling. In the decoder path, transposed convolution takes place at each of the levels. As mentioned in the architecture, the skip connections start with 64 filters at the first level and the number of filters increases by a factor of two. In the final layer, the original dimensions of images ( $256 \times 256$ ) are restored.

```

def MultiResUnetBP(height, width, n_channels):

    inputs = Input((height, width, n_channels))

    mresblock1 = MultiResBlock(32, inputs)
    pool1 = MaxPooling2D(pool_size=(2, 2))(mresblock1)
    mresblock1 = ResPath(32*2, 4, mresblock1)

    mresblock2 = MultiResBlock(32*2, pool1)
    pool2 = MaxPooling2D(pool_size=(2, 2))(mresblock2)
    mresblock2 = ResPath(32*4, 3, mresblock2)

    mresblock3 = MultiResBlock(32*4, pool2)
    pool3 = MaxPooling2D(pool_size=(2, 2))(mresblock3)
    mresblock3 = ResPath(32*8, 2, mresblock3)

    mresblock4 = MultiResBlock(32*8, pool3)
    pool4 = MaxPooling2D(pool_size=(2, 2))(mresblock4)
    mresblock4 = ResPath(32*16, 1, mresblock4)

    mresblock5 = MultiResBlock(32*16, pool4)

    up6 = concatenate([Conv2DTranspose(
        32*8, (2, 2), strides=(2, 2), padding='same')(mresblock5), mresblock4], axis=3)
    mresblock6 = MultiResBlock(32*8, up6)

    up7 = concatenate([Conv2DTranspose(
        32*4, (2, 2), strides=(2, 2), padding='same')(mresblock6), mresblock3], axis=3)
    mresblock7 = MultiResBlock(32*4, up7)

    up8 = concatenate([Conv2DTranspose(
        32*2, (2, 2), strides=(2, 2), padding='same')(mresblock7), mresblock2], axis=3)
    mresblock8 = MultiResBlock(32*2, up8)

    up9 = concatenate([Conv2DTranspose(32, (2, 2), strides=(
        2, 2), padding='same')(mresblock8), mresblock1], axis=3)
    mresblock9 = MultiResBlock(32, up9)

    conv10 = conv2d_bn(mresblock9, 1, 1, 1, activation='sigmoid')

    MultiResModel = Model(inputs=[inputs], outputs=[conv10])

    return MultiResModel

```

**Fig. 4.10 Defining the MultiRes model**

```
MultiResModel = MultiResUnetBP(height=256, width=256, n_channels=3)
MultiResModel.summary()
```

Model: "model"

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 256, 256, 3)]	0	[]
conv2d_1 (Conv2D)	(None, 256, 256, 8)	216	['input_1[0][0]']
batch_normalization_1 (Batch Normalization)	(None, 256, 256, 8)	24	['conv2d_1[0][0]']
activation_1 (Activation)	(None, 256, 256, 8)	0	['batch_normalization_1[0][0]']
conv2d_2 (Conv2D)	(None, 256, 256, 17)	1224	['activation_1[0][0]']
batch_normalization_2 (Batch Normalization)	(None, 256, 256, 17)	51	['conv2d_2[0][0]']
activation_1 (Activation)	(None, 256, 256, 17)	0	['batch_normalization_2[0][0]']
conv2d_3 (Conv2D)	(None, 256, 256, 26)	3978	['activation_1[0][0]']
batch_normalization_3 (Batch Normalization)	(None, 256, 256, 26)	78	['conv2d_3[0][0]']
activation_2 (Activation)	(None, 256, 256, 26)	0	['batch_normalization_3[0][0]']
concatenate_12 (Concatenate)	(None, 256, 256, 51)	0	['activation_52[0][0]', 'activation_53[0][0]', 'activation_54[0][0]']
batch_normalization_78 (Batch Normalization)	(None, 256, 256, 51)	153	['concatenate_12[0][0]']
batch_normalization_82 (Batch Normalization)	(None, 256, 256, 51)	204	['batch_normalization_78[0][0]', 'batch_normalization_82[0][0]']
add_18 (Add)	(None, 256, 256, 51)	0	['batch_normalization_78[0][0]', 'batch_normalization_82[0][0]']
activation_55 (Activation)	(None, 256, 256, 51)	0	['add_18[0][0]']
batch_normalization_83 (Batch Normalization)	(None, 256, 256, 51)	204	['activation_55[0][0]']
conv2d_56 (Conv2D)	(None, 256, 256, 1)	51	['batch_normalization_83[0][0]']
batch_normalization_84 (Batch Normalization)	(None, 256, 256, 1)	3	['conv2d_56[0][0]']
activation_56 (Activation)	(None, 256, 256, 1)	0	['batch_normalization_84[0][0]']
Total params: 9,906,494			
Trainable params: 9,876,980			
Non-trainable params: 29,514			

**Fig. 4.11 Brief Model Summary**



```

history = MultiResModel.Fit(x = train_X_scaler, y = train_Y_scaler, validation_data = (val_X, val_Y), batch_size = 8, epochs = 100, verbose = 1)

Epoch 1/100
104/104 [=====] - 213s 1s/step - loss: 0.5083 - IOU: 0.4230 - mcc: 0.5080 - dice_coef: 0.5787 - accuracy: 0.8161 - val_loss: 0.6078 - val_IOU: 0.4110 - val_e
Epoch 2/100
104/104 [=====] - 198s 1s/step - loss: 0.5220 - IOU: 0.5455 - mcc: 0.6563 - dice_coef: 0.6811 - accuracy: 0.8911 - val_loss: 0.5029 - val_IOU: 0.6348 - val_e
Epoch 3/100
104/104 [=====] - 198s 1s/step - loss: 0.5020 - IOU: 0.5982 - mcc: 0.7089 - dice_coef: 0.7368 - accuracy: 0.9308 - val_loss: 0.4051 - val_IOU: 0.7028 - val_e
Epoch 4/100
104/104 [=====] - 198s 1s/step - loss: 0.4911 - IOU: 0.6193 - mcc: 0.7208 - dice_coef: 0.7525 - accuracy: 0.9191 - val_loss: 0.4045 - val_IOU: 0.7509 - val_e
Epoch 5/100
104/104 [=====] - 198s 1s/step - loss: 0.4790 - IOU: 0.6557 - mcc: 0.7580 - dice_coef: 0.7817 - accuracy: 0.9304 - val_loss: 0.4379 - val_IOU: 0.7676 - val_e
Epoch 6/100
104/104 [=====] - 198s 1s/step - loss: 0.4683 - IOU: 0.6775 - mcc: 0.7889 - dice_coef: 0.7886 - accuracy: 0.9354 - val_loss: 0.4303 - val_IOU: 0.7525 - val_e
Epoch 7/100
104/104 [=====] - 198s 1s/step - loss: 0.4587 - IOU: 0.6978 - mcc: 0.7888 - dice_coef: 0.8141 - accuracy: 0.9415 - val_loss: 0.4349 - val_IOU: 0.7657 - val_e
Epoch 8/100
104/104 [=====] - 198s 1s/step - loss: 0.4509 - IOU: 0.7122 - mcc: 0.8018 - dice_coef: 0.8135 - accuracy: 0.9464 - val_loss: 0.4483 - val_IOU: 0.8089 - val_e
Epoch 9/100
104/104 [=====] - 198s 1s/step - loss: 0.4423 - IOU: 0.7289 - mcc: 0.8155 - dice_coef: 0.8372 - accuracy: 0.9508 - val_loss: 0.4317 - val_IOU: 0.8451 - val_e
Epoch 10/100
104/104 [=====] - 198s 1s/step - loss: 0.4333 - IOU: 0.7522 - mcc: 0.8332 - dice_coef: 0.8532 - accuracy: 0.9557 - val_loss: 0.3982 - val_IOU: 0.8618 - val_e

Epoch 90/100
104/104 [=====] - 198s 1s/step - loss: 0.1817 - IOU: 0.9413 - mcc: 0.9645 - dice_coef: 0.9690 - accuracy: 0.9876 - val_loss: 0.2053 - val_IOU: 0.9451 - val_e
Epoch 91/100
104/104 [=====] - 198s 1s/step - loss: 0.1818 - IOU: 0.9381 - mcc: 0.9618 - dice_coef: 0.9679 - accuracy: 0.9865 - val_loss: 0.2046 - val_IOU: 0.9486 - val_e
Epoch 92/100
104/104 [=====] - 198s 1s/step - loss: 0.1760 - IOU: 0.9429 - mcc: 0.9657 - dice_coef: 0.9708 - accuracy: 0.9882 - val_loss: 0.2090 - val_IOU: 0.9437 - val_e
Epoch 93/100
104/104 [=====] - 198s 1s/step - loss: 0.1780 - IOU: 0.9418 - mcc: 0.9638 - dice_coef: 0.9688 - accuracy: 0.9869 - val_loss: 0.2066 - val_IOU: 0.9526 - val_e
Epoch 94/100
104/104 [=====] - 198s 1s/step - loss: 0.1788 - IOU: 0.9377 - mcc: 0.9708 - dice_coef: 0.9643 - accuracy: 0.9857 - val_loss: 0.2088 - val_IOU: 0.9473 - val_e
Epoch 95/100
104/104 [=====] - 198s 1s/step - loss: 0.1800 - IOU: 0.9288 - mcc: 0.9516 - dice_coef: 0.9579 - accuracy: 0.9866 - val_loss: 0.2081 - val_IOU: 0.9407 - val_e
Epoch 96/100
104/104 [=====] - 198s 1s/step - loss: 0.1778 - IOU: 0.9331 - mcc: 0.9598 - dice_coef: 0.9681 - accuracy: 0.9858 - val_loss: 0.2075 - val_IOU: 0.9388 - val_e
Epoch 97/100
104/104 [=====] - 198s 1s/step - loss: 0.1787 - IOU: 0.9317 - mcc: 0.9681 - dice_coef: 0.9688 - accuracy: 0.9863 - val_loss: 0.2062 - val_IOU: 0.9398 - val_e
Epoch 98/100
104/104 [=====] - 198s 1s/step - loss: 0.1758 - IOU: 0.9311 - mcc: 0.9578 - dice_coef: 0.9610 - accuracy: 0.9866 - val_loss: 0.2064 - val_IOU: 0.9353 - val_e
Epoch 99/100
104/104 [=====] - 198s 1s/step - loss: 0.1750 - IOU: 0.9374 - mcc: 0.9601 - dice_coef: 0.9688 - accuracy: 0.9857 - val_loss: 0.2068 - val_IOU: 0.9484 - val_e
Epoch 100/100
104/104 [=====] - 198s 1s/step - loss: 0.1734 - IOU: 0.9351 - mcc: 0.9587 - dice_coef: 0.9625 - accuracy: 0.9851 - val_loss: 0.2033 - val_IOU: 0.9525 - val_e

```

**Fig. 4.12 Training the model for 100 epochs**

Fig. 4.11 provides a brief model summary of the MultiRes UNet model. The images have been resized to  $256 \times 256$  dimensions as mentioned in 4.2.3. After tweaking out the hyperparameters, we found out the following values to give the best results for our model. The model training as given in fig. 4.12 started with an initial IoU value of 42% and gradually increased to 75% in the next 10 epochs. From the above training, we see that there is no generalization gap between validation IoU and training IoU, and hence the model is not overfitting. In a similar way, the loss value has reduced significantly from 0.5220 in the first epoch to 0.1734 in the final epoch. We also notice MCC and Dice Coefficient to follow similar trends.



**Table 4.1 Hyperparameter values used in training our MultiRes UNet model**

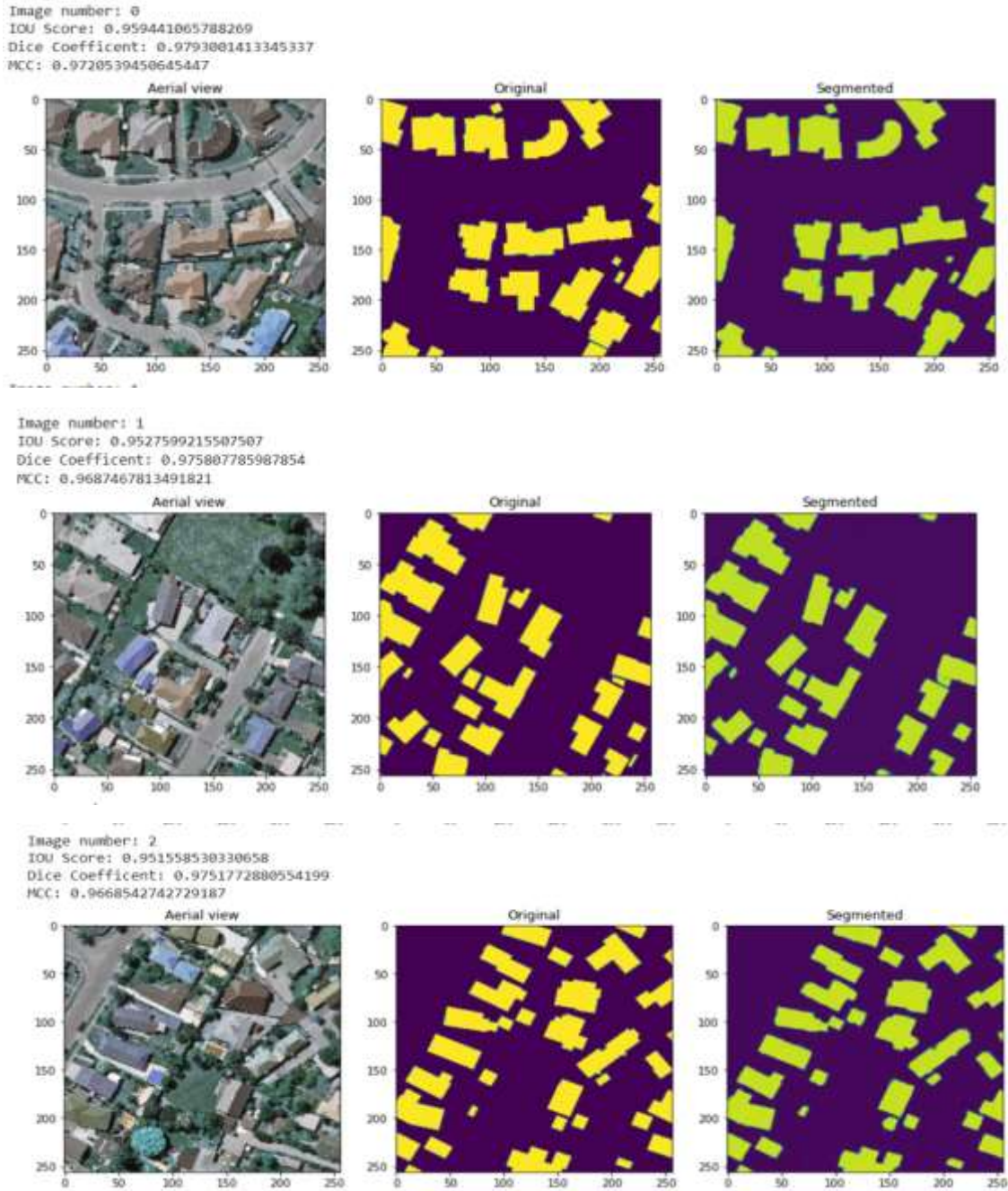
Hyper parameters	Values
Learning rate	0.0001
Epochs	100
Batch size	8
Image dimensions	$256 \times 256$
Optimizer	Adam
Loss	Binary cross entropy
Activation Function	ReLU (in all convolution layers) Sigmoid (in the output layer)

$$\text{Loss} = -\frac{1}{\text{output size}} \sum_{i=1}^{\text{output size}} y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log (1 - \hat{y}_i) \quad (4.1)$$

### Binary cross entropy loss

The performance graphs of the MultiRes model on building segmentation is described in the following section along with various performance metrics on training and validation set. Fig. 4.13 compares ground truth images with our segmented images. The results show that on an average an IoU score of 95% is obtained and we notice that even small buildings are correctly segmented by the trained model. Dice Coefficient/ F1 score hovers

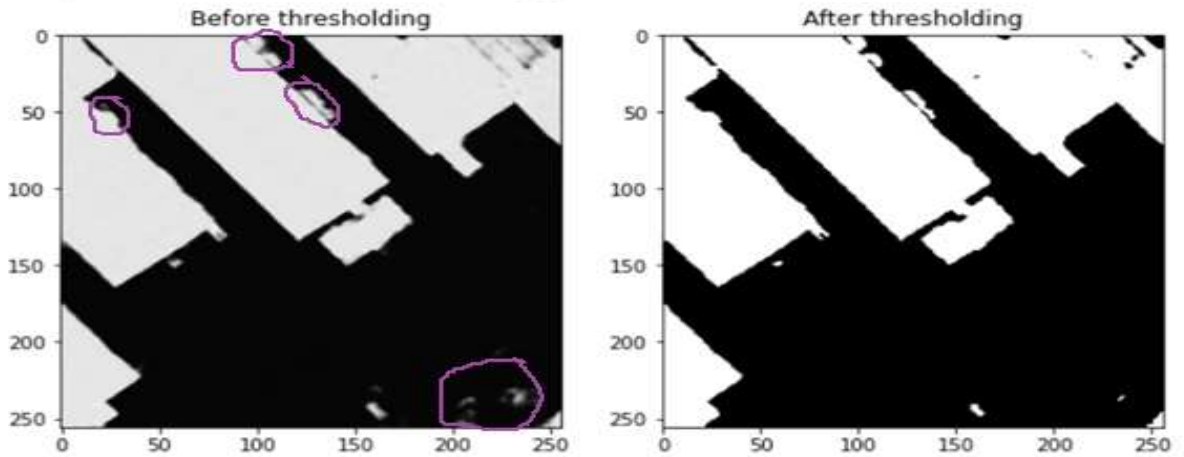
around 97.5% for the following figure. From fig. 4.13, we also notice that the corners and different layouts of multiple buildings are evenly segmented out.



**Fig. 4.13 Results of building segmentation with MultiRes UNet model**

#### 4.2.5 Applying threshold

Thresholding is a method where pixels are partitioned based on their intensity value. Here, pixel value 0 indicates black color while pixel value 1 indicates white color. After training the images with the MultiRes UNet model, some pixel values lie in the range between (0, 1) and these pixels do not clearly indicate whether they belong to foreground or background. In order to resolve this issue and delineate the boundaries of buildings properly, we use simple thresholding. The threshold value is set to 0.5, and all values greater than 0.5 are assigned to the foreground, while values less than 0.5 are assigned to the background. In fig. 4.14, the violet color indications show the difference in segmentation before and after the threshold is applied. After applying thresholding, the images are clearly delineated. This aids in the separation of two structures that are quite close to one another.



**Fig. 4.14 Comparison of results before and after applying threshold**

Fig. 4.15 provides a comparison between the ground truth images with our prediction masks before and after applying threshold. After thresholding, our buildings are more effectively demarcated.

Image number: 0  
 IOU Score: 0.959441065788269  
 Dice Coefficient: 0.9793001413345337  
 MCC: 0.9720540046691895

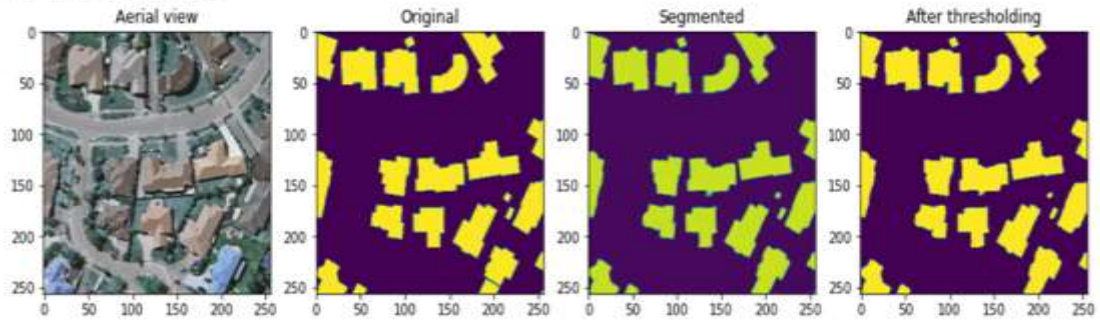


Image number: 1  
 IOU Score: 0.9527599215507507  
 Dice Coefficient: 0.975807785987854  
 MCC: 0.9687466621398926

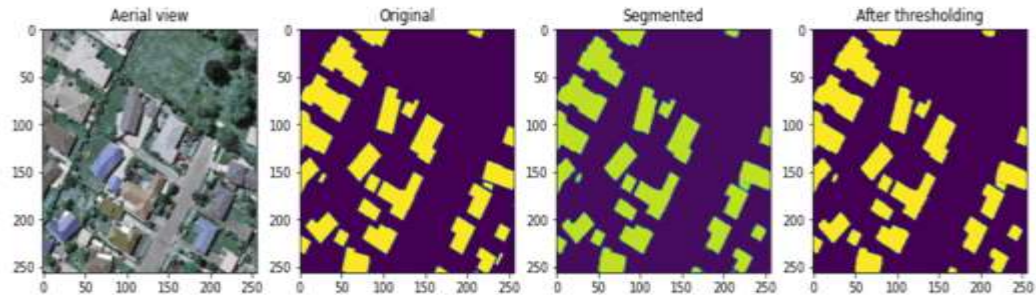
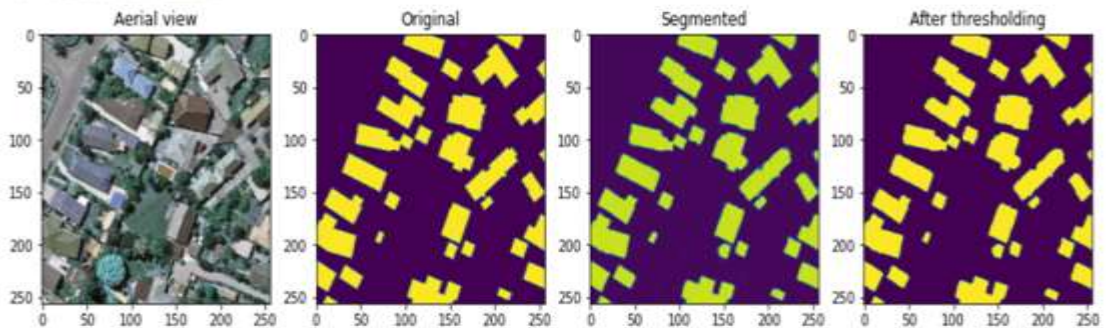


Image number: 2  
 IOU Score: 0.951558530330658  
 Dice Coefficient: 0.9751772880554199  
 MCC: 0.9668542742729187



**Fig. 4.15 Comparison of ground truth images with our predictions before and after applying threshold**



## 4.2.6 Rooftop Extraction

### 4.2.6.1 Get contours

Contours are curves that connect all continuous points of the same hue or intensity. They are used to detect the borders of objects and localize them easily in an image. The segmented building masks are first transformed to grayscale and then binary thresholding is applied on it following which the contours of buildings are identified. For each image, the number of buildings in the given aerial satellite image is identified by this method.

In fig. 4.16, findContours function is used to draw contours around the detected buildings. The mode used here is cv2.RETR\_EXTERNAL which retrieves external or outer contours of the building. The algorithm used for contour detection is CHAIN\_APPROX\_SIMPLE that removes all redundant points and compresses the contour, thereby saving memory. The result is the detected contours where each contour is stored as a vector of points.

#### Draw bounding boxes on the binary segmented mask & perform color filling

- Convert the image to grayscale.
- Apply threshold.
- Get contours.
- Fill bounding box with color
- Store {x,y,w,h} in a vector of list

```
# read image
img_mask = cv2.imread(train_path_label)
print(type(img))
# convert to grayscale
gray = cv2.cvtColor(img_mask,cv2.COLOR_BGR2GRAY)

# threshold
thresh = cv2.threshold(gray,128,255,cv2.THRESH_BINARY)[1]

# get contours
result = img_mask.copy()
seg_with_bounded_box = img_mask.copy()
contours = cv2.findContours(thresh, cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)
contours = contours[0] if len(contours) == 2 else contours[1]
print("No. of identified buildings: {}".format(len(contours)))
bb = []
```

**Fig. 4.16** Code implementation of drawing contours

#### 4.2.6.2 Draw bounding boxes

Bounding boxes are used to highlight the regions of interest in the image. The detected contour is enclosed within a rectangle using the `boundingRect` function. The top-left (x, y) coordinates of the bounding box are found along with the width and height of each box. Each of the bounding box coordinates are stored in a vector.

For the purpose of drawing rectangle around the segmented masks, the `cv2.boundingRect()` function is used. This function is used mainly to highlight the region of interest after obtaining contours from an image.

```
# draw bounding boxes
for cnt in contours:
    x,y,w,h = cv2.boundingRect(cnt)
    cv2.rectangle(result, (x, y), (x+w, y+h), (0, 0, 255), 2)
    cv2.rectangle(seg_with_bounded_box, (x, y), (x+w, y+h), (255, 255, 255), -1)
    # print("x,y,w,h:",x,y,w,h)
    bb.append([x,y,w,h])
print(bb)
```

```
<class 'numpy.ndarray'>
No of identified buildings: 36
[[4398, 9761, 57, 48], [2593, 6474, 179, 134], [2794, 6393, 39, 47], [2518, 6162, 300, 227], [4495, 5982, 48, 49], [4548, 5860, 140, 123], [3544, 5015, 81, 73],
```



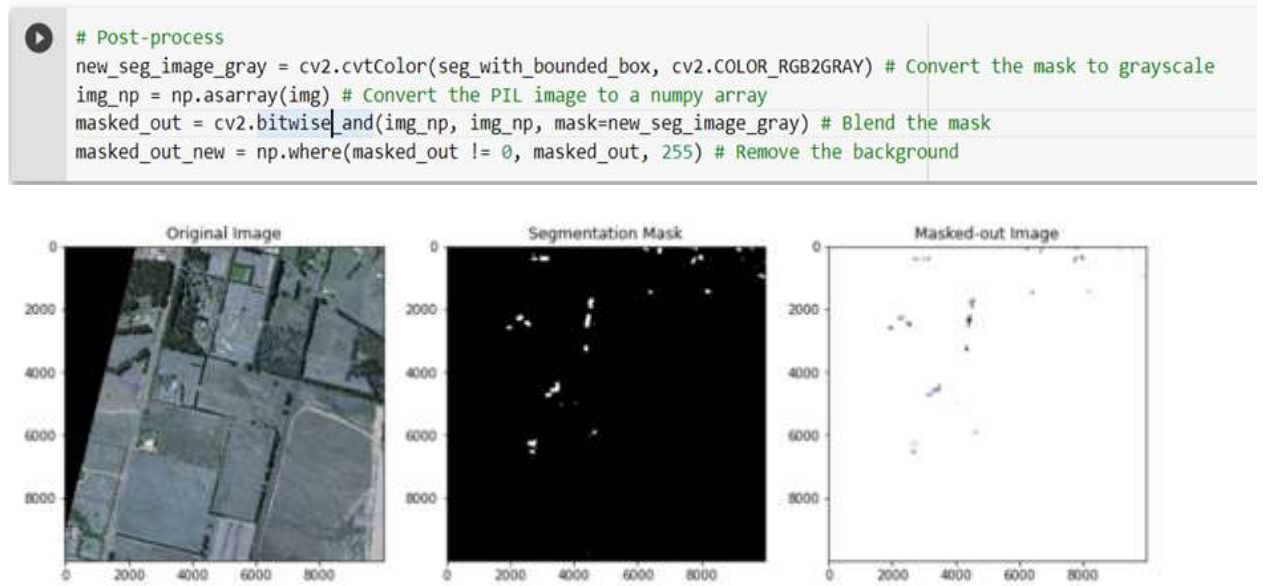
**Fig. 4.17** Bounding boxes drawn around the segmented mask images

After performing building detection, we draw bounding boxes as mentioned above and fig. 4.17 is an illustration where rectangular boxes are drawn after finding the contours.

#### 4.2.6.3 Background Subtraction

Background subtraction refers to a technique where an image's foreground is extracted for further processing like object recognition, classification. In our case, the foreground is the detected buildings with contours. The aerial satellite image is subtracted from the bounding box drawn image and rooftops are extracted.

##### Background Subtraction



**Fig. 4.18 Original satellite image (left), masked image after training on MultiRes UNet model (center), extracted rooftops (right)**

The rooftops are extracted by the above process as shown in fig. 4.18 and are stored in a database which is later used in the second stage for classification.

### 4.3 ROOFTYPE CLASSIFICATION AND BOUNDARY DETECTION

#### 4.3.1 Preparation of dataset and labeling roof type

The second phase in the pipeline is classifying the type of roofs and drawing boundaries on the rooftop. After conducting rooftop extraction in the preceding step, a dataset including 1115 rooftop photos is populated into three different categories: Flat, Gable, and Hip. After reading numerous publications, we discovered that the authors used manual roof type labeling in every case. Hence, we resorted to the same technique of manual labeling of roof tops as there was no supervised dataset for roof type classification available. The task of manual labeling does not require any prior knowledge of the field and roof types in most cases can be easily identified by laymen. Thus, we referred to the sample images provided by M. Buyukdemircioglu et al. for labeling our dataset.

The dataset was cleaned up by removing roof types with poor resolution. A balance between the several roof type classifications was sought as much as feasible when compiling the dataset. Despite this, the number of rooftops of type gable is considerably higher than the other two classes due to the varying numbers of instances for each roof type in the research area. The distribution of different roof types are presented in table 4.2.

**Table 4.2 Distribution of roof type dataset**

<b>Roof Type</b>	<b>Training (91%)</b>	<b>Validation (5%)</b>	<b>Testing (4%)</b>	<b>Total</b>
Flat	303	14	18	335
Gable	410	23	18	451
Hip	307	13	9	329

**Table 4.3 Samples of images of each class from Christchurch, New Zealand**












<b>FLAT</b>			
<b>GABLE</b>			
<b>HIP</b>			

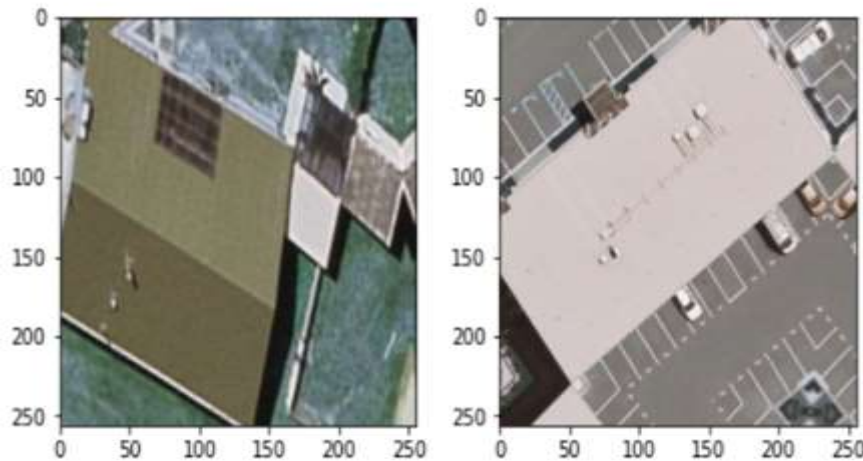
Table 4.3 gives a sample representation of the different roof tops belonging to each of the 3 classes. In order to improve the performance of model training, rooftops of different layouts and orientations were carefully picked, analyzed and labeled.

### ***4.3.2 Data Augmentation***

For training neural network models, the size of the dataset needs to be large. However, owing to time constraints we were able to manually label only 1115 images. As a result, data augmentation approaches were employed to increase the dataset size for better model training. Rotation, shifting and flipping are the three main techniques used here.

- i. Rotation - This is used to randomly rotate images in the range present. We set the rotation range to 7 degrees here.
- ii. Flipping - In flipping, images are randomly flipped in both horizontal and vertical direction.
- iii. Shifting - Images are shifted randomly based on the percentage mentioned in horizontal and vertical direction.

Fig. 4.19 shows two data augmentation techniques applied to roof top images. Horizontal flipping is applied to the one on the left while width shifting is used for the image on the right.



**Fig. 4.19 Results of image after performing data augmentation**

### 4.3.3 Roof type classification

Following data augmentation, the model must be trained to recognize different types of roofs. CNN model along with additional transfer learning models like VGG16, ResNet50, and EfficientNetB4 are used for training. Following this, majority voting is applied as the ensembling approach to combine learnings from different models for a better result.

#### 4.3.3.1 Shallow CNN Model

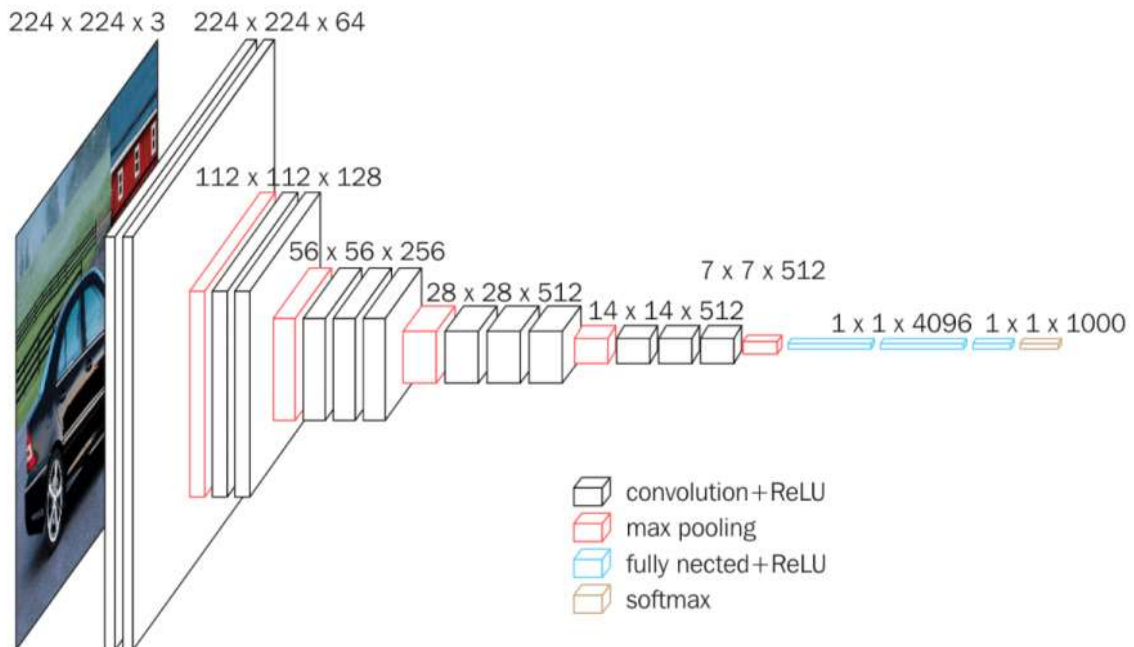
The shallow CNN model used here has 6 convolution layers and 6,751,809 trainable parameters. For the convolutional layers, a 3x3 kernel filter size was selected (Conv2D). In the pooling layers, the default pool size (2x2) was employed (MaxPooling2D). A dropout of 0.2 is used to prevent the model from overfitting. ReLU is used for the activation function with categorical cross entropy as the loss function. The detailed model summary is displayed in fig. 4.20. Both RMSProp and Adam optimizers were tried for training the model and results indicate RMSProp as a better choice as it gave better results. With a learning rate of 0.0001 and a batch size of 16, the model was trained for 100 epochs.

Model: "sequential"			conv2d_5 (Conv2D)	(None, 64, 64, 64)	36928
Layer (type)	Output Shape	Param #	activation_5 (Activation)	(None, 64, 64, 64)	0
conv2d_0 (Conv2D)	(None, 256, 256, 64)	1792	max_pooling2d_2 (MaxPooling2D)	(None, 30, 30, 64)	0
activation_0 (Activation)	(None, 256, 256, 64)	0	dropout_2 (Dropout)	(None, 30, 30, 64)	0
conv2d_1 (Conv2D)	(None, 256, 256, 128)	63470	conv2d_6 (Conv2D)	(None, 30, 30, 32)	18464
activation_1 (Activation)	(None, 256, 256, 128)	0	activation_6 (Activation)	(None, 30, 30, 32)	0
max_pooling2d (MaxPooling2D)	(None, 127, 127, 128)	0	conv2d_7 (Conv2D)	(None, 32, 32, 32)	9344
dropout (Dropout)	(None, 127, 127, 128)	0	activation_7 (Activation)	(None, 32, 32, 32)	0
conv2d_2 (Conv2D)	(None, 127, 127, 64)	83244	max_pooling2d_3 (MaxPooling2D)	(None, 14, 14, 32)	0
activation_2 (Activation)	(None, 127, 127, 64)	0	dropout_3 (Dropout)	(None, 14, 14, 32)	0
conv2d_3 (Conv2D)	(None, 128, 128, 64)	63568	flatten (Flatten)	(None, 872)	0
activation_3 (Activation)	(None, 128, 128, 64)	0	dense (Dense)	(None, 1024)	442368
max_pooling2d_1 (MaxPooling2D)	(None, 62, 62, 64)	0	activation_8 (Activation)	(None, 1024)	0
dropout_1 (Dropout)	(None, 62, 62, 64)	0	dense_1 (Dense)	(None, 3)	3078
conv2d_4 (Conv2D)	(None, 62, 62, 64)	83440	activation_9 (Activation)	(None, 3)	0
activation_4 (Activation)	(None, 62, 62, 64)	0	Total params: 6,751,809		
			Trainable params: 6,751,809		
			Non-trainable params: 0		

**Fig. 4.20 Model summary of CNN model**

### 4.3.3.2 VGG16

VGG16 is one of the most famous transfer learning models trained on the ImageNet database. The architecture of VGG goes as follows: A stack of multiple convolution layers of filter size  $3 \times 3$ , stride one, and padding 1, followed by a max-pooling layer of size  $2 \times 2$ , is the basic building block for all of these configurations. There are a total of 13 convolutional layers and 3 fully connected layers in VGG16 architecture. Image is passed through the first stack of 2 convolution layers of the very small receptive size of  $3 \times 3$ , followed by ReLU activations. Each of these two layers contains 64 filters. The activations then flow through a similar second stack, but with 128 filters as against 64 in the first one. This is followed by the third stack with three convolutional layers of kernel size 256 and a max pool layer. This is followed by two stacks of three convolutional layers, with each containing 512 filters. Fig. 4.21 gives an insight on the architecture of the VGG16 model.



**Fig. 4.21 Architecture of VGG16 model**

In this project, the fully connected layers of the three networks are replaced with the fully connected layer block with 64, 128, 256, 512 nodes. The last layer has three nodes representing the three classes for our classification. The model is trained for 50 epochs with a learning rate of 0.0001 and batch size of 4. RMSProp Optimizer is used here.

#### 4.3.3.3 RESNET50

The architecture of ResNet50 as shown in fig. 4.22 contains a convolution with a kernel size of 7 x 7 and 64 different kernels all with a stride of size 2. This is followed by max pooling with a stride size of 2. In the next convolution there is a 1 x 1, 64 kernel following this a 3 x 3, 64 kernel and at last a 1 x 1, 256 kernel. These three layers are repeated a total of three times, giving us nine layers in this step.

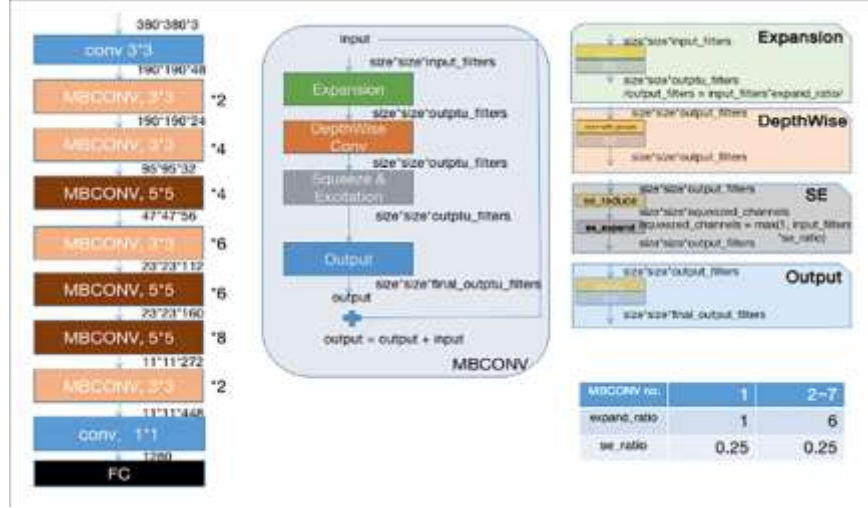
layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		$1.8 \times 10^9$	$3.6 \times 10^9$	$3.8 \times 10^9$	$7.6 \times 10^9$	$11.3 \times 10^9$

**Fig. 4.22 Architecture of ResNet50 model**

The model is trained for 50 epochs with a learning rate of 0.0001 and batch size of 8 and RMSProp optimizer is used.

#### 4.3.3.4 EFFICIENTNETB4

EfficientNetB4 is yet another best pre-trained model. There are five modules in total in EfficientNetB4. Module 3 is used as a skip connection to all the sub-blocks and module 4 for combining the skip connections in the first sub-blocks. The architecture of EfficientNetB4 is shown in fig. 4.23.



**Fig. 4.23 Architecture of EfficientNetB4 model**

All the layers after top\_bn are set to trainable. There are totally 125,207 non-trainable parameters. The model is trained for 50 epochs with a learning rate of 0.0001 and batch size of 8. EfficientNetB4 works well with Adam Optimizer and hence is used here.

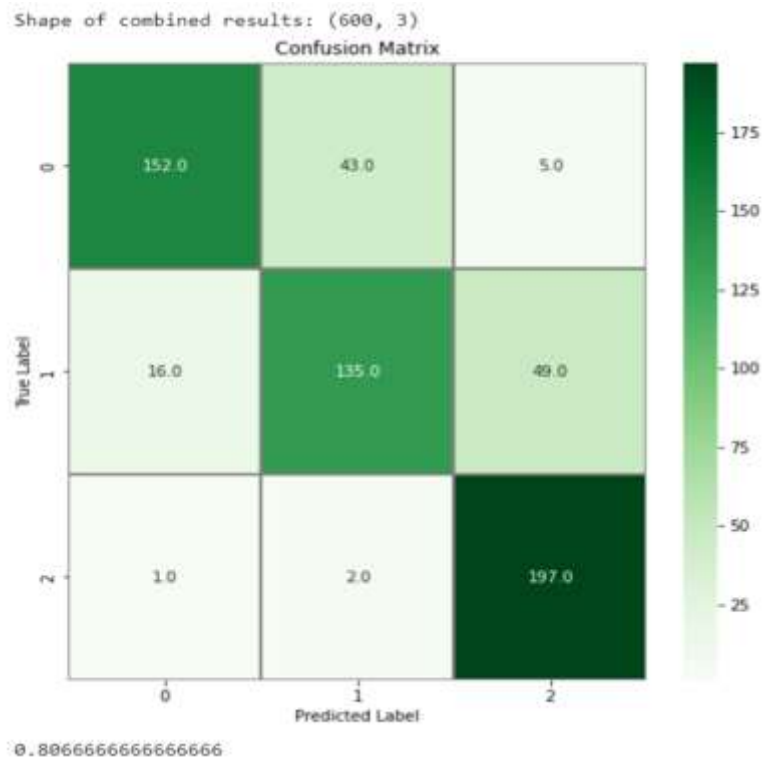
The performance results of different deep learning models and a comparative analysis between the transfer learning models are presented in section 5.5.2 along with the validation on Potsdam dataset.

#### 4.3.4 Majority Voting

From the performance results in section 5.5.2, it is inferred that transfer learning methods result in higher classification accuracy when compared to the shallow CNN model. However in these transfer learning models, it can be

noticed that each model performs classification well on only some parts. Hence an ensembling approach is used as it can make better predictions and achieve better performance than any single contributing model. Ensembling also reduces the spread or dispersion of the predictions and model performance. The ensembling method applied here is majority voting.

Fig. 4.24 reveals that VGG16 was efficient in identifying hip roof types. Similarly, ResNet very well classified flat roofs from other rooftop types and EfficientNet was able to classify a large chunk of gable roof types correctly. Majority voting is thus incorporated to further boost the performance. The accuracy has increased by 5.67% and the overall classification accuracy is now 80.67%.



**Fig. 4.24 Results on Potsdam dataset after applying majority voting**

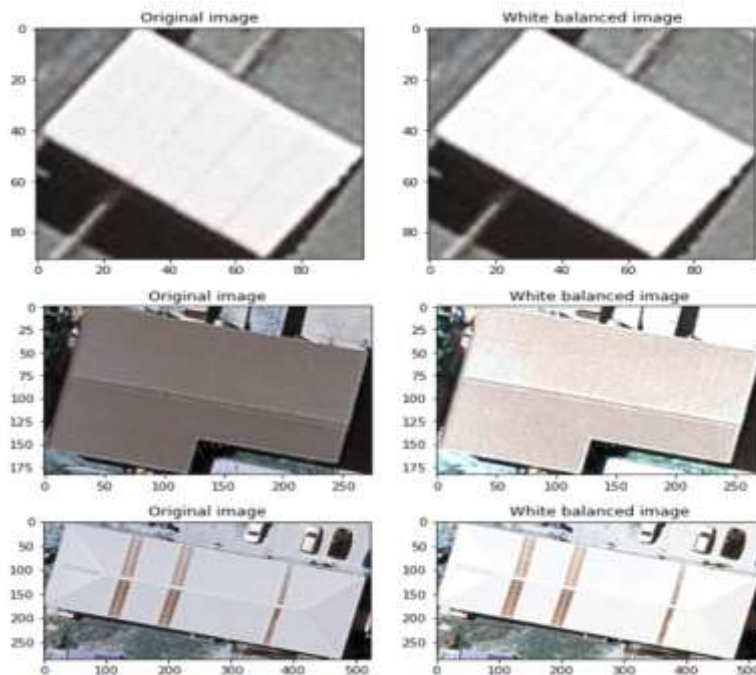


### ***4.3.5 Boundary detection***

#### ***4.3.5.1 White Balancing***

White balancing is the initial step in boundary detection of roof tops. This step must be completed first as it is color dependent. White balancing is used to remove any colored haze from an image so that it seems to be under white light. Edge recognition methods can be hampered by haze on an image since it reduces local contrast, which can lead to edges being ignored or incorrectly recognized. This is a major issue with satellite images as they are taken with varying incident angles for the light, or may be exacerbated by weather and atmospheric variances.

Due to the highly variable lighting conditions under which satellite imagery is gathered, this is an extremely impactful step on the resulting quality of edge detection. Fig 4.25 shows the rooftop images before and after white balancing. It can be seen that images appear to be under white light after this step is performed.



**Fig. 4.25 Images after performing white balancing to remove haze**



#### ***4.3.5.2 Gaussian Blur***

Gaussian blurring is a smoothing technique that reduces various types of noise or accentuates edge contrast within the image. Gaussian blur does not introduce ringing into the filtered image.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (4.2)$$

#### **Gaussian Blur**

The formula in equn. 4.2 is used to construct a convolution matrix, which is then applied to the original image. Here, a kernel size of 3×3 is used with  $\sigma = 1.5$ .

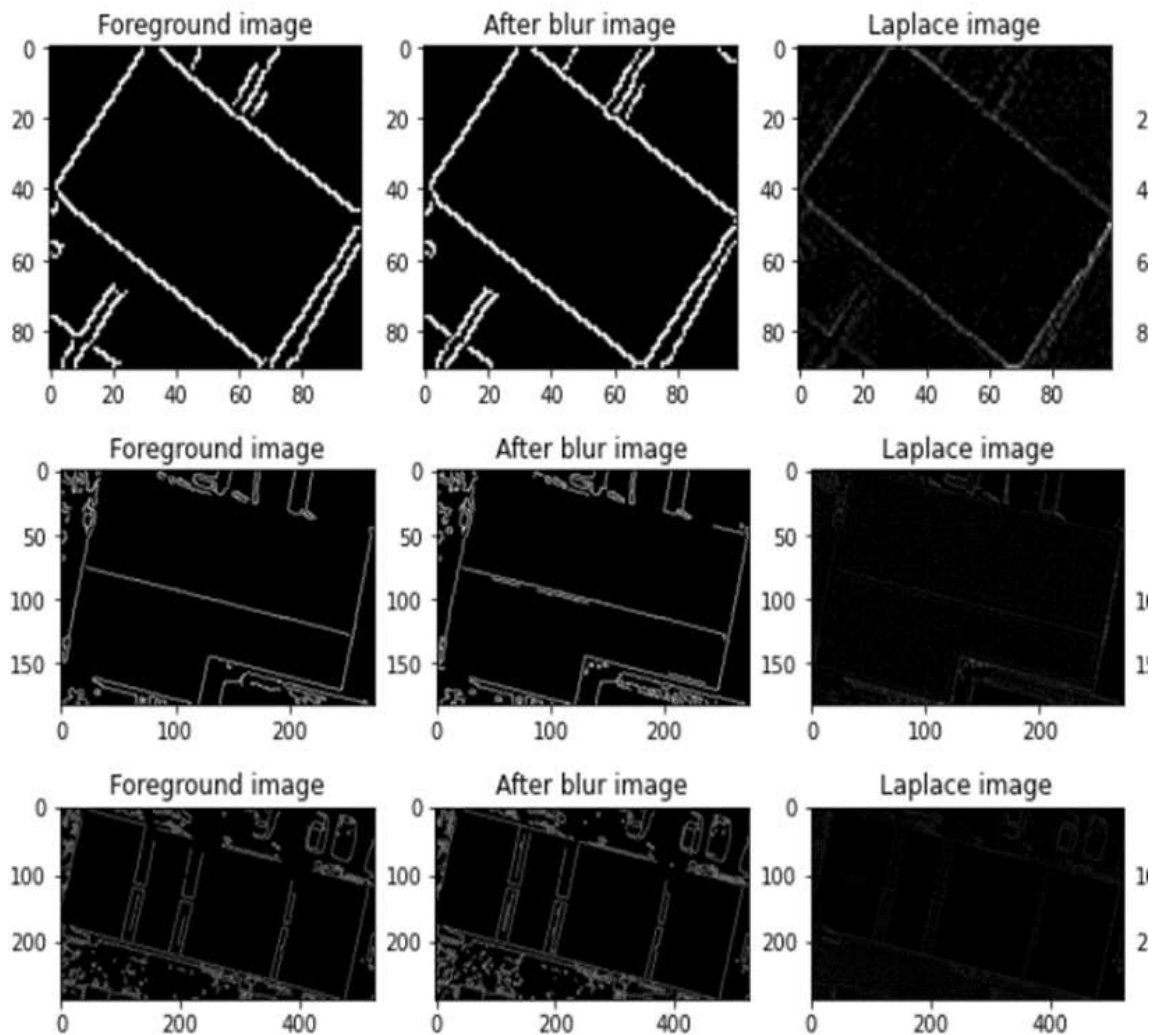
The new value for each pixel is a weighted average of the pixels in its immediate vicinity. The value of the original pixel gets the most weight (since it has the highest Gaussian value), while nearby pixels get lesser weights as their distance from the original pixel grows. This results in a blur that preserves boundaries and edges better than other, more uniform blurring filters.

#### ***4.3.5.3 Auto Canny Edge Detection***

Canny edge detection algorithm is a popular edge detection algorithm that is multi-staged involving noise reduction, finding intensity gradients and hysteresis suppression. After white patching and converting the image to grayscale, Gaussian blur is applied. This is followed by auto canny edge detection. The median of single channel pixel intensity is calculated. The

upper and lower threshold is calculated by using mean and variance of an image intensity.

Laplacian operator is a derivative operator which is used to find edges in an image. It is a second order derivative mask. Here, the Laplacian operator is also applied on the gray image and then edges are detected. From fig 4.26, we can infer that applying auto canny edge detection after Gaussian blur provides better results comparatively.



**Fig. 4.26 Boundaries detected on rooftops by applying different techniques**

## 4.4 PV MODULE FITTING

The first stage in PV module installation is to identify the contours and detect obstacles on the rooftop using the edge sharpened image acquired by auto canny edge technique in the previous step. We use the `contourArea` function in openCV to compute the area of each contour drawn. Since most of the obstacles on top of roof top occupy a smaller area, all contours whose area is less than 5, are marked as obstacles or patches having small areas where panels cannot be fitted and these points are not considered for panel placement.

This is followed by obtaining the dimensions of solar panels including length, width and tilt angle from the user. The tilt angle for gable and hip roofs is typically kept between 0° and 10°. For rooftops of flat type, the preferred tilt angle lies around 20°- 45°.

### 4.4.1 *Rotation of panels*

PV panels that have zero tilt angle can be mounted directly on rooftops and moved. However, for panels that have non- zero tilt angle, a projection of panels needs to be made. Let us assume the initial coordinates of PV panels to be (x, y). The rotated coordinates (x', y') are determined depending on the tilt angle using equn. 4.3. Fig. 4.27 shows the implementation of the rotation method.

$$\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix} \quad (4.3)$$

**Rotating points based on rotation matrix**

```

▶ # Rotate coordinates based on rotation matrix
def rotation(center_x, center_y, points, ang):
    angle = ang * math.pi / 180 # Convert to radians
    rotated_points = []
    for p in points:
        x, y = p
        x, y = x - center_x, y - center_y
        x, y = (x * math.cos(angle) - y * math.sin(angle), x * math.sin(angle) + y * math.cos(angle))
        x, y = x + center_x, y + center_y
        rotated_points.append((x, y))
    return rotated_points

```

**Fig. 4.27 Code implementation of rotation of points**

#### ***4.4.2 PV Panel Placement***

The rotated points are now checked to see if they lie within the boundary of rooftops. For each of the rotated solar PV panels, pointpolygontest is undertaken to ensure that all points of solar PV panels lie inside the rooftop. If the rotated panel coordinates lie outside, then the panels are not placed. Otherwise, the PV panels are placed inside the roof tops.

This is then followed by using Bresenham's algorithm to identify points in straight line segments so that PV panels can be placed one after the other along the line segment. This procedure can be used to remove smaller patches of area and other obstructions. The number of points between two end points are determined by the above step and a threshold value of 10 is set, over which indicates the presence of a large number of intermediary points, which in most circumstances could constitute impediments. Fig. 4.28 gives the code implementation of PV panel placement as described here.

```

# Check for if rotated points go outside of the image
if (rotated_patch_points > 0).all():
    solar_polygon = Polygon(rotated_patch_points)
    polygon_points = np.array(solar_polygon.exterior.coords, np.int32)

# Appending points of the image inside the solar area to check the intensity
patch_intensity_check = []

# Point polygon test for each rotated solar patch area
for j in range(rows):
    for k in range(cols):
        if cv2.pointPolygonTest(polygon_points, (k, j), False) == 1:
            patch_intensity_check.append(high_reso[j, k])

print("Rotated patch points type: {0}, {1}, {2}, {3}, {4}, {5}".format(type(rotated_patch_points), rotated_patch_points[0], rotated_patch_points[1],
rotated_patch_points[2], rotated_patch_points[3], rotated_patch_points[4], rotated_patch_points[5]))

# Check for the region available for Solar Panels
if np.mean(patch_intensity_check) == 255:

    # Moving along the length of line to segment solar panels in the patch
    solar_line_1 = createLineIterator(rotated_patch_points[0], rotated_patch_points[1], high_reso)
    solar_line_1 = solar_line_1.astype(int)
    solar_line_2 = createLineIterator(rotated_patch_points[3], rotated_patch_points[2], high_reso)
    solar_line_2 = solar_line_2.astype(int)
    line1_points = []
    line2_points = []

    print("Solar line 1: {0}, {1}\n Solar line 2: {2}".format(solar_line_1, len(solar_line_2), solar_line_2))

```

```

if len(solar_line_2) > 10 and len(solar_line_1) > 10:

    # Remove small unwanted patches
    cv2.fillPoly(high_reso, [rotated_patch_points], 0)
    cv2.fillPoly(high_reso_new, [rotated_patch_points], 0)
    cv2.polylines(high_reso_orig, [rotated_patch_points], 1, 0, 2)
    cv2.polylines(high_reso_new, [rotated_patch_points], 1, 0, 2)

    cv2.fillPoly(high_reso_orig, [rotated_patch_points], (0, 0, 255))
    cv2.fillPoly(high_reso_new, [rotated_patch_points], (0, 0, 255))

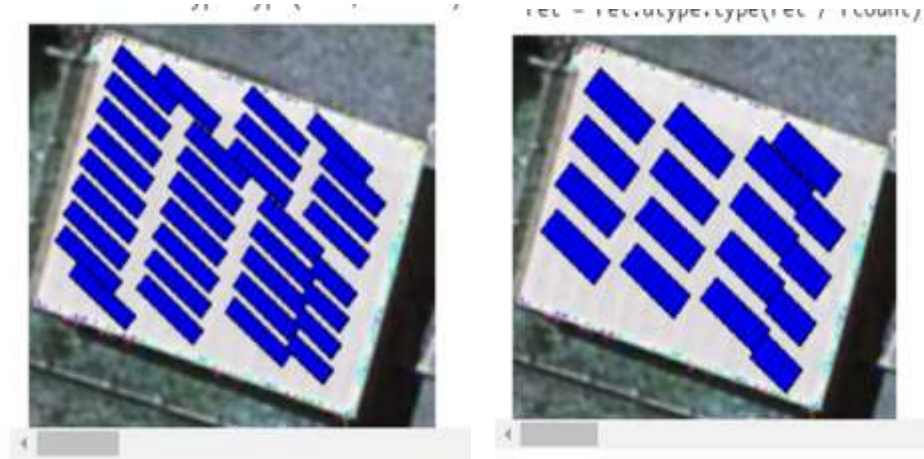
    for i in range(5, len(solar_line_1), 5):
        line1_points.append(solar_line_1[i])
    for i in range(5, len(solar_line_2), 5):
        line2_points.append(solar_line_2[i])

# Segmenting Solar Panels in the Solar Patch
for points1, points2 in zip(line1_points, line2_points):
    x1, y1, _ = points1
    x2, y2, _ = points2
    cv2.line(high_reso_orig, (x1, y1), (x2, y2), (0, 0, 0), 1)
    cv2.line(high_reso_new, (x1, y1), (x2, y2), (0, 0, 0), 1)

```

**Fig. 4.28 Code implementation for region availability of solar panels**

A simulation of PV panel arrangement on flat roof tops is shown in fig. 4.29. It can be observed from fig. 4.29 that the number of solar panels and their position/alignment change as the width of PV panels change.



**Fig. 4.29 PV panels with dimensions length - 20mm; tilt angle - 40°; width - 10 mm (left) and length - 20mm; tilt angle - 40°; width -20mm (right).**

The effectiveness of our algorithm in placing PV panels on rooftops when there are obstacles can be noticed from fig. 4.30. In this case, the tree occupies a large portion of the rooftop and becomes the obstacle. The PV panels are oriented in such a way to ensure there is maximum energy consumption and does not occupy the place where there are obstacles.



**Fig. 4.30 Placement of PV panels when there are obstacles.**

## 4.5 PERFORMANCE METRICS

### 4.5.1 For building detection segmentation:

#### 4.5.1.1 IoU - Intersection over Union /Jaccard Coefficient

To quantify the accuracy of our model for building detection, we use Jaccard coefficient which is to measure the similarity between detected regions and ground truth regions. Jaccard Similarity Index (JSI) measures the similarity for the two sets of pixel data, with a range from 0% to 100%. The higher the percentage, the more precise the prediction is. It is defined as follows:

$$JSI = \frac{r_d \cap r_g}{r_d \cup r_g} \quad (4.4)$$

#### Formula for IoU

where  $r_d$  denotes the predicted masked region for building detection, and  $r_g$  indicates the groundtruth region for building segmentation.

#### 4.5.1.2 DICE Coefficient

DICE coefficient is used to compare the pixel-wise agreement between a predicted segmentation and its corresponding ground truth. DICE coefficient is twice the area of overlap divided by the total number of pixels in both the images. The formula is given by:

$$\frac{2 * |X \cap Y|}{|X| + |Y|} \quad (4.5)$$

#### Formula for Dice coefficient

where X is the predicted set of pixels and Y is the ground truth.

#### **4.5.1.3 MCC - Matthews Correlation Coefficient**

MCC is a standard measure of a binary classifier's performance, where values are in the range  $-1.0$  to  $1.0$ , with  $1.0$  being perfect building segmentation,  $0.0$  being random building segmentation, and  $-1.0$  indicating building segmentation is always wrong. The expression for computing MCC is below, where TP is the fraction of true positives, FP is the fraction of false positives, TN is the fraction of true negatives, and FN is the fraction of false negatives, such that  $TP+FP+TN+FN= 1$ .

$$\frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (4.6)$$

**Formula for MCC**

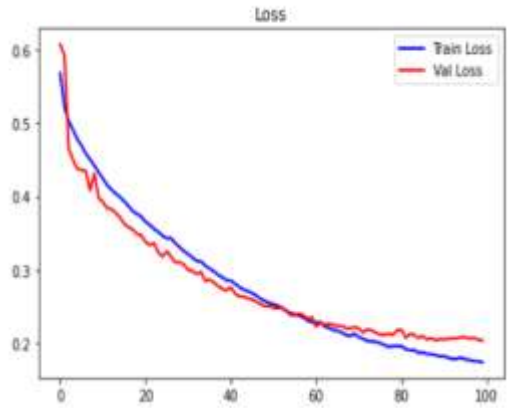
#### **4.5.1.4 Accuracy**

Accuracy is the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.

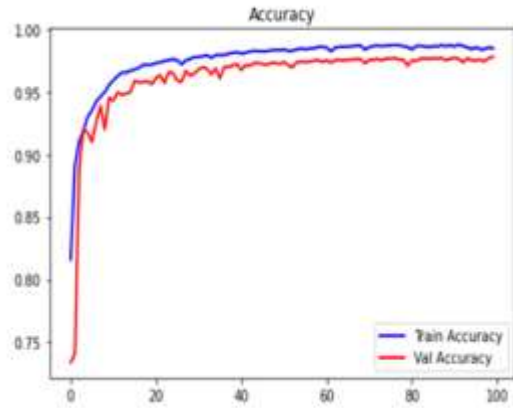
$$\text{accuracy} = \frac{\text{correct predictions}}{\text{all predictions}} \quad (4.7)$$

**Formula for accuracy**

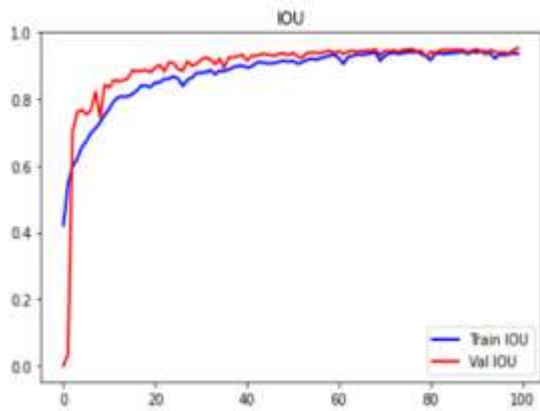




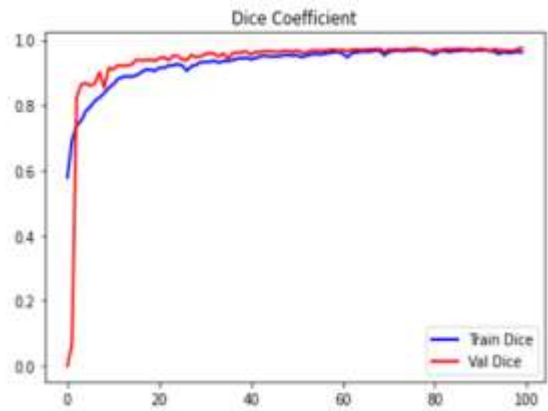
(a) Model Loss



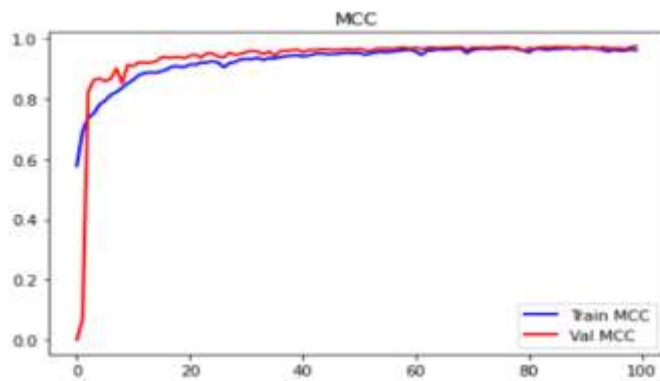
(b) Model Accuracy



(c) Model IOU



(d) Model Dice Coefficient



(e) Model MCC

**Fig. 4.31 Graphs of various performance metrics for building segmentation**

**Table 4.4 Performance results of the model on training & validation set**

<b>Performance Metrics</b>	<b>Training Set</b>	<b>Validation Set</b>
IOU (%)	93.53	95.25
Dice Coefficient (%)	96.25	97.56
MCC (%)	95.87	96.74
Accuracy (%)	98.51	97.83
Loss	0.1734	0.2033

The graphs in fig. 4.31 allows us to easily interpret the training of the MultiRes UNet model. The blue line indicates the training details while the red line shows the validation details. The MCC value is close to 98% at the end of 100 epochs which indicates that buildings have been identified correctly. IoU, which penalizes heavily than dice coefficient, has achieved a remarkable value of 96.27% at the end of 100 epochs. It can be seen that the Dice coefficient starts with 58% at the beginning of training and gradually attains a value of 98.5% with no major fluctuations. As can be seen from fig.4.31, the training and validation curve coincides with each other and the difference in loss is very meager which shows the model does not overfit.

Furthermore, from table 4.4, we infer that the model has performed remarkably well as the training loss hovers around 0.173 and validation loss is close to 0.2. This is a reasonably good value which indicates that our model performs well in segmenting the buildings accurately. Average IoU after training 1548 images for 100 epochs comes to 93.53% and the validation IoU of 95.25% demonstrates that the model performs well even after the images are resized and normalized.

#### **4.5.2 For roof type classification:**

##### ***4.5.2.1 Classification Report***

The classification report is used to measure the quality of predictions from a classification algorithm. Precision, Recall and F1 scores are calculated on a per-class basis based on True Positives, True Negatives, False Positives, and False Negatives. The above values are calculated for each of the classes, namely: Flat, Gable, and Hip.

##### ***4.5.2.2 AUC-ROC***

The Receiver Operator Characteristic (ROC) is a probability curve that plots the TPR (True Positive Rate) against the FPR (False Positive Rate) at various threshold values and separates the ‘signal’ from the ‘noise’. The Area under the Curve (AUC) is the measure of the ability of a classifier to distinguish between classes. Higher the values of AUC-ROC, better is the performance of the classification algorithm.

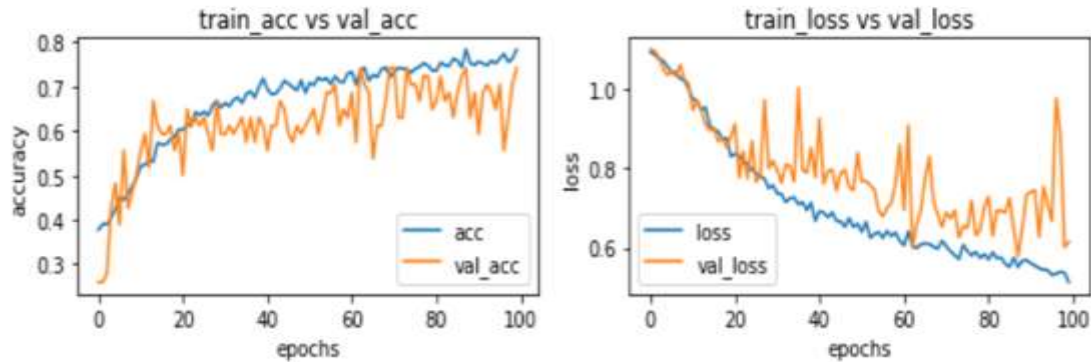
#### ***4.5.3 Performance results on roof type classification***

##### ***4.5.3.1 Performance of Shallow CNN Model***

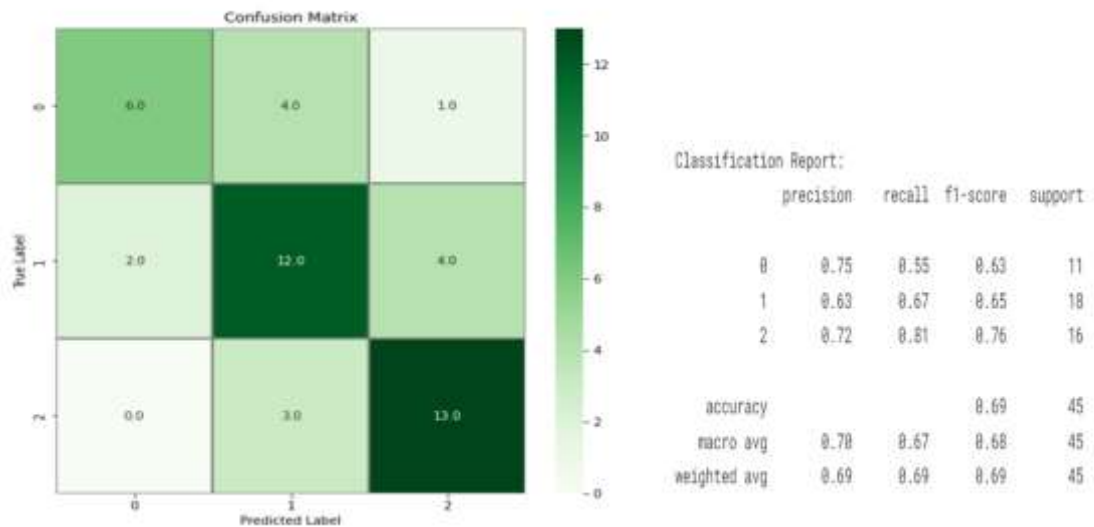
The CNN model achieved an accuracy of 78% on the training set and 74% on the validation set. From fig. 4.32, it can be inferred that there is no generalization gap because of dropout and data augmentation. The model was likewise trained for 50 epochs at first, but because it didn't produce satisfactory results, the model was further trained for another 50 epochs.

The confusion matrix in fig. 4.33 reveals that gable and hip classes are accurately classified to a larger extent than other classes. An equal number of flat classes, on the other hand, are misclassified as gables. The results also show that hip class has the highest recall value of 81% followed by gable

(67%) and then flat (55%). The overall accuracy on the test samples accounts to 69%.



**Fig. 4.32** Graphs showing training & validation accuracy, loss for CNN

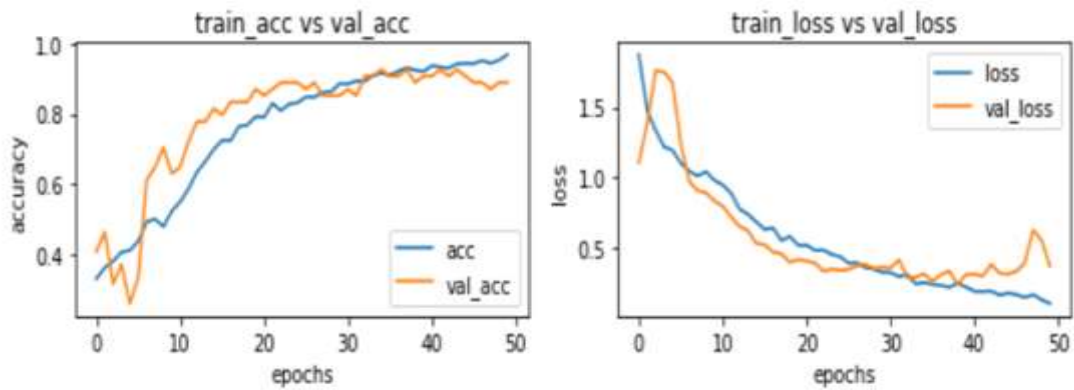


**Fig. 4.33** Confusion matrix & classification report on test data for CNN

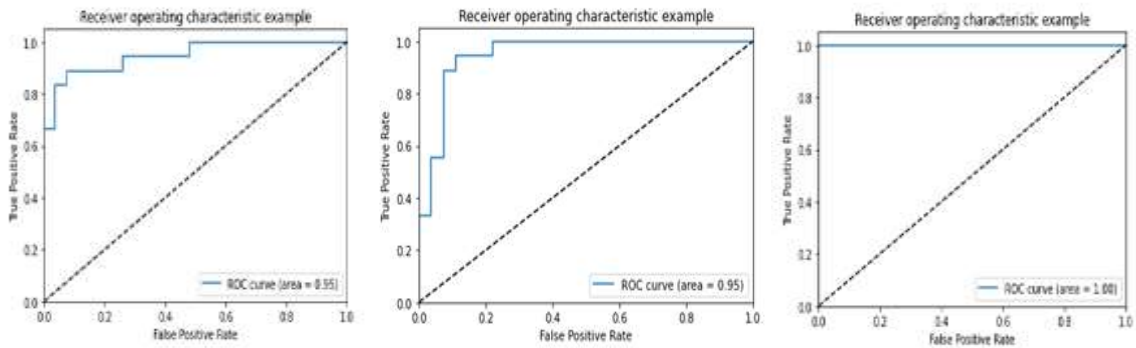
#### 4.5.3.2 Performance of Fine-tuned ResNet-50 model

The ResNet-50 has yielded a very good accuracy of 91.59% on the training set and 92.5% on the validation set. From fig. 4.34, it can be witnessed that the loss is also very less at only 0.245 on the training set. All transfer learning models are only trained for 50 epochs as training for 100 epochs led to overfitting.

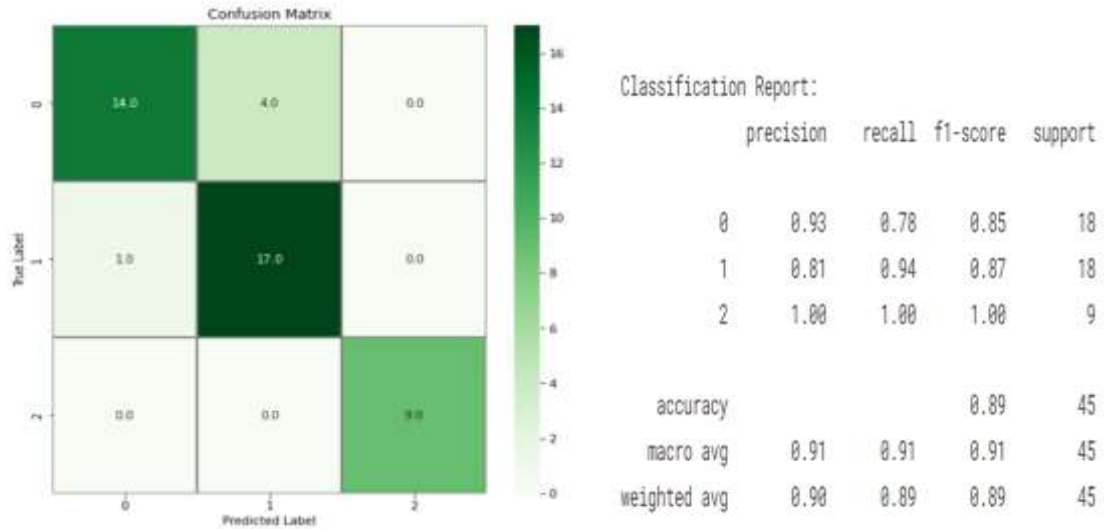
From the confusion matrix in fig. 4.36, it can be inferred that the hip class is 100 % accurately classified and nearly all gable classes are correctly identified. The classification of flat classes is better compared to the previous CNN model and the misclassification of a few flat classes as gable could be due to the slight tilt angle of a few rooftop slopes, making it difficult for the model to classify it to the appropriate class. The overall accuracy on the test samples is 89% with F-1 score averaging to 91%. The higher the AUC, the better the performance of the model at distinguishing between the positive and negative classes. The ROC curve (fig. 4.35) has a value of 1 for hip class as all rooftops are correctly classified.



**Fig. 4.34** Graphs showing training & val accuracy, loss for ResNet50



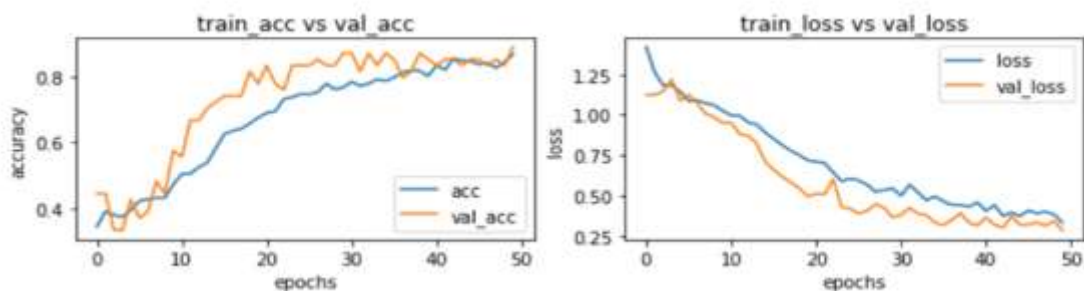
**Fig. 4.35** AUC-ROC metrics for ResNet-50 model



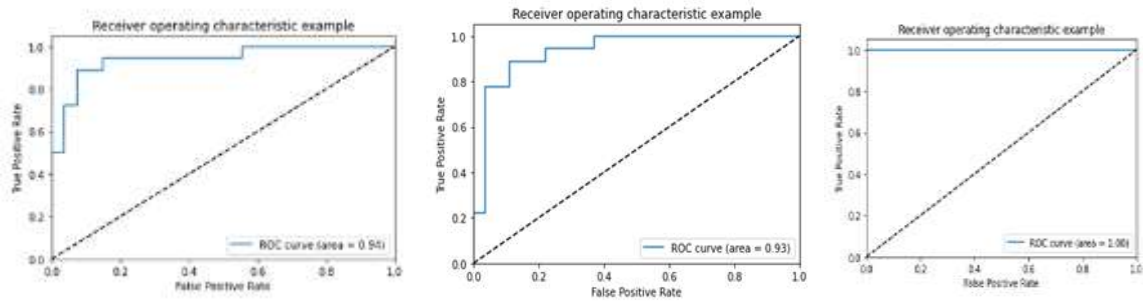
**Fig. 4.36 Confusion matrix & classification report on test data for ResNet-50**

#### 4.5.3.3 Performance of Fine-tuned EfficientNetB4 model

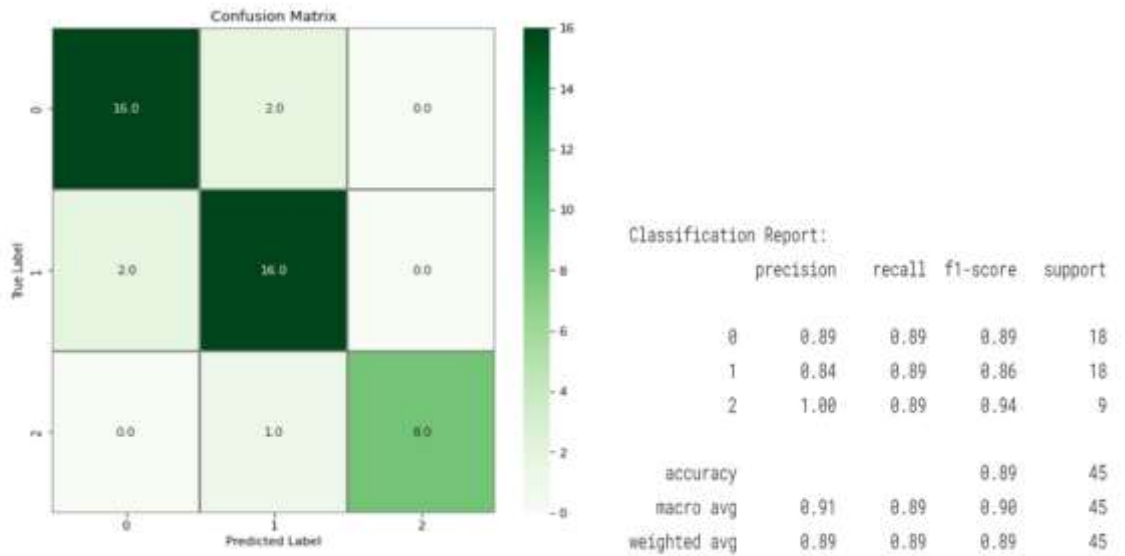
The graphs in fig. 4.37 exhibits that the accuracy of the model has steadily grown from 40% to 89% and the model loss has significantly reduced from 1.25 to 0.33 at the end of 50 epochs. When compared with the ResNet-50 model which had higher accuracy in correctly classifying hip class, EfficientNetB4 model has achieved higher F-1 score and accuracy in identifying flat roof tops (fig. 4.39). The average F-1 score, precision, and recall hover around 91%. The AUC-ROC score for flat and gable classes is 0.93 and is a perfect 1.0 for hip as seen in fig. 4.38.



**Fig. 4.37 Graphs showing training & validation accuracy, loss for EfficientNet**



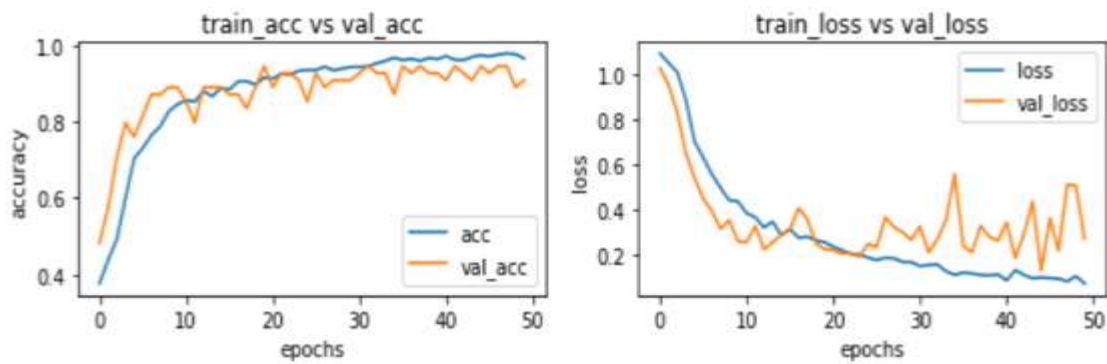
**Fig. 4.38 AUC-ROC metrics for EfficientNetB4 model**



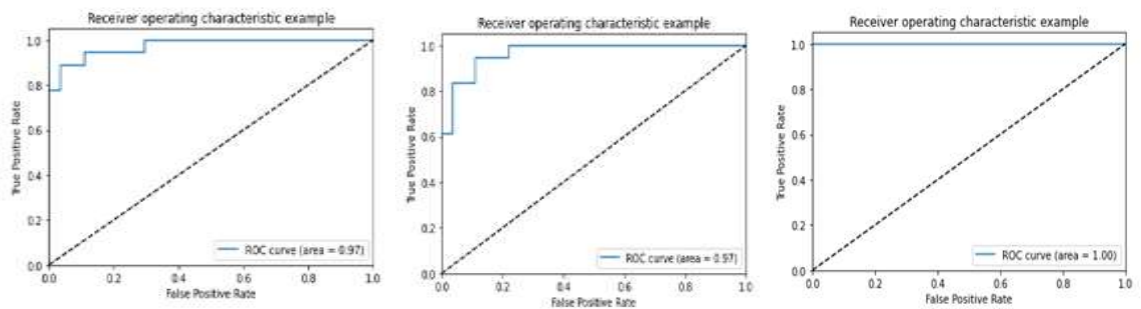
**Fig. 4.39 Confusion matrix & classification report on test data for EfficientNet**

#### 4.5.3.4 Performance of Fine-tuned VGG-16 model

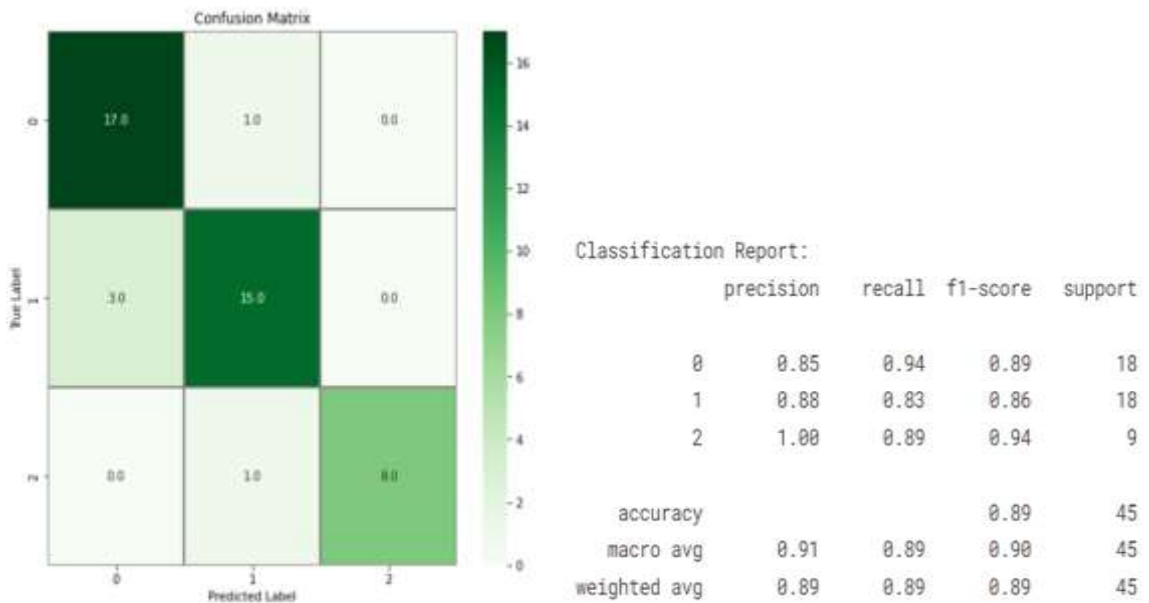
VGG-16 model has achieved the highest accuracy compared to the other three models. The classification accuracy on the validation set is 94.45% while the training accuracy is around 97%. In fig. 4.40, it can be noticed there is a gradual decline in training loss over 50 epochs while there are some fluctuations with validation loss. From classification report in fig. 4.42, it is observed that the F-1 score for flat class is the highest with VGG16 compared to other models. The overall classification accuracy has totalled to 89%.



**Fig. 4.40** Graphs showing training & val accuracy, loss for VGG16



**Fig. 4.41** AUC-ROC metrics for VGG-16 model



**Fig. 4.42** Confusion matrix & classification report on test data for VGG16



**Table 4.5 Results of different models on our manually labeled dataset**

<b>Model</b>	<b>Learning rate &amp; Optimizer</b>	<b>Batch Size</b>	<b>Loss</b>	<b>Accuracy (%)</b>	<b>F-1 score</b>
Shallow CNN	0.0001 and RMSProp	16	0.7233-Train 0.8067-Test	69.87-Train 60 - Test	0.69
ResNet- 50	0.00001 and RMSProp	8	0.2450-Train 0.7123-Test	92.59-Train 88.89 - Test	0.91
Efficient NetB4	0.00001 and Adam	16	0.3359-Train 0.343-Test	86.70-Train 89 - Test	0.91
VGG16	0.0001 and RMSProp	4	0.2021-Train 0.4894-Test	96.50-Train 88.89 - Test	0.89

Table 4.5 provides a summary of the results obtained from various models. It is seen that the average accuracy of all the three transfer learning models comes around 91.93% when compared to 69.87% of the shallow CNN model and the loss value hovers around an average of 0.356. This brings us to the conclusion that the pre-trained model performed better in terms of prediction due to the large amount of training data in the model building phase. The shallow CNN model also exhibited high accuracy despite the small dataset size and can be improved with more data.

#### ***4.5.3.5 Validation on Potsdam dataset***

As mentioned in section 5.3, we have resorted to manual labeling of roof types. In order to validate our dataset, the trained models were tested on a labeled dataset of roof types from Potsdam in Germany. The dataset contains

2000 samples of images belonging to three classes - Flat, Gable and Hip and we tested our model on 600 images. The results are listed in table 4.6.






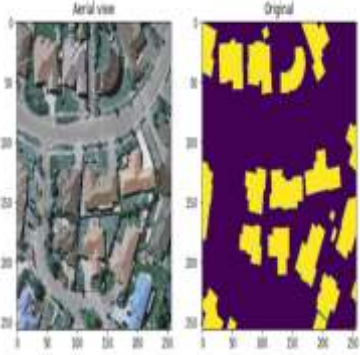
**Table 4.6 Results of testing our trained models on Potsdam dataset**

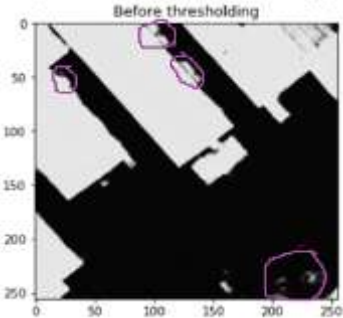
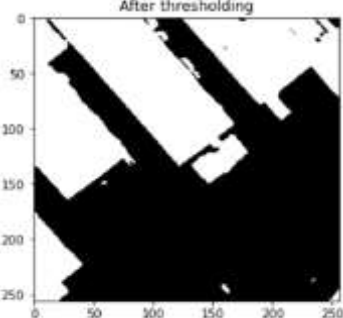
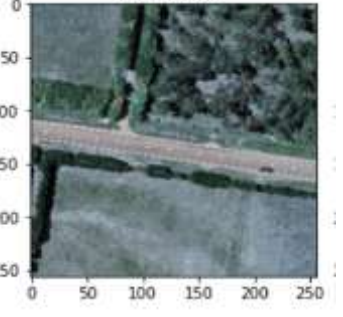
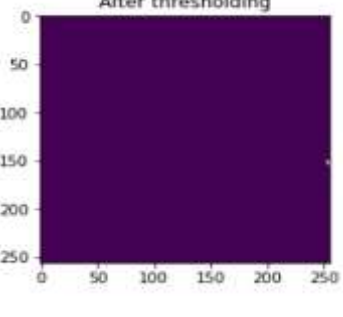
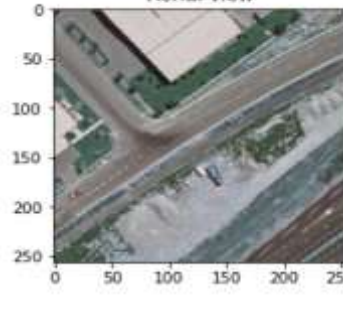
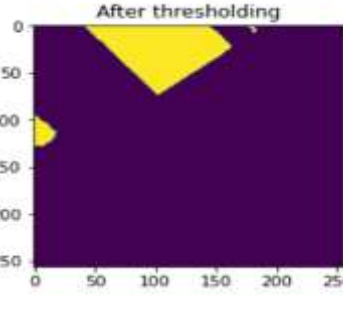
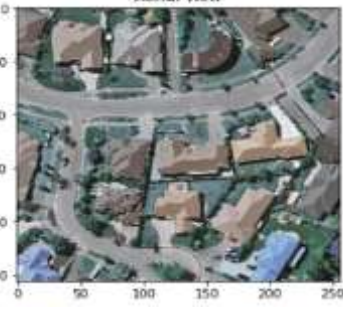
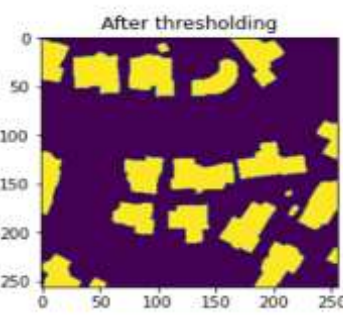
<b>Model</b>	<b>Learning rate &amp; Optimizer</b>	<b>Batch Size</b>	<b>Loss</b>	<b>Accuracy (%)</b>	<b>F-1 score</b>
Shallow CNN	0.0001 and RMSProp	16	1.2829	45.67	0.38
ResNet- 50	0.00001 and RMSProp	8	0.9248	70.67	0.71
EfficientNetB4	0.00001 and Adam	16	0.7110	65.17	0.66
VGG16	0.0001 and RMSProp	4	0.8507	74	0.73

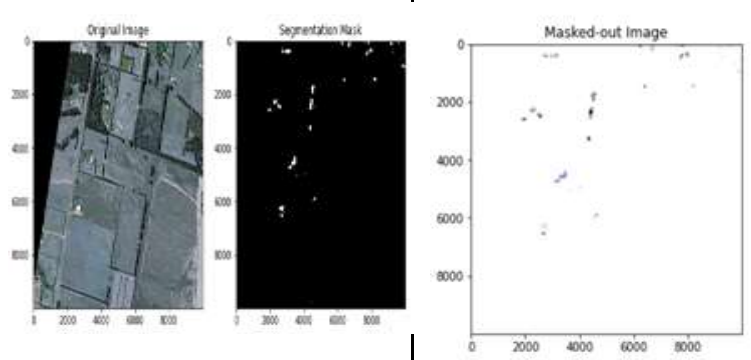
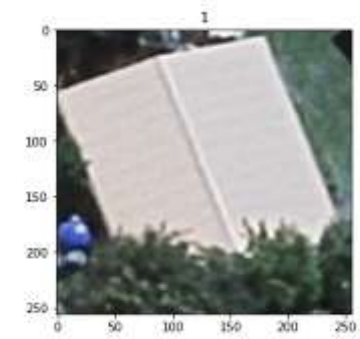
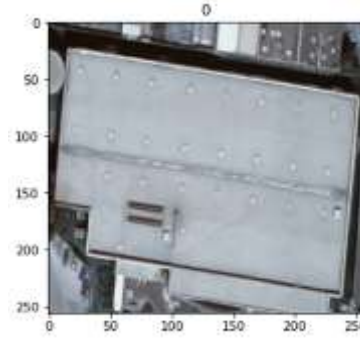
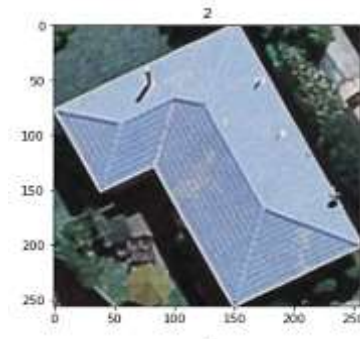
From the findings, we notice that the accuracy on the Potsdam dataset is hovering around 70% on average when compared to the results on manually labeled dataset. There are several factors which we think could be the reason for the same:



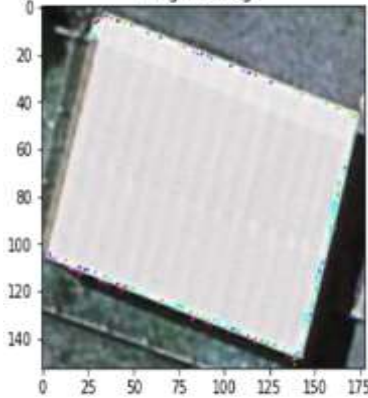
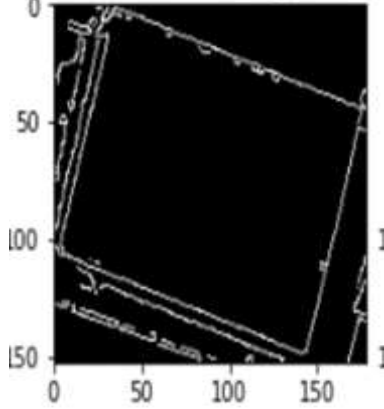
- i. CNN and other transfer learning models were trained on high resolution images whereas the images of the Potsdam dataset are not of very good resolution.
- ii. Nearly all rooftops in Potsdam are brown in color while the colors of rooftops in Christchurch are significantly bright.
- iii. With the aforementioned reasons in mind, we can see that an average of 75% accuracy is still a pretty respectable result, especially considering the model was trained on one dataset and evaluated on another.

## 4.6 TESTCASES


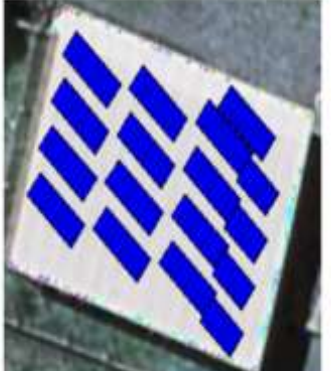
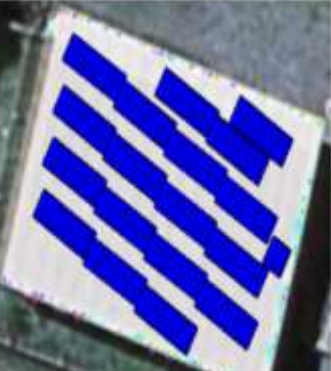
TEST CASE ID	TEST CASE DESCRIPTION	TEST INPUT	TEST OUTPUT
TC_01	Clipping of aerial images.		
TC_02	Clipping of ground truth masks.		
TC_03	Normalization		



TEST CASE ID	TEST CASE DESCRIPTION	TEST INPUT	TEST OUTPUT
TC_04	Applying Threshold	 <p>Before thresholding</p>	 <p>After thresholding</p>
TC_05	Building segmentation results- with no buildings.	 <p>Aerial view</p>	 <p>After thresholding</p>
TC_06	Building segmentation of a single building.	 <p>Aerial view</p>	 <p>After thresholding</p>
TC_07	Building segmentation of multiple buildings.	 <p>Aerial view</p>	 <p>After thresholding</p>

TC_08	Rooftop extraction by background subtraction.	 <p>The figure consists of three subplots. The first, 'Original Image', shows an aerial view of a city grid with a coordinate system from 0 to 8000 on both axes. The second, 'Segmentation Mask', shows the same area with a black background and white pixels representing rooftops, with axes from 0 to 8000. The third, 'Masked-out Image', shows the background with the rooftop areas removed, leaving white shapes on a black background, with axes from 0 to 8000.</p>	
TC_09	Classification of gable image.	 <p>An aerial photograph of a gable roof, labeled '1' at the top. The roof is light-colored and triangular. The image has a coordinate system from 0 to 250 on both axes.</p>	ResNet50 - 1 (Gable) Shallow CNN - 1 (Gable) EfficientNetB4 - 1 (Gable) VGG16 - 1 (Gable)
TC_10	Classification of flat image.	 <p>An aerial photograph of a flat roof, labeled '0' at the top. The roof is light-colored and rectangular. The image has a coordinate system from 0 to 250 on both axes.</p>	ResNet50 - 0 (Flat) Shallow CNN - 1 (Gable) EfficientNetB4 - 0 (Flat) VGG16 - 0 (Flat)
TC_11	Classification of hip image.	 <p>An aerial photograph of a hip roof, labeled '2' at the top. The roof is light-colored and has a complex, multi-faceted shape. The image has a coordinate system from 0 to 250 on both axes.</p>	ResNet50 - 2 (Hip) Shallow CNN - 2 (Hip) EfficientNetB4 - 2 (Hip) VGG16 - 2 (Hip)

TEST CASE ID	TEST CASE DESCRIPTION	TEST INPUT	TEST OUTPUT
TC_12	Majority Voting (Unequal predictions)		<p> CMN Result: 2  ResNet Result: 1  EfficientNetB4-Adam Result: 1  VGG Result: 1  [[2 1 1 1]]  [0, 3, 1] 1  Image christchurch_173_343.jpg belongs to class: 1 </p> <p>3 models predict as Gable while 1 model predicts as Flat</p>
TC_13	Majority Voting (Equal predictions)		<p> CMN Result: 2  ResNet Result: 2  EfficientNetB4-Adam Result: 2  VGG Result: 2  [[2 2 2 2]]  [0, 0, 4] 2  Image christchurch_173_395.jpg belongs to class: 2 </p> <p>All models predict as class hip</p>
TC_14	Auto Canny Edge detection	 	



TEST CASE ID	TEST CASE DESCRIPTION	TEST INPUT	TEST OUTPUT
TC_15	PV Panel placement for flat roof.	Roof type: Flat Length: 20mm Width: 10 mm Tilt Angle: 40°	
TC_16	PV Panel placement for flat roof with different width.	Roof type: Flat Length: 20mm Width: 20 mm Tilt Angle: 40°	
TC_17	PV Panel placement for flat roof with varying tilt angle.	Roof type: Flat Length: 20mm Width: 20 mm Tilt Angle: 30°	

TEST CASE ID	TEST CASE DESCRIPTION	TEST INPUT	TEST OUTPUT
TC_18	PV Panel placement for gable roof with tree as obstacle.	Roof type: Gable Length: 20mm Width: 20 mm Tilt Angle: 10°	
TC_19	PV Panel placement for flat roof with tree as obstacle.	Roof type: Flat Length: 10mm Width: 10 mm Tilt Angle: 30°	



## 4.7 COMPARATIVE ANALYSIS

Table 4.7 provides a comparative analysis of building segmentation with the MultiRes UNet baseline model and the model we have implemented after performing resizing and normalization. From the results, we can infer that despite reducing the image to lower dimensions and applying normalization, the model has outperformed the baseline model with respect to Dice Coefficient and MCC values.

**Table 4.7 Comparative analysis for building segmentation using MultiRes UNet model**

	<b>MultiRes baseline</b>	<b>MultiRes with normalization</b>
<b>IoU (%)</b>	93.14	95.25
<b>MCC (%)</b>	94.91	96.74
<b>Dice coefficient (%)</b>	96.45	97.56

Similarly, for roof type classification the results with different DL models are compared. The comparison shows that our CNN model performs a bit poorly when compared with the CNN model employed by M. Buyukdemircioglu et al. However, we can observe that ResNet50 and EfficientNetB4 have outperformed the AIRS dataset when compared to the other dataset by around 10%. We also infer that the VGG16 model performs equally well on both datasets.

**Table 4.8 Comparative analysis of different DL models on AIRS  
dataset and dataset from Ankara, Turkey**

<b>Model</b>	<b>Results on AIRS dataset</b>			<b>Results based on dataset in Turkey</b>		
	Precision	Recall	F1 score	Precision	Recall	F1 score
<b>Shallow CNN</b>	70	67.67	68	81	81.67	81
<b>ResNet50</b>	91.34	90.67	90.67	82.34	86.34	84.34
<b>EfficientNetB4</b>	91	89	89.67	81	84	82.34
<b>VGG16</b>	91	88.67	89.67	84.67	89.67	86.67

## CHAPTER 5

### CONCLUSION AND FUTURE WORK

#### 5.1 CONCLUSION

A 3-step pipeline was proposed in this research to automate the process of PV panel placement on rooftops given a satellite image covering a wide range and different types of buildings. The first phase involved building segmentation on AIRS dataset with the MultiRes UNet model that utilizes MultiRes block to adapt spatial features from various scales and Res path for transferring encoder features to decoder features using a set of convolution operations. Despite resizing the images to  $256 \times 256$  dimensions after clipping the large satellite images, the findings show that the model works extremely well, with an average IoU of 95.25% and a Dice Coefficient of 97.56%.

Proper segmentation of buildings led to proper extraction of different roof types which were manually categorized into 3 classes: Flat, Gable and Hip. Different deep learning models were trained on the 1115 images and a comparative analysis was performed. With an F1 score of 0.69, the shallow CNN model achieved a classification accuracy of 79.25 %. Fine-tuned transfer learning models, on the other hand, fared well in classifying roof types. ResNet50 and EfficientNetB4 both achieved an F1 score of 0.91 while VGG16 obtained a score of 0.89.

We discovered that different models performed well in classifying different classes of roof tops with ResNet performing well in identifying flat roofs, EfficientNetB4 in classifying gable roof tops and VGG16 with hip roofs and hence majority voting was used as an ensembling technique in combining

the predictions of different learning models. Validation was performed on the supervised Potsdam dataset, with an accuracy of 80.34% after applying majority voting. The comparatively low accuracy could be due to the fact that the training was done on high-resolution satellite photos from the AIRS dataset, whereas the images on Potsdam were somewhat low-resolution. The classification results can further be improved by using more data.

Finally, boundaries were drawn on rooftops using auto canny edge detection mechanism. Based on the type of roof and user input dimensions (length, width, tilt angle), a simulation of PV panels fitted on top of roofs are produced as the final end output. We discovered that the algorithm was pretty good in identifying obstacles or trees (shadows) on rooftops and hence not placing PV panels on obstructions. This module provides an additional flexibility of trying out the simulation with different customized values.

To summarize, the proposed 3-step mechanism takes a single satellite image as input, performs segmentation to detect buildings from the image, extracts rooftops alone using background subtraction technique, classifies the roof type as flat, gable, or hip, and then provides a simulation of PV panel placement on roof tops.

## **5.2 FUTURE WORK**

The MultiRes model was trained with only 1548 images owing to memory and hardware constraints. Training with even more images can lead to an even better accuracy. The classification of roof types is one of the areas of anticipated future development. We primarily seek to improve the outcomes of shallow CNN models by expanding the dataset size, either through labeling or by obtaining equivalent training data from other sources. The final module so far only considers the length, width, tilt angle and

obstructions on roof tops. It is planned to further include additional parameters like solar irradiation angle, climatic conditions and other necessary factors and validate the results from the final module with a solar expert.

## REFERENCES

- [1] Abdollahi, A., Pradhan, B., Gite, S., & Alamri, A. (2020). Building footprint extraction from high resolution aerial images using generative adversarial network (GAN) architecture. *IEEE Access*, Article 209517-209527. <http://dx.doi.org/10.1109/ACCESS.2020.3038225>.
- [2] Axelsson, M., Soderman, U., Berg, A. and Lithen, T., 2018, April. Roof Type Classification Using Deep Convolutional Neural Networks on Low Resolution Photogrammetric Point Clouds from Aerial Imagery. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)(pp.1293-1297).IEEE, <https://doi.org/10.1109/ICASSP.2018.846>.
- [3] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder–decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 39, 2481–2495.
- [4] B. Chatterjee and C. Poullis, "On Building Classification from Remote Sensor Imagery Using Deep Neural Networks and the Relation Between Classification and Reconstruction Accuracy Using Border Localization as Proxy," (2019) 16th Conference on Computer and Robot Vision (CRV), 2019, pp. 41-48, doi: 10.1109/CRV.2019.00014.
- [5] Buyukdemircioglu, Mehmet & Can, Recep & Kocaman, Sultan. (2021). Deep learning based roof type classification using VHR aerial imagery, *The*

International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. XLIII-B3-2021. 55-60. 10.5194/isprs-archives-XLIII-B3-2021-55-2021.

- [6] Chen, Mengge and Jonathan Li. (2019) “Deep convolutional neural network application on rooftop detection for aerial image.” *ArXiv* abs/1910.13509.
- [7] Edun, Ayobami & Harley, Joel & Deline, Chris & Perry, Kirsten. (2021). Unsupervised azimuth estimation of solar arrays in low-resolution satellite imagery through semantic segmentation and Hough transform. *Applied Energy*. 298. 10.1016/j.apenergy.2021.117273.
- [8] Jordan M Malof, Rui Hou, Leslie M Collins, Kyle Bradbury, and Richard Newell. 2015. Automatic solar photovoltaic panel detection in satellite imagery. In 2015 International Conference on Renewable Energy Research and Applications (ICRERA). IEEE, 1428–1431.
- [9] Kumar, Akash & Sreedevi, Indu. (2018). Solar Potential Analysis of Rooftops Using Satellite Imagery. *ArXiv* abs/1812.11606.
- [10] Nahid Mohajeri, Dan Assouline, Berenice Guiboud, Andreas Bill, Agust Gudmundsson, Jean-Louis Scartezzini, (2018). A city-scale roof shape classification using machine learning for solar energy applications, *Renewable Energy* Volume 121. Pages 81-93. ISSN 0960-1481. doi:10.1016/j.renene.2017.12.096.
- [11] Peiran Li, Haoran Zhang, Zhiling Guo, Suxing Lyu, Jinyu Chen, Wenjing Li, Xuan Song, Ryosuke Shibasaki, Jinyue Yan. (2021). Understanding rooftop PV panel semantic segmentation of satellite and aerial images for better using machine learning. *Advances in Applied*

Energy, Elsevier. Volume 4, 100057, ISSN 2666-7924. doi: 10.1016/j.adapen.2021.100057.

- [12] Qi, Chen & Wang, Lei & Wu, Yifan & Wu, Guangming & Guo, Zhiling & Waslander, Steven. (2018). Aerial Imagery for Roof Segmentation: A Large-Scale Dataset towards Automatic Mapping of Buildings. *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier. Volume 147, pp. 42-55.
- [13] Q. Li, Y. Feng, Y. Leng and D. Chen, " SolarFinder: Automatic Detection of Solar Photovoltaic Arrays," (2020) 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), 2020, pp. 193-204, doi: 10.1109/IPSN48710.2020.00024.
- [14] V. Golovko, S. Bezobrazov, A. Kroshchanka, A. Sachenko, M. Komar and A. Karachka, (2017), "Convolutional neural network based solar photovoltaic panel detection in satellite photos," 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), pp. 14-19, doi: 10.1109/IDAACS.2017.8094501.
- [15] X. Li, Y. Jiang, H. Peng and S. Yin, (2019), "An aerial image segmentation approach based on enhanced multi-scale convolutional neural network," IEEE International Conference on Industrial Cyber Physical Systems (ICPS), pp. 47-52, doi: 10.1109/ICPHYS.2019.8780187.