Semantic Crop Segmentation Using Deep Leaning Technique

MSc Research Project
MSc. in Data Analytics

Venkatesh Mukhopadhyay
Student ID: 20107790

School of Computing
National College of Ireland

Supervisor: Dr. Barry Haycock

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Venkatesh Mukhopadhyay
Student ID:	x20107790
Programme:	Msc Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Dr. Barry Haycock
Submission Due Date:	16/12/2021
Project Title:	Crop Segmentation
Word Count:	5777
Page Count:	19

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

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

Signature:	venkatesh
Date:	30th January 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

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

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

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

Semantic Crop Segmentation Using Deep Leaning Technique

Venkatesh Mukhopadhyay
x20107790

Abstract

Hunger and Malnutrition are two of the major concerns that the whole world is currently dealing with. Hunger can be dealt effectively by understanding Agricultural growth and studying the food system, that may provide valuable insight before applying advanced agriculture analytics. Plenty of papers in this field, work with the land images but do not study the effects of irregular land shapes, cloud coverage as well as small crop segments, which in turn provides an outcome which is driven using fewer practical data. This study focuses on applying data mining methodologies as well as deep learning methods such as Modified UNet with MobileNet V2 encoder, to identify crop segmentation by using images taken from Sentinel 2 satellite. Dataset is subjected to image augmentation which creates a balance in classes and provides more scalable data to work with. All results are evaluated based on Accuracy, precision, and Cross entropy loss and are compared with previous studies for a better understanding on the subject. This study will help in identifying the core issues or improvement areas to advise a food system that is effective in increasing the crop yield along with creating fitting schemes and policies for farmers in need.

Keywords: ***Crop Segmentation, Deep learning, Semantic image segmentation, Modified UNet, MobileNet V2 Encoder, Vegetation Indices, Transfer learning.***

1 Introduction

1.1 Background and Motivation

The world we know now is divided between Pre and Post pandemic conditions. We are still getting used to the worsening of some already critical issues such as food insecurity, hunger, malnutrition, and declining agricultural output over the last two years. According to UN, the world is facing economic slowdown which in turn is causing acute food insecurity. UN provides supporting data to show the increment of 60 million undernourished people since 2014, which brings the count as 690 million undernourished people worldwide by 2019. This count is expected to increase by 161 million more people due to the COVID by end of 2021 Coscieme et al. (2021). With the advancement of technology over the years, these issues can be analysed deeply, and solutions can be applied if proved fit. Understanding food system and different crop yields may further help in reducing the issues. We are going to research same issues in this study.

The dataset used for the research consists of more than 3K+ images for crop segmentation. This dataset contains Sentinel 2 Satellite images of a region in Kenya. All the images were captured for a year covering both winters and summer months across 11 time slices. In total 7 crops are identified and are used for the segmentations, these crops are Cotton, Dates, Grass, Lucerne, Maize, Pecan and Vineyard. Along with this some blank fields as well as Intercrop of Vineyard and Pecan is also considered for segmentation. The data is subjected to Data preparation, pre-processing, transformation, creating data pipeline, labelling and Augmentation before feeding the data into customized autoencoder build using MobileNetV2 as encoder for feature extraction and Unet as the decoder. The results are compared with the output from a pretrained UNet architecture. For a reference we can see the original and masked images in the Figure 1, below.

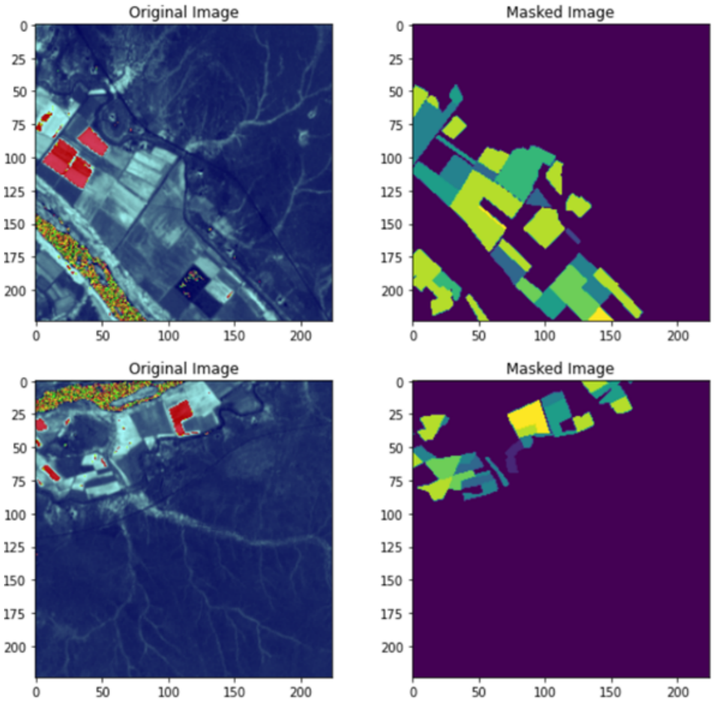


Figure 1: Original and masked crop images.

1.2 Research Question and Objective

1.2.1 Research Question

How can we improve the accuracy of pixel based multi-crop segmentation using advanced methods like Data Augmentation with CNN and Pretrained MobileNetV2 as encoder?

1.2.2 Research Objective

Research objectives for this study are mentioned below:

- Preparation and Pre-processing of Dataset and application on Data augmentation to achieve higher classification accuracy.
- Understanding the impact of adding additional spectral indices channels on performance parameters of the proposed model.

• Verifying all the results by training UNet and Model based on MobileNet V2 +UNet.

• Understanding the effect of different epochs on the performance of the model.

• Evaluating the results and comparing them with old studies on three evaluation matrix criteria's, i.e Accuracy, Cross Entropy and Precision.

1.3 Limitation and Challenges

Study of this kind comes with its own set of challenges. Few of them emerge from the low availability of data, or imbalanced classes as both these issues introduce bias which can result in underfitting of the model. At the same time, computational issue cannot be neglected as implementing a model correctly may take time and increase the overall cost of the project.

The history of related research is discussed in Literature review which is explained in section 2. This is followed by Research Methodology in section 3. Further discussion of Design and Implementation Specifications are discussed in section 4. Research results and previous study results are then Evaluated in section 5, which is then followed by Conclusion and Discussion in section 6. The report is ending with references in section 7.

2 Related Work

This section covers the history of some previous studies done on crop segmentation and how they are a benefit to us, as well as what these studies lack and how it could be improved with this study. Computer vision technology is getting more and more developed day by day, and as it progresses, lot more algorithms are coming up for crop segmentation. These algorithms are broadly classified in two sections by Ye et al. (2015), the first one is colour index based and second is learning based. Another study done by Lu et al. (2015) shows three categories, by adding colour model-based algorithm to the previous mentioned categories. Three more categories were recognised by Hamuda et al. (2016), which were colour index based, learning based and threshold based. These algorithms have proven efficient in picking out crop segments, but as not all the algorithms are covered in these categories, they do not define boundaries. To increase the range, we have considered following categories: -

2.1 Spectral Indices based Algorithms

When two or more spectral bands are taken into consideration and their pixels values are combined, we get Spectral Indexes. These are used to highlight all the pixels revealing land cover lack or abundance of vegetation in an image. In previous studies Bhandari et al. (2012), Rehman et al. (2019), we have seen the use of Normalized Difference Vegetation Index (NDVI) for the feature extraction as well as for predicting the crop yield. Whereas Phadikar and Goswami (2016), Rhyma et al. (2020) used both NDVI and Soil-Adjusted Vegetation Index (SAVI) for detecting Mangrove vegetation patterns and Diseases in Rice plants. It was noticed that both [2, 3-Our paper] faced lower accuracy rate on using just NDVI. While on the other hand, when another spectral index SAVI was used along with NDVI, it provided better accuracy for Rhyma et al. (2020)Phadikar and Goswami (2016). Some other studies were done to identify biophysical characteristics of any vegetation using Wide Dynamic Range Vegetation Index (WDRVI) in comparison with NDVI Gitelson (2004).

Sakamoto et al. (2011), Fawakherji et al. (2019) have completed their study using WDRVI spectral Index to Monitor growth indicators in sugar beet and to estimate the gross primary maize production, respectively. Study done by Bhandari et al. (2012) focuses on trying to find a spectral signature for Vegetation Index, Concrete surfaced structure, other textured area along with land cover classification. The use of NDVI is done to create artificial colour composition of the objects which are getting classified. The region used for study is of Jabalpur India. On the other hand Rehman et al. (2019) have conducted a study to use NDVI to assess the yield of rice crop for a region in California. Both the studies provide promising results and observations. NDVI worked well in identifying the spectral signature for Bhandari et al. (2012) varying density of vegetation. Due to the saturation of NDVI on higher values of crop N status, it just provided (r2 = 0.58), which could have been improved with the use of some other spectral indices. For differentiating the vegetation cover from the mangrove, the use of both NDVI and SAVI is done by Rhyma et al. (2020). This study observes that NDVI and SAVI have similar behaviour and result as they both have correlation value of 0.99. Similar concept of differentiating area using two spectral techniques is studied by Phadikar and Goswami (2016), to identify the disease infected area or brown spot in rice plants. This study provides a comparative study of four vegetation indexes over Grey image-based method. The results show that vegetation index can identify the brown spots by 2% more accuracy then grey image-based method, by giving the accuracy of 84%. The use of modified NDVI was used in the study done by Gitelson (2004), in which they have used the weighing coefficient (a) having the values between the range of 0.1-0.2, which turns NDVI into WDRVI. This is used to monitor the spatial distribution of different biophysical as well as different physiological features of vegetation.Sakamoto et al. (2011) has used Moderate Resolution Imaging Spectroradiometer WDRVI to assess the changes in daily gross primary production (GPP) for maize crops. While Gitelson (2004) finds WDRVI much effective while using the same band as NDVI, Sakamoto et al. (2011) further takes the same thought and observes the effect of software rendition with WDRVI provides better results (r2=0.86) compared to smoothed WDRVI (r2=0.83). Cao et al. (2020) agrees with NDVI's saturation issue due to high ground biomass highlighted by Rehman et al. (2019), It also highlights that when the near infrared band is having reflectance higher than 40%, its contribution to NDVI is quite less. Instead of using the same range as Gitelson (2004), this paper considers the value of weight coefficient between 0.05 to 0.5. The results for this study seem to be quite promising, as they received r2=0.963. It implements that using WDRVI is effective in reducing the saturation in growth indicator of beetroot when the biomass is high, this in turn increases the accuracy of the growth monitoring.

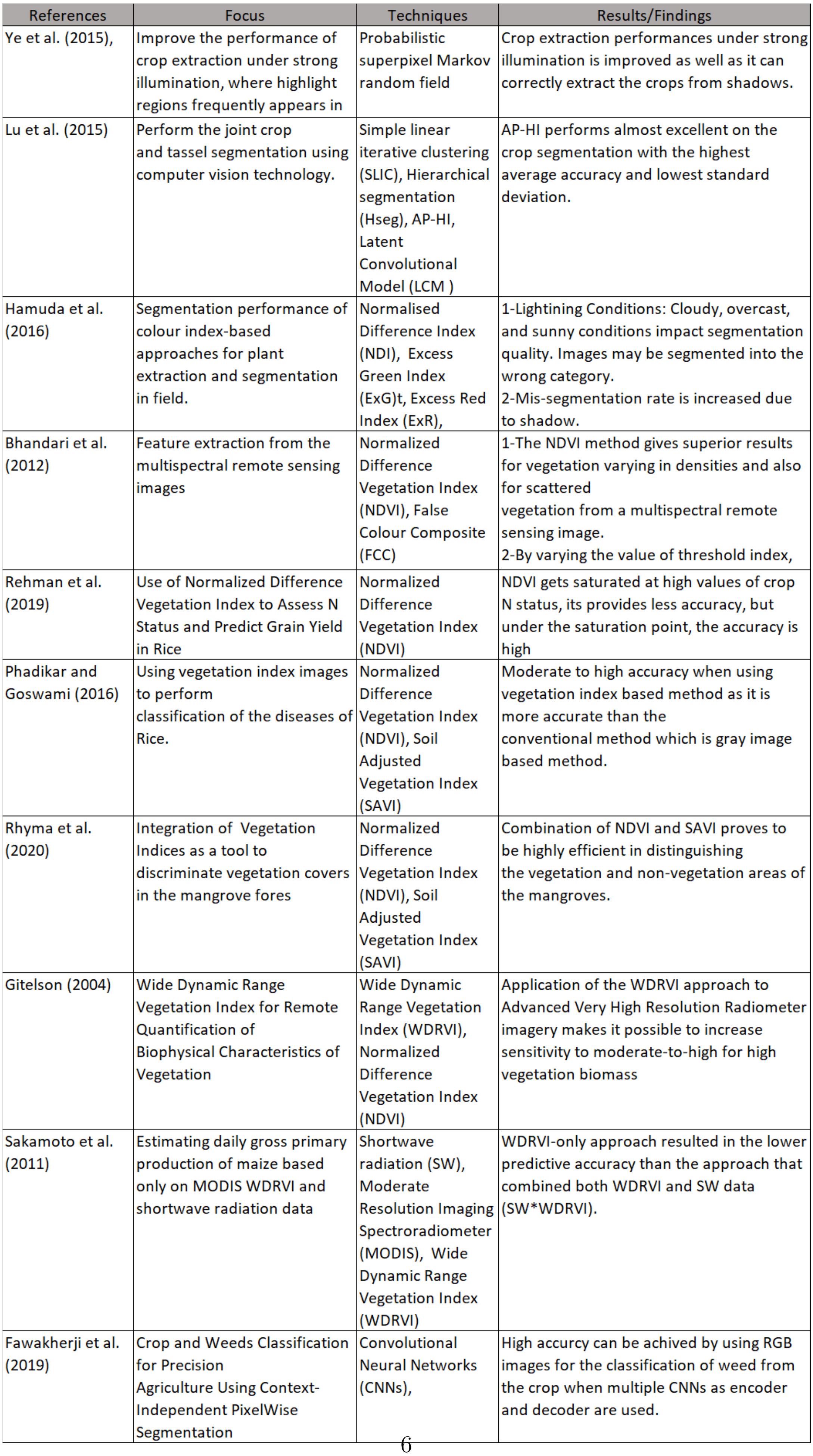
2.2 Pixel-based and Region Based Algorithms

A pixel is minute area of an image which contains information of the image, this information sometimes can be limited if colour is not involved. Study done by Hamuda et al. (2016) works on the similar fundamentals of pixel-based algorithms, to performs segmentation between crop and weed. On the other hand, the study done by Lu et al. (2016) uses the region-based approach for the segmentation of maise tassel and joint crop. The effectiveness of both the methods can be evaluated asLu et al. (2016) uses a super pixel-based algorithm for the segmentation along with graph-based algorithm to produce an edge preserved and region smoothed algorithm. Hamuda et al. (2016) talks about working in three steps, these steps start with a robust pixel-wise segmentation while the second step is of extracting the images patches that are containing plants. The third step is of applying Convolutional neural network based on VGG-16 for the classification of crop and weed. The results for this provide the accuracy of 90% with sensitivity as 94%, while precision and specificity being 87% and 88%, respectively. This study is not as effective due to the presence of inaccuracies in binary masks that are generated due to the segmentation, due to which some soil pixels are getting mixed up with weed pixels. Region based algorithm applied in Lu et al. (2016) judges its results based on effectiveness, efficiency and robustness, the crop segmentation experiences overall accuracy as 79% while maize tassel segmentation experiences overall accuracy as 74%. This study does not prove effective when maize tassel's colour is considered due to its tendency to change with time, as it is greener in beginning and turns yellow later. This gives us understanding that pixel-based algorithm work better than region based algorithms.

2.3 Semantic Segmentation Based Algorithms

Use of machine learning and deep learning is done in many ways to figure out the best possible solution for crop segmentation. Most of the artificial networks used for segmentation of soil from crop use multi or hyper spectral data to understand the radiometric indices. Dyson et al. (2019) has used a directional mathematical filter to reconsider radiometric filter. This paper has a generalized NDVI filter, which is used with a gradient cleaned dataset, in hopes to identify grass much effectively. The results for this study are favourable to combination DSM and NDVI instead of simple NDVI. The results seem to be better by 4 time as of normal NDVI. It is observed by Dyson et al. (2019) that the method proposed by them can help in identifying and segmenting soil from crop effectively. In a similar study done by Sodjinou et al. (2021) for the segmentation of weed and crops, the use of K-means algorithm is done with the combination of semantic segmentation. Images used are coloured and are taken from two different datasets to create more data. These images are first subjected to thresholding technique, in which just the plants were kept, and rest of the image was removed. Segmentation was done in two stages which started first with applying UNet for semantic segmentation and the results were later subjected to K-means algorithm. Using these results a comparative study was completed, which highlighted the effects of using UNet and k-means as positive. Both the studies deal with similar concepts but different approach. Study done by Sodjinou et al. (2021), provides much tangible aspects of using the combination of selected algorithms. This cannot be neglected that the image quality decreased due to the use of semantic segmentation in the research paper Sodjinou et al. (2021), and the absence of multi spectral images show that the study is done on basic concepts. This provides a base step for us to follow in this study.

With the help of these papers, we have understood the issues concerning with crop segmentation and have used the information to improve on the similar issues. We have taken multi-spectral data along to keep some of the quality of image, we have used a pixel-based algorithm to cover a huge biomass as region-based algorithms do not provide the significant efficiency. We also observed that using NDVI along with few other spectral indices' algorithms may provide a better accuracy in crop segmentation, this inspired us to use SAVI and WDRVI for the segmentation.



References	Focus	Techniques	Results/Findings
Ye et al. (2015),	Improve the performance of crop extraction under strong illumination, where highlight regions frequently appears in	Probabilistic superpixel Markov random field	Crop extraction performances under strong illumination is improved as well as it can correctly extract the crops from shadows.
Lu et al. (2015)	Perform the joint crop and tassel segmentation using computer vision technology.	Simple linear iterative clustering (SLIC), Hierarchical segmentation (Hseg), AP-HI, Latent Convolutional Model (LCM)	AP-HI performs almost excellent on the crop segmentation with the highest average accuracy and lowest standard deviation.
Hamuda et al. (2016)	Segmentation performance of colour index-based approaches for plant extraction and segmentation in field.	Normalised Difference Index (NDI), Excess Green Index (ExG)t, Excess Red Index (ExR),	1-Lightining Conditions: Cloudy, overcast, and sunny conditions impact segmentation quality. Images may be segmented into the wrong category. 2-Mis-segmentation rate is increased due to shadow.
Bhandari et al. (2012)	Feature extraction from the multispectral remote sensing images	Normalized Difference Vegetation Index (NDVI), False Colour Composite (FCC)	1-The NDVI method gives superior results for vegetation varying in densities and also for scattered vegetation from a multispectral remote sensing image. 2-By varying the value of threshold index,
Rehman et al. (2019)	Use of Normalized Difference Vegetation Index to Assess N Status and Predict Grain Yield in Rice	Normalized Difference Vegetation Index (NDVI)	NDVI gets saturated at high values of crop N status, its provides less accuracy, but under the saturation point, the accuracy is high
Phadikar and Goswami (2016)	Using vegetation index images to perform classification of the diseases of Rice.	Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI)	Moderate to high accuracy when using vegetation index based method as it is more accurate than the conventional method which is gray image based method.
Rhyma et al. (2020)	Integration of Vegetation Indices as a tool to discriminate vegetation covers in the mangrove fores	Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI)	Combination of NDVI and SAVI proves to be highly efficient in distinguishing the vegetation and non-vegetation areas of the mangroves.
Gitelson (2004)	Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation	Wide Dynamic Range Vegetation Index (WDRVI), Normalized Difference Vegetation Index (NDVI)	Application of the WDRVI approach to Advanced Very High Resolution Radiometer imagery makes it possible to increase sensitivity to moderate-to-high for high vegetation biomass
Sakamoto et al. (2011)	Estimating daily gross primary production of maize based only on MODIS WDRVI and shortwave radiation data	Shortwave radiation (SW), Moderate Resolution Imaging Spectroradiometer (MODIS), Wide Dynamic Range Vegetation Index (WDRVI)	WDRVI-only approach resulted in the lower predictive accuracy than the approach that combined both WDRVI and SW data (SW*WDRVI).
Fawakherji et al. (2019)	Crop and Weeds Classification for Precision Agriculture Using Context-Independent PixelWise Segmentation	Convolutional Neural Networks (CNNs),	High accurcy can be achived by using RGB images for the classification of weed from the crop when multiple CNNs as encoder and decoder are used.

Figure 2: Literature Review Summary.

3 Methodology

The selection of a methodology for the research is done based on its effects on overall project implementation, as well as its ability to be flexible with the requirements along with the domain. A research is not just applying the models to the dataset, but the whole process of data gathering as well as data transformation for the model to be applied. We have divided our whole research in following sections:

Data Gathering
Data Processing and Transformation
Image processing
Image Augmentation
Modelling
Evaluation

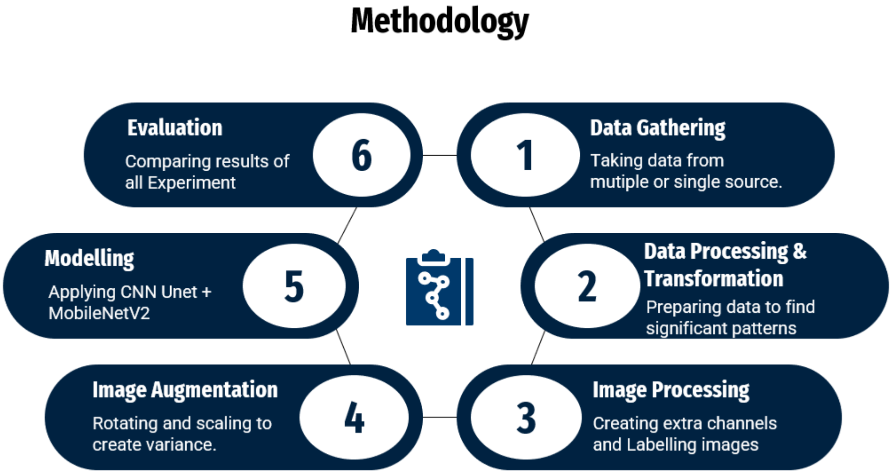

Methodology
Evaluation
Comparing results of all Experiment
6
1
Data Gathering
Taking data from mutiple or single source.
Modelling
Applying CNN Unet + MobileNetV2
5
2
Data Processing & Transformation
Preparing data to find significant patterns
Image Augmentation
Rotating and scaling to create variance.
4
3
Image Processing
Creating extra channels and Labelling images

Figure 3: Adopted Methodology.

3.0.1 Data Gathering

The very first step of any research is to gather the data and then prepare it for further processing and transformation. The data we used for the research is a Sentinel 2 satellite data, consisting of multi temporal and multi spectral images taken over a region in South Africa to understand the crop types grown in the area. Once the raw data is gathered from different or single source, it is the then prepared to be processed and transformed in the format useful for the research.

3.0.2 Data processing and Transformation

For the preparation of data, we have applied Time stamp processing in order to categorize the data based on correct timestamps. This was considered as this dataset contains multi temporal time series data which can be divided easily according to the time the image was taken by the satellite. We have then calculated multiple spectral indices values, which helps us in differentiating the crop and vegetation from soil as well as land cover. Using this technique, we can detect and visualize all vegetation areas and observe any abnormalities that may be faced in the growth. As our data was consisting of images having irregular plot sizes and blank spaces, we have applied Invalid geometry removal using GeoPanda data frame. We have used the geopanda data-frame to deal with empty geometrics which are geometries having no coordinates or area. By using this we are trying to find if there are any empty geometries present in our dataset. Once the data is cleaned and processed, we transform our data is format that is easier to deal with or can provide much more information when compared to its original form. As our dataset consisted of JFP format images, we have converted the images into PNG format. This is done in order to save the special meta data of the images. We have also created tiles of 224*224 resolution out of a high-resolution image dataset as it has reduced our overall processing time and has kept the details intact too. The special meta data acquired from the data transformation is later used for the image augmentation.

3.0.3 Image Processing

Now we have clean, processed, and transformed data that will be subjected to image processing, which will help us in interpreting the valuable information contained by images. We have also created additional channels by combing and subtracting the existing bands from the images.to be specific we have calculated the NDVI, SAVI and WDRVI vegetation indices Later these channels are concertinaed to form 1 channel. These indices help in increasing the sensitivity of the model towards the vegetation characteristics and reduces other noises such as refection from background soil and atmospheric effects. Another requirement before feeding the images to the model is to create labels for the images. This process helps in determining the correct categories of the images as well as create different segments. We have saved the images in geopanda data frame and labelled them using GeoTiff. What it does is, take the information embedded in an image file and geo-reference it for satellite images. This helps in understanding the area much better while trying to segment the crop types.

3.0.4 Image Augmentation

Data augmentation is done with majority of image datasets due to the availability of data. This is an important step while dealing with image data as more the data, better is the performance of the Model We have performed the augmentation in order to gain more data as well as create variance in the dataset. We have using image rotation as one of the augmentation techniques by rotating the data horizontally and vertically. We have also used scaling, in which we resize the image according to our requirements. As the acquired images were high resolution, reducing them to a smaller size helped in processing them quickly as well as reduced the chances of overfitting. Now the images are ready to be fed to the model to train it , which is explained in the next section.

3.0.5 Modelling

We have worked with two major algorithms in this section, CNN-UNet and MobileNetV2. The modelling process is explained below: The dataset consists of more than 3700 images which are divided into 3 parts namely training, validation and testing in the ratio of 80:10:10 respectively. We have kept the testing data less as it will help us to deal with the imbalance nature of dataset. We have designed couple of experiment in order to perform crop segmentation. Cross Entropy/Focal loss is used to determine the performance of these experiment as the dataset is imbalance in nature and focal loss uses different parameters to measure the final loss provided by each class to reduce bias. The experiments also include training the proposed models with and without data augmentation and comparing the output. The first experiment is performed using a pretrained CNN architecture call U-net. Which helps in masking the crop field according to the crop type. The results from this architecture where not too exciting, as we were getting a lot of noises in the segmented output image and the cross entropy/focal loss were also high. For the second experiment we have modified the U-net architecture and integrated another pretrained encoder called MobileNetV2. When compared with other pre-existing pretrained models MobileNetV2 has lesser number of hyperparameters to train. The MobileNetV2 architecture is explained in the implementation section of this paper in details. MobileNetV2 requires less amount of storage space and the processing power is also hight compared to other encoder and also provides higher accuracy and better end segmentation results.

3.0.6 Evaluation

There are two sets of experiments done for the research, and these results are compared with each other in based on accuracy, F1 score and cross entropy. We have observed that the second experiment was much more successful and provided better results as expected. All these results are further discussed in the Results and discussion section.

4 Design Specification

In order to implement the proposed architecture and run all the experiments we have used google colab. The version of python used is Python 3.8 and the GPU provided by google colab is nvidia sims. Keras library is predominately used to build the proposed architecture which is built on Tenserflow 2.x. Multiple libraries have been used like geopandas, sklearn , rasterio.

5 Implementation

This section covers the implementation of the proposed deep learning architecture methodology to segment crop into 9 different classes which is explained in details followed by systematic flow diagram. Two approaches are designed to perform sematic image segmentation and the performance of both the architecture are compared to evaluate which model is more accurately able to segment the crops. The steps that are followed for implementation of the planned deep learning architecture are as follow : Data preparation for each model, visualizing the data, perform some exploratory data analysis , data pipeline setup and creating the model, and defining the parameters, training the model , finding out the appropriate losses and reduce the losses , testing the model , assessing model outcomes, Checking the designed architecture for over or underfitting , as well as optimizing model parameters for consistent results. All of the models that have been built are based on neural networks but having different model architecture in order to optimize the results. Firstly, the data is collected from appropriate data source. The data consists of around 3000 images which include 9 classes of crops which include inter crops as well as vacant fields. The images are taken from sentinel 2 satellite and are time series data with 11 time slices spread across 1 year. The data is split into 3 parts for training , validation and testing in the ratio of (80:10:10) respectively. The data is processed and all the images from single month are put into one label in order to reduce variance in the data. The images from sential-2 stallite are of hight resolution and have 13 channels containing red, green , blue , infrared, ultra infrared , cloud fliter etc. For the purposes of crop segmentation 3 new vegetation indexes like (NDVI,SAVI and WDRVI)are calculated using the band 2, 3,4 and 8 channels which helps in masking of the crop images. The provided shapefile is loaded into a geopanads dataframe which is then used to check the invalid plot shapes and the local CRS is set. The meta data is used to construct the label image array. The images are converted into titles of (244 x 244) 10% of the complete data was used as validation data during training of the model to track accuracy during validation, which is obtained at the completion of each epoch. All the models execute for a specified number of epochs. Because the model can get over-fit when epoch value is high, different callback settings were utilized to halt the training early. The callback value is used to monitor the validation losses , once the validation losses stops improving the training is also stopped . Once the model is trained ‘leaning curve’ between the valuation loss and epoch is plotted to understand the growth or reduction of looses over time. Once we have achieved satisfactory loss value the model is used to make prediction using the testing data.

5.1 Spectral Indexes calculation for studying vegetation

In this research work 3 different types of spectral indexes are used in order to help with the correct crop class prediction. Spectral indexes can be defined as the combination of 2 or more different wavelengths that are reflected from objects present on earth surface and are captured by the remote sensing satellite, which can be used to determine the characteristics of the vegetation.

5.1.1 NVDI - Normalized Difference Vegetation Index

Vegetation greenness can be quantified using NDVI spectral index. NDVI is helpful in understanding plant health and assessing the density of vegetation. NDVI is traditionally determined as a ratio of the red color wavelength (R) and near infrared (NIR) values.
NDVI = (NIR - RED) / (NIR + RED)

5.1.2 SAVI - Soil-Adjusted Vegetation Index

SAVI is a vegetation index that uses a soil-brightness correction factor to try to minimize soil brightness impacts. This is commonly utilized in arid places with medium to little vegetation.

SAVI = (NIR Red) (1 + L) / (NIR + Red + L)

where, L is a canopy background adjustment factor

5.1.3 WDRVI- Wide Dynamic Range Vegetation Index

The Wide Dynamic Range Vegetation Index, or WDRVI, measures levels of red light wavelength and near-infrared light in a similar way to the NDVI, but it is more precise because it can detect slight changes in the crop canopy in medium to high vegetation cover, which is especially important for mature crops and crops with dense canopies.

$$\text{WDRVI} = (\text{a} * \text{rho(NIR-rho(red))}/(\text{a} * \text{rho(NIR)} + \text{rho(red)})$$

The value of a is a constant which rages from 0.1-0.2

5.2 Model Architecture

5.2.1 U-Net Implementation

An autoencoder is an deep neural network that can be used to segment images. An autoencoder model is made up of two parts: an encoder and a decoder. The encoder half of an convolutional deep neural network is used for feature extraction from the input image, and the decoder part reconstructs the image from the features. U-net is specially designed autoencoder that is particularly good at semantic segmentation. Brox and Ronneberger created it for segmentation of medical image Ronneberger et al. (2015). The two primary sections of this U-shaped architecture are the encoder or contracting part and the decoder or expansion part. Figure 4 depicts a general U-net model diagram. The masked images and the original images are feed as input into the Unet architecture. The left section of the Unet architecture is composed of encoder or the upsampling path on the left side as shown in Fig. 4 which constitutes of multiple section of layers. Two convolutional layers of dimension 3 x 3 are followed by a layer max pooling having dimension of 2 x 2 in each section. The feature extracted by the preceding layer is feed to the next layer as input for for each stages. The features exacted from the encoder section of Unet is passed into the decoder part for prediction. The decoder or the expanding path on the right side of Fig. 3 is constitutes of multiple sections of layers. Two sets of convolution layers of dimension 3x3 which are followed by an upsampling layer of dimension 2x2 in each section. Input image from the encoder is feed in to the decoder in which it reconstrues the original input image on the basis of contextual information Input . The horizontal link from the encoder section to the decoder section are one of the most important elements of the U-net architecture Guo et al. (2021). This horizontal or skip link passes all the features that were execrated from the encoder to the decoder side which helps in improving segmentation accuracy . Another important advantage of the skip link is it helps in avoiding the gradient descent issue . Feature maps having same dimensions are combined in the concatenation layer . The final layer in the decoder section is a convolutional layer of 1x1 dimension and it transfer the feature numpy array from preceding layer to a layer which gives binary output .

5.2.2 MobileNetV2 Implementation

Mobilenet are specifically designed CNN network to be using in mobile devices.. Depth-wise separable filters is a unique feature in a MobileNet architecture . Networks using Depth-wise separation filters have lesser number of computation making the architecture computationally less intensive compared to standard convolution network . Srinidhi et al. (2021)This feature is important for models having many convolutional layers. The MobileNetV2 model comes under MobileNet series . MobileNetV2 comes with two extra functionalities. It is designed to have series of linear bottleneck layers with direct linkage between them. The bottleneck layer filters the amount of data that can travel across

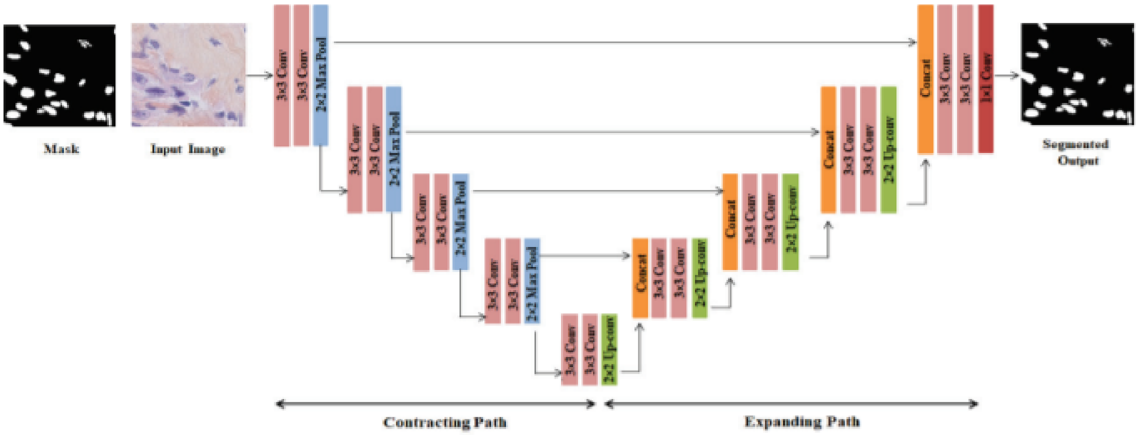


Figure 4: U-net Architecture.

the network. MobileNetV2 architecture is made up of multiple layers. First layer of the MobilenetV2 consists of convolutional layer having 32 feature extractor , next layer is made up consists of 17 layers of having bottleneck feature The MobileNetV2 design is thoroughly explained in Sandler et al. (2018). The MobileNetV2 model's structure is shown in Fig 5. The bottleneck blocks and their respective input sizes are described in Table I of the MobileNetV2 design. Where c is the count of output channels, n denotes the count of repeated blocks, and s denotes the count of strides.

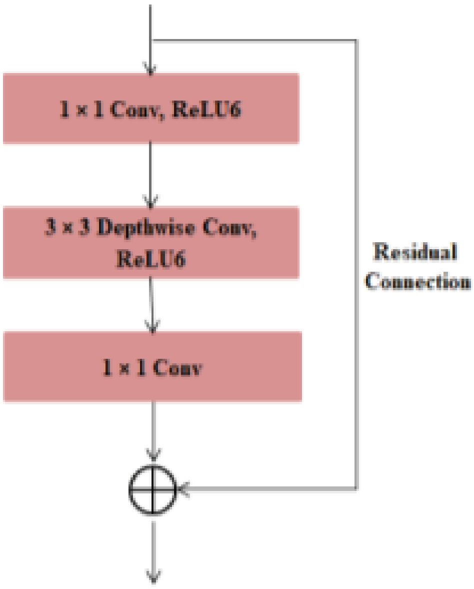


Figure 5: Bottleneck Residual Block of MobileNetV2.

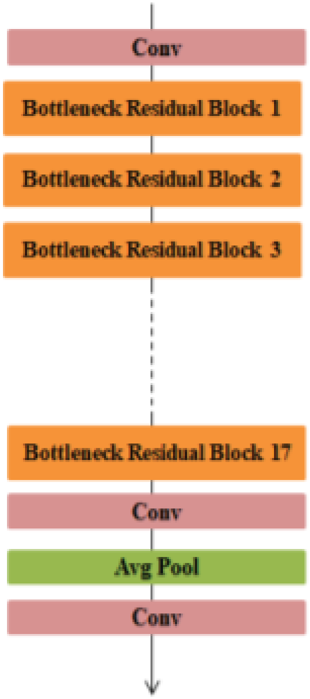

Conv
Bottleneck Residual Block 1
Bottleneck Residual Block 2
Bottleneck Residual Block 3
Bottleneck Residual Block 17
Conv
Avg Pool
Conv

Figure 6: MobileNetV2 Architecture.

5.2.3 Modified Unet(U-net+ MobileNetV2)

In this architecture, pretrained Mobilenetv2 is used as encoder instead of using encoder from the Unet rest of the architecture remains the same as Unet. Figure 8 shows a graphical depiction of the suggested model. The upsampling and downsampling parts are connected via a horizontal connector. We are using a pretrained Mobilenet encoder which consists of 17 layers of bottleneck connections. Output from 1,3 ,6 and 13 bottleneck layers of encoder section are taken and feed into the Unet encoder. The decoder section consists of multiple layers segmented into 4 groupsGeffray et al. (2016). Every group consistis of an upsampling layer of dimension 2 x 2 , Concatenation operation is performed after this layer which is then followed by convolutional network of dimension 3 x 3 .The decoder or the expanding section gets the input from block number 13 of bottleneck layer which has dimension of (16 x 16 x 192) this output is then passes onto the first group. Similarly skip connection are used to pass on the output from bottleneck layer of the encoder to the decoder having proper dimensions (32 × 32, 64 × 64, 128 × 128, 256 × 256).

The output from the bottleneck cell number 6 is feed to group 1 which has dimension of (32 × 32 × 96) . Similarly, the output from bottleneck cell 3 and 1 are passed to group number 2 and 3 respectively. The dimension of the output from bottleneck cell 3 and 1 are (64 × 64 × 48) and (s 128 × 128 × 48) respectively.The input image is feed into the group number 4 of the concatenation layer which is a part of the decoder.The final layer consists of a convolutional layer having dimension of 1 x 1 which helps in generation the output segmentation mask .

We have also performed some experiments with and without training the modified Unet model on the crop dataset.Figure 8 shows the output when the model is not trained on the crop dataset and data augmentation is not performed. we can see there is a lot

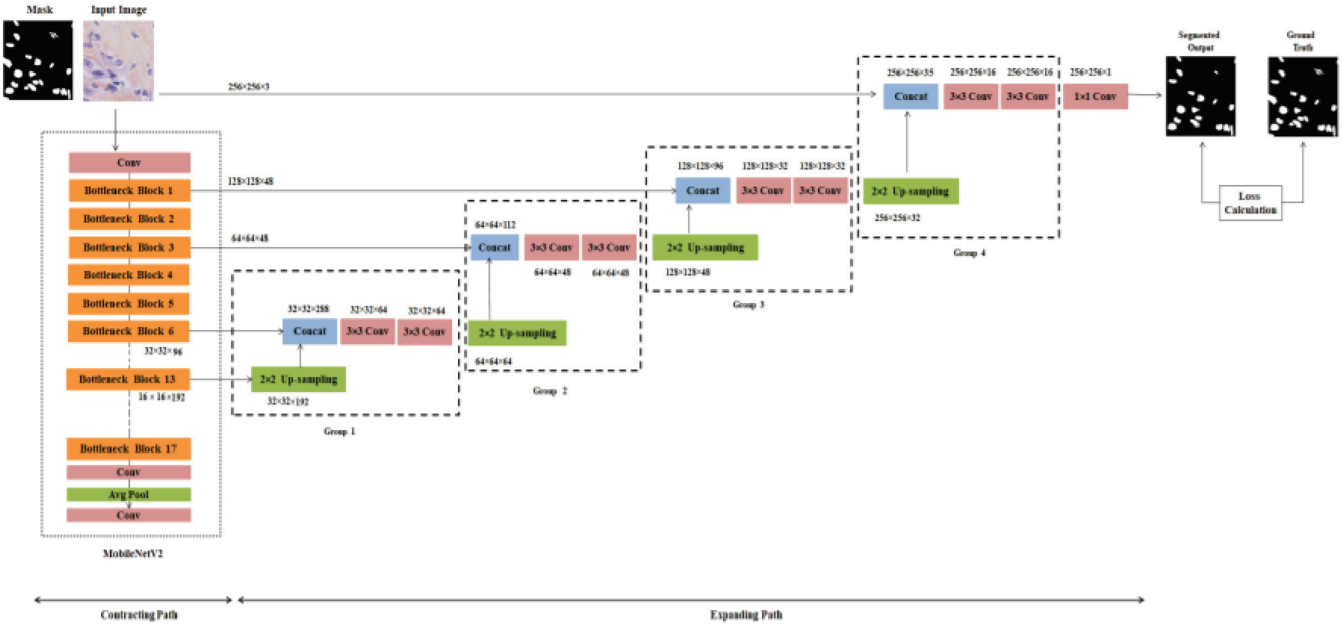


Figure 7: Proposed U-net Model with MobileNetV2 as Encoder.

of noise in the predicted mask.Figure 9 shows the output when the model is trained and data augmentation is also used.

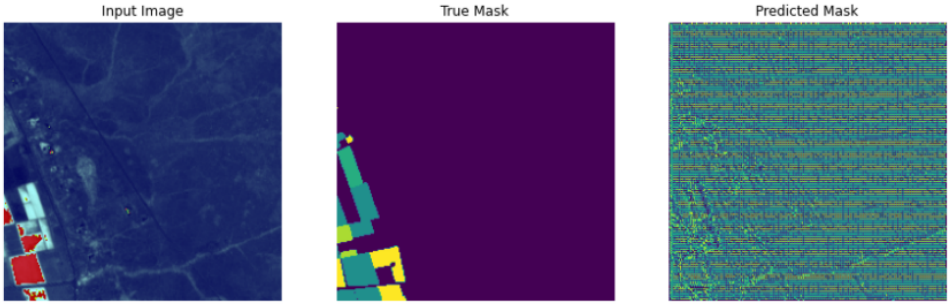


Figure 8: Output when no data augmentation used and model is not trained.

6 Experiment and Evaluation Metrics

6.1 Training

The dataset is split into 3 sets (training , validation and testing) in an 80:10:10 ratio after augmentation. During training, the network is fed with the Crop images from the training dataset as well as the ground truth masks. After training , validation dataset is feed into the same model architecture to determine the accuracy, validation losses and hyper tune the parameters in order to improve the end results. Once we have achieved satisfactory results the testing dataset is feed into the model which will predict the segmentation or mask of the crop fields. The training of the model is done by running 100 epoch. Adam

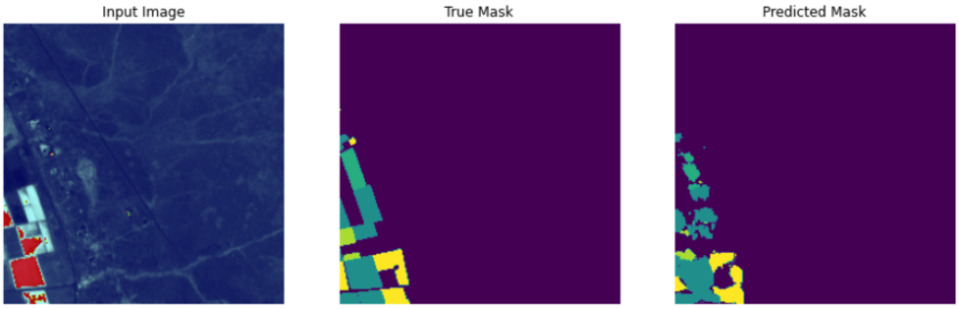


Figure 9: Output when data augmentation is implemented and model is trained.

optimizer is implemented in order to optimize the learning rate and seed up the training process. The training is accelerated using the Adam optimizer, an adaptive learning rate optimization technique. The loss function for the model is focal cross entropy, as illustrated in (7).

6.2 Evaluation

Evaluation metrics used for this research work are the accuracy, F1 score , cross entropy losses.

6.2.1 Accuracy

The accuracy score of an image segmentation model is calculated on per pixel basis. It is the ratio of the pixels that are correctly segmented divided by the total pixels in the input image. The parameters used to define accuracy are as follow

a. **True positive** – Number of crop pixels that are correctly predicted.

b. **True negative** (T N) – Number of background pixels that are correctly predicted.

c. **False positive** (F P) – number of incorrect predictions in which crop pixels are predicted as background pixels

d. **False negative** (F N) - number of incorrect predictions in which background pixels are predicted as crop pixels.An accuracy value closer to 1 implies a better segmented result

Accuracy can be defined by below equation:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

6.2.2 F1 score

F1 score is the calculated to understand the performance of an model. It is defined as the harmonic mean of precision and recall. Higher the value of F1 score , model performance is also better. The reason behind using F1 score as performance metrics is the presence of slight imbalance in the dataset. F1 score takes account of the distribution of the data in the dataset.

6.2.3 Cross Entropy loss/ Focal loss

Focal loss is an improvement upon the cross entropy loss which attempts to address the issue of class imbalance in the dataset by assigning higher weight values to the classes that are easily misclassified and less weight are assigned to that classes that are difficult to misclassify .

7 Results and Discussion

The proposed model's experiment outcomes are discussed in this section. In this paper, we design and compare a U-net and a modified Unet model based on pretrained encoder as MobileNetV2 with and without data augmentation. We have also compared the results of the out come with and with out training the model with the crop image dataset. The results of the experiments are shown in the Figure 6. To assess the effectiveness of the proposed research architecture, we have compared the performance of the final modified Unet model and original Unet model with and without data augmentation. We have also performed some experiments with and without training the modified Unet model on the crop dataset. Figure 10 shows graph between the training and validation loss over Epoch. We can observe that the training and validation losses tends to decreases as the epoch approaches to 100.

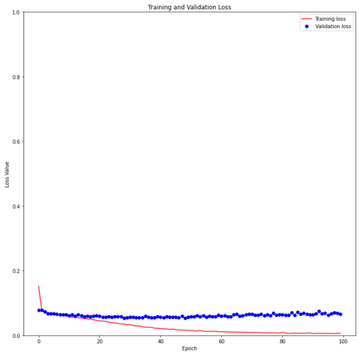
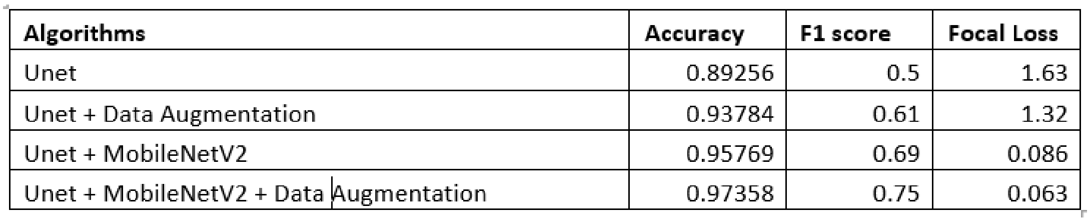


Figure 10: Training and validation loss vs Epoch

From the data visualization and segmentation results it can be clearing seen that the performance of the original Unet model is quite low and full of noise when it is not trained with the crop dataset and data augmentation is not used. Once we have trained the modified Unet model and used data augmentation with the proposed same the performance of the model gets better. This can be seen from figure 11. Based on the evaluation parameters we can see the modified Unet model with MobileNetV2 as pretrained encoder gives the best segmentation results when it is trained, and data augmentation is used.



Algorithms	Accuracy	F1 score	Focal Loss
Unet	0.89256	0.5	1.63
Unet + Data Augmentation	0.93784	0.61	1.32
Unet + MobileNetV2	0.95769	0.69	0.086
Unet + MobileNetV2 + Data Augmentation	0.97358	0.75	0.063

Figure 11: Segmentation Results

A normalized confusion matrix is drawn to understand the segmentation performance of the model.From figure 12 we can see that the model is doing a very good job of identifying no crop i.e presence of soil and segmentation of crop id 3,5,6 and 8 are also performed with high accuracy.

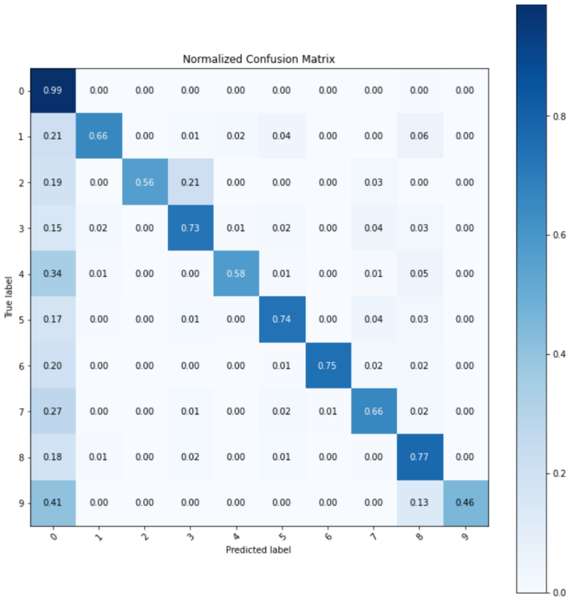


Figure 12: Normalized Confusion Matrix

8 Conclusions and Future Work

In this research work, we have built a model based on based on pretrained MobilkeNetV2 and Unet for the purpose of crop image segmentation from crop field images. For the encoder part of the architecture we have used the pretrained MobileNetV2 instead of the original Unet and for the decoder part we have used the Unet architecture itself. An augmented high resolution images of crop field dataset which is taken from sentinel -2 satellite are used for training and testing. The model architecture proposed by us is then compared with few more existing segmentation algorithms with similar dataset and results are noted. This results obtained from the proposed segmentation model outperforms all the other algorithms.

The system was capable of accurately detecting and segmenting different classes of crops, as well as producing photos with sharp edges. It's worth noting that the image's quality suffers as a result of the segmentation process. As a result, in future work, quality enhancement methods will be used to increase the output image quality. In multi-spectral images, the proposed method will also be employed for crop segmentation. We can also deal with the atmospheric noise in a better way in the future work.

References

Bhandari, A., Kumar, A. and Singh, G. (2012). Feature extraction using normalized difference vegetation index (ndvi): A case study of jabalpur city, *Procedia technology*

6: 612–621.

Cao, Y., Li, G. L., Luo, Y. K., Pan, Q. and Zhang, S. Y. (2020). Monitoring of sugar beet growth indicators using wide-dynamic-range vegetation index (wdrvi) derived from uav multispectral images, *Computers and Electronics in Agriculture* **171**: 105331.

Coscieme, L., Mortensen, L. F. and Donohue, I. (2021). Enhance environmental policy coherence to meet the sustainable development goals, *Journal of Cleaner Production* **296**: 126502.

Dyson, J., Mancini, A., Frontoni, E. and Zingaretti, P. (2019). Deep learning for soil and crop segmentation from remotely sensed data, *Remote Sensing* **11**(16): 1859.

Fawakherji, M., Youssef, A., Bloisi, D., Pretto, A. and Nardi, D. (2019). Crop and weeds classification for precision agriculture using context-independent pixel-wise segmentation, *2019 Third IEEE International Conference on Robotic Computing (IRC)*, IEEE, pp. 146–152.

Geffray, S., Klutchnikoff, N. and Vimond, M. (2016). Illumination problems in digital images. a statistical point of view, *Journal of Multivariate Analysis* **150**: 191–213. **URL:** *https://www.sciencedirect.com/science/article/pii/S0047259X16300203*

Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation, *Journal of plant physiology* **161**(2): 165–173.

Guo, Y., Duan, X., Wang, C. and Guo, H. (2021). Segmentation and recognition of breast ultrasound images based on an expanded u-net, *PLOS ONE* **16**: e0253202.

Hamuda, E., Glavin, M. and Jones, E. (2016). A survey of image processing techniques for plant extraction and segmentation in the field, *Computers and Electronics in Agriculture* **125**: 184–199.

Lu, H., Cao, Z., Xiao, Y., Li, Y. and Zhu, Y. (2015). Joint crop and tassel segmentation in the wild, *2015 Chinese Automation Congress (CAC)*, IEEE, pp. 474–479.

Lu, H., Cao, Z., Xiao, Y., Li, Y. and Zhu, Y. (2016). Region-based colour modelling for joint crop and maize tassel segmentation, *Biosystems Engineering* **147**: 139–150.

Phadikar, S. and Goswami, J. (2016). Vegetation indices based segmentation for automatic classification of brown spot and blast diseases of rice, *2016 3rd International Conference on Recent Advances in Information Technology (RAIT)*, IEEE, pp. 284–289.

Rehman, T. H., Borja Reis, A. F., Akbar, N. and Linquist, B. A. (2019). Use of normalized difference vegetation index to assess n status and predict grain yield in rice, *Agronomy Journal* **111**(6): 2889–2898.

Rhyma, P., Norizah, K., Hamdan, O., Faridah-Hanum, I. and Zulfa, A. (2020). Integration of normalised different vegetation index and soil-adjusted vegetation index for mangrove vegetation delineation, *Remote Sensing Applications: Society and Environment* **17**: 100280.

Ronneberger, O., Fischer, P. and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation, *in* N. Navab, J. Hornegger, W. M. Wells and A. F. Frangi (eds), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer International Publishing, Cham, pp. 234–241.

Sakamoto, T., Gitelson, A. A., Wardlow, B. D., Verma, S. B. and Suyker, A. E. (2011). Estimating daily gross primary production of maize based only on modis wdrvi and shortwave radiation data, *Remote Sensing of Environment* **115**(12): 3091–3101.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520.

Sodjinou, S. G., Mohammadi, V., Mahama, A. T. S. and Gouton, P. (2021). A deep semantic segmentation-based algorithm to segment crops and weeds in agronomic color images, *Information Processing in Agriculture* .

Srinidhi, C. L., Ciga, O. and Martel, A. L. (2021). Deep neural network models for computational histopathology: A survey, *Medical Image Analysis* **67**: 101813.
URL: *https://www.sciencedirect.com/science/article/pii/S1361841520301778*

Ye, M., Cao, Z., Yu, Z. and Bai, X. (2015). Crop feature extraction from images with probabilistic superpixel markov random field, *Computers and Electronics in Agriculture* **114**: 247–260.