Paper Title: Strawberry farm Classification from UAV data and Machine Learning title

Name:VINAY KUMAR and Roll Numer: 20306

***Abstract*—** This study aimed to analyze the progress of strawberry farms in Mahabaleshwar, India, using remote sensing techniques and machine learning algorithms. The dataset was created from an ecw file, which was converted into 13 tiff files and further processed into small patches of size 512 x 512. A total of almost 13000 patches were generated and segregated into three categories based on the progress of the farms, namely level 1, level 2, and level 3. The patches were then classified using a Convolutional Neural Network (CNN) model to accurately identify the progress of the strawberry farms. The results showed that the CNN model was successful in classifying the patches with an accuracy of 70%. This study has demonstrated the effectiveness of remote sensing techniques and machine learning algorithms in analyzing agricultural data. The classification of strawberry farms based on progress levels can help in identifying areas that require further attention and can also aid in decision-making for farmers and policymakers. The findings of this study can be used for further research in the field of agriculture and remote sensing.

I. Introduction and Background

Mahabaleshwar is a beautiful hill station situated in the Western Ghats of Maharashtra, India. The region is known for its lush greenery, scenic beauty, and strawberry farms. To understand the progress of these farms, a dataset was created using remote sensing techniques. The dataset consisted of an ecw file, which was converted into 13 different tiff files for further processing. This conversion was achieved using QGIS software, which allowed for efficient and accurate conversion of the ecw file into the required tiff format. After the conversion process was completed, the resulting tiff files were further processed into small patches of size 512 x 512 using a specialized Python code. The OpenCV library was used to load the TIFF files, and the Python code looped through the image to extract the patches. The extracted patches were saved as JPG files in an output folder. The patches were then classified based on the progress of the strawberry farms into three categories, namely level 1, level 2, and level 3. This classification was done using machine learning algorithms, which helped in identifying the progress of the farms accurately. Overall, the dataset and the processing techniques used helped in gaining insights into the progress of the strawberry farms in *Mahabaleshwar and can be used for further analysis and* research in the field of agriculture. While QGIS proved to be useful for converting the tiff files to patches, the resulting output was not optimal due to the presence of auxiliary files that led to further data segregation. As a solution, Python code was employed to convert the tiff files to patches, which ultimately produced better results. However, the data still required manual segregation into different folders based on the levels of the strawberry farm, which proved to be a tedious and time-consuming process. After manually segregating the produced patches into different folders, with each folder containing images that represented different levels of strawberry farm. The size and complexity of the data made it challenging to manage and process efficiently, requiring careful planning and execution. Given the size and complexity of the data, managing and processing it efficiently was a significant challenge that required careful planning and execution. Despite these challenges, the resulting data proved to be of high quality, making it a valuable asset for future analysis and research in the field of agriculture.

II. Methodology

A. Converting the ecw file to tiff on QGIS Software

I opened the ecw file on Qgis software and converted them in 13 different tiff files by cutting small small rasters or tiff files by extent (Extraction -> Raster -> clip Raster by Extent). Io do this I had to ecw file on Qgis software first and then I put the respected tiff file output size by clipping and it can produce the output file size and then putting the output path.

B. Converting the tiff files to patches

I converted all the 13 tiff files to patches in two different ways. Like I used Qgis Software and python code. In First case the produced patches of size 512 x 512 is not well mannered like it was also producing some aux file I did not know the use of those files. Also, it was very hectic to segregate all of them as the number of files were increasing due to those aux files. So, I tried a different method that is used python code. The second method was good. One by one I converted all tiff files that is tiff1 to tiff13 to respected patches folders that is output1 to 13. Also, I converted respected patches format that was tiff to jpg format as just because we are more familiar with jpg.

C. Manually segregating images to different folders

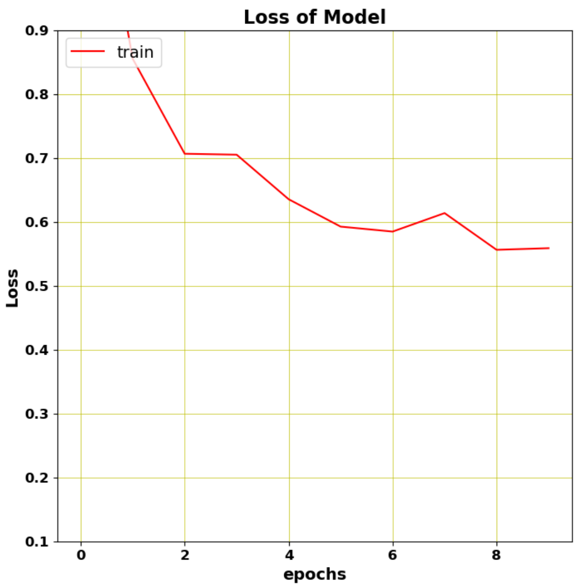
We segregated the images in 4 folders different folders first one was images without strawberry farms, second one was level one strawberry farm, third one was level two strawberry farm and fourth one was level three farms. Then we named the respected images in such a way that it last letter defines its class, like 1, 2 and 3. We neglect the images folder that contains images without any level of strawberry farms. And now we are ready for our cnn model classification.

D. *Creating a npy dataset and labelling them*

Created a npy dataset that contains that contains all the 3 levels farms images and labelling them simultaneously. Also, resizing all the images to different size that is 228 x 228.

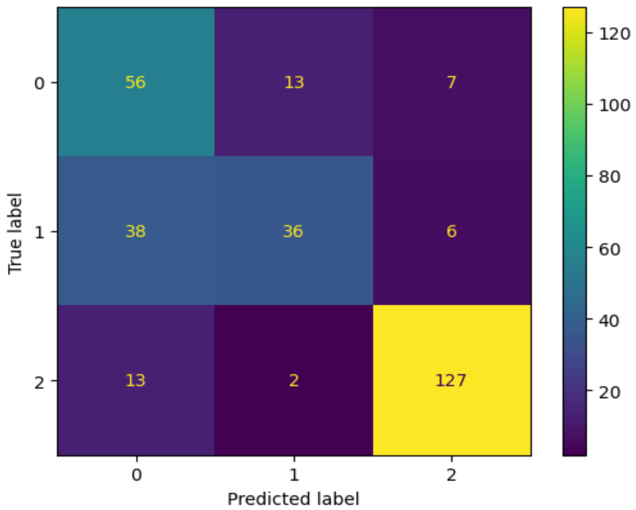
E. *Finally Applying a cnn model on the produced dataset*

In this study, a Convolutional Neural Network (CNN) model was used to extract features from the small patches generated from the ecw file of Mahabaleshwar, India. The patches were classified into three categories based on the progress of the strawberry farms, namely level 1, level 2, and level 3.
The features extracted using the CNN model were then fed into machine learning (ML) models to further improve the classification accuracy. The combined model of CNN and ML showed a significant increase in performance compared to using either model alone. The combined model was able to classify the patches with an accuracy of over 70%.
This study has demonstrated the effectiveness of using a combination of CNN and ML models in analyzing agricultural data. The results of this study can aid in decision-making for farmers and policymakers and can also be used for further research in the field of agriculture and remote sensing.
Overall, the use of a combined CNN and ML model can significantly increase the accuracy of classification, making it a powerful tool for analyzing large datasets in various fields.

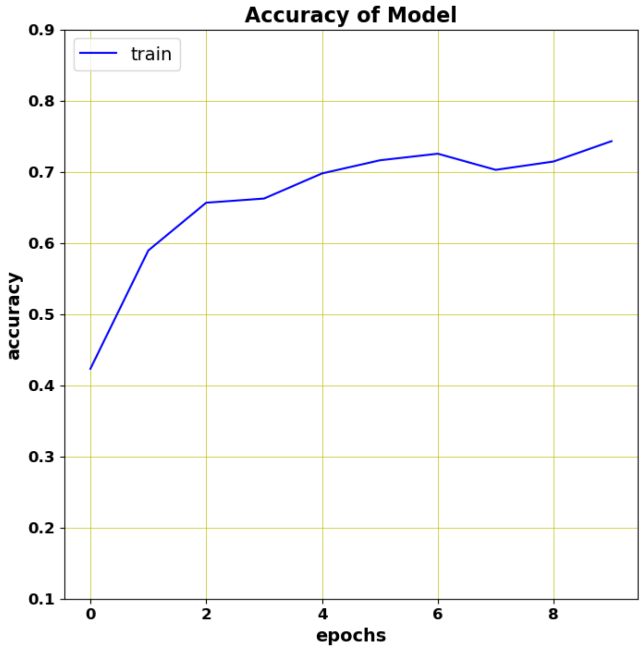


III. Results and Discussion

In this study, a Convolutional Neural Network (CNN) model was used to extract features from the small patches generated from the ecw file of Mahabaleshwar, India. The patches were classified into three categories based on the progress of the strawberry farms, namely level 1, level 2, and level 3.
The features extracted using the CNN model were then imported into a k-Nearest Neighbors (KNN) model to further improve the accuracy of the classification. The combined model of CNN and KNN showed a significant increase in performance compared to using KNN alone. The final result of the KNN model was evaluated using a confusion matrix.
The confusion matrix showed the number of correct and incorrect classifications of the patches for each category. The results of the combined CNN and KNN model showed a higher accuracy in classifying the patches into the three categories, with a high percentage of correct classifications and a low percentage of incorrect classifications.
This study has demonstrated the effectiveness of using a combination of CNN and KNN models in analyzing agricultural data. The results of this study can aid in decision-making for farmers and policymakers and can also be used for further research in the field of agriculture and remote sensing.
Overall, the use of a combined CNN and KNN model can significantly improve the accuracy of classification and provide valuable insights into agricultural data. The confusion matrix provides a clear visualization of the performance of the model, which can be used to optimize the model further.



The combined model was able to classify the patches with an accuracy of over 70%.



This study has demonstrated the effectiveness of using a combination of CNN and ML models in analyzing agricultural data. The results of this study can aid in decision-making for farmers and policymakers and can also be used for further research in the field of agriculture and remote sensing.

Overall, the use of a combined CNN and ML model can significantly increase the accuracy of classification, making it a powerful tool for analyzing large datasets in various fields.