Monitoring and Prediction of Agriculture Field in Embedded System with Machine Learning

Aditi Khekale[1], Drakshyani Desai[2]
[1]*Research Scholar, JJT University*
[2]*Assitant Professor, JJT University*

***Abstract*—For thousands of years, farmers have turned to the skies and tracked clouds to determine how a cropping season will turn out. But there is a new cloud on the horizon. Farmers are increasingly taking the help of big data and embedded systems to undertake precision farming. In these technologically advanced farms, autonomous combines harvest crops with the help of the Global Positioning System. Drones flying overhead, map the field and send the data to the cloud, where it is processed. The farmer gets all this data in a tablet (connected to the internet). He can see the areas in need of his intervention. He makes corrections and the data is immediately fed to the choppers and tractors which automatically correct their courses.**

***Index Terms*—Embedded System, Smart Agriculture, Monitoring, Machine Learning, Data analysis.**

I. INTRODUCTION

Pest damage and development are affected by the rise in global temperature brought by climate change. When the temperature rises, the metabolic rate of insects increases, driving them to consume more food and inflict more damage. Growth rates of several insect species are also affected by temperature. For each degree of average global warming of the earth's surface, worldwide agricultural losses due to insect pests are expected to increase by 10% to 25%. Pesticides and chemical treatments have long been used by farmers to keep pests away. The use of pesticides for crop protection is on the rise, with negative Agriculture which is considered as one of the prime occupation of human being has 64% of the total available land occupied by agriculture out of which it consumes 85% of available fresh water [1].The problems faced by the traditional instrumentation based on discrete and wired method has been solved through the implementation of wireless method which provides with a low-cost wireless controlled irrigation solution and a real-time monitoring of the field [2]. Many new ideas are being taken up to allow agricultural automation to flourish and deliver its full potential. The introduction to this modern technology not only consumes less time but it also allows the farmer to carry out his agricultural activities more efficiently. In this system, all the devices work on their own with the help of the inputs received from the sensors. The sensors which are placed in the field provide us with the data in regard to the field which is then controlled and monitored by programmable controller. consequences for human health and increased environmental damage to soil and groundwater. On the other hand, this also increases the risk of pests developing pesticide resistances.

II. PROCEDURE FOR PAPER SUBMISSION

Data gathering, data pre-processing (i.e., data preparation that includes feature extraction), and ML classification models are the basic steps of ML applications.

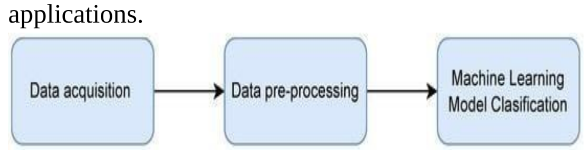

Figure 1. Stages used for Data processing

Data Acquisition

Data acquisition is the process of gathering data from various sources systems [6]. Previous studies gather their data various sources to be used for ML techniques. Some of them produce their own images by taking pictures of plants in greenhouses, such as in the studies from Gutierrez et al. and Raza et al. [5]. However, image data acquisition using manual processes, as done by many, generally results in small image data-sets, which can compromise the development of effective ML-based models. Weather data collection is also proposed in the literature using for instance sensors in greenhouses. Meteorological data can also be obtained from weather stations of regional areas, which typically store records for a longer period of time.

Images can be collected using search engines on their own. This approach can get a large number of images, but ground truth must be checked by domain experts, and data cleaning is frequently used to filter out images that do not meet the requirements.

Remote sensing images from satellites and drones have the advantage of being able to retrieve image data for large agricultural areas. Remote sensing data from satellites typically consists of multi-temporal and hyper-spectral imagery data, which can be used to assess the development of the crops. This task can be performed by monitoring the evolution of vegetation indices, which provide important information about the development status of the crop fields. Spectral imagery can be used for computing different vegetation indexes, such as those proposed in, which are robust to variations on the sun illumination, an important advantage when compared to visible light spectrum imagery.

Images retrieved from drones can also be used, but have additional needs: to define the path of the device; to coordinate the drone position with the camera for image acquisition; and to correct geometric distortions on each acquired image in order to merge the different acquired images in order to reconstruct a larger image of the whole field.

Therefore, it can be stated that data consists of different modalities and variables. With ML-based and data analysis techniques it can be possible to understand their interaction and how they relate to a studied outcome. In the context of the cultivation fields, the questions are usually: which disease is affecting crops? What pest is causing damage? What is the relation between weather data and disease and pest occurrence?

Freely available data-sets can also be utilized for the development of ML-based applications. This enables researchers to directly compare the performance of different ML techniques and approaches. The data conditions a significant impact on the performance of ML models. Data-sets should be representative and include enough records for the model to perform an effective generalization.

III. LITERATURE REVIEW

H.Sabrol 2021 proposed a classification tree model based on supervised learning techniques. The extracted features from the segmented image are fed to this classification tree, which classifies tomato plant disease with 97.3% of accuracy. Merits include classification tree results with good accuracy. Demerits include the latest classification techniques that can be used.[8]

Mohammed Brahimi et al, 2020 proposed Convolutional Neural Network for improved classification. The significance of CNN is an automatic feature extraction from raw input images. Thus the trained model achieves 99.18% of accuracy. Merits include high performance. Demerits include computation and the size of deep NN can be reduced.[9]

Alvaro Fuentes et al, 2021 proposed an object detection method with Faster r-CNN, R-FCN, and SSD algorithms merged with the latest feature extractors of deep CNN. This proposed model, R-CNN with VGG-16 shown better recognition results. The merit is false positives are reduced in the training phase. The demerit is slag in performance.[10]

Halil Durmas et al, 2022 proposed Deep Learning-based network architectures namely sequent and Alex Net for tomato leaves disease detection. Alex Net provided better classification accuracy of 95.65%. The merit includes, squeeze Net is light in weight requires low computational needs. Demerits are long training time and small batch size.

Jia Shijie et al, 2022 proposed CNN network-based VGG16 model. The VGG is good in image classification and localization problems. Overall 89% of classification accuracy is achieved. Merit includes, the proposed model is very effective. Demerits include relatively high-quality test images that are only used.

Santosh Adhikari et al, 2022 proposed a machine learning model with CNN architecture, and data augmentation reduces overfitting during training the model. Thus overall accuracy of 89% is obtained. Merits include data augmentation that gives better performance. Demerits include transfer learning is required to classify all diseases.[13]

Konstantinos P. Ferentinos 2022 proposed deep learning CNN based architecture models. The VGG model achieved a high success rate of 99.53% after fine-tuning and training. Merits include the proposed method is highly potential and more robust. Demerits include, works only for trained input values with the same database.[14]

Endang Suryawati et al, 2023 proposed the VGGNet model. The proposed model achieved an accuracy of 95.24%, which supports disease detection effectively. Merits include model accuracy. Demerits include the absence of model robustness.[15]

Aravind Krishnaswamy Rangarajan et al, 2023proposed pre-trained deep learning-based architectures for tomato disease classification. The Alex Net model achieved 97.49% of classification accuracy with minimum execution time. Merits include better accuracy with less execution time. Demerits include an increase in mini-batch size decreases accuracy.

Belal A.M.Ashqar 2023 proposed deep CNN with grayscale and full-color models. Thus proposed model achieves 99.84% of accuracy with a full-color model. The gray-scale only learned the shape and patterns of the leaf and the color model learned to identify diseased leaves. Merits include highly feasible ones. Demerits include recognition slightly lacks.[17]

KekeZhang et al, 2023 proposed the optimization methods with a deep neural network. The optimal SGD with the ResNet technique provides the highest accuracy of 97.28%. Merits include, fine-tuning saves computational resources and time. Then good performance is achieved. Demerits are time-consuming.

IV. METHODLOGY

Data Pre-Processing:
Data pre-processing techniques vary according to the used ML-based approach. In the case of image data, feature extraction can be done manually by applying computer vision algorithms, or automatically using deep learning.

Manual feature extraction processes typically demand pre-processing steps such as noise reduction or contrast enhancement. The researchers have to decide and select which feature extractors are more suitable for the problem at hand. When using deep learning, pre-processing is typically focused on data augmentation, enriching the training data-set in order to achieve a better model generalization. Deep learning shows better results when directly analyzing the originally acquired images when compared with the use of images converted to grey-scale or subject to background removal. This is a useful finding because background removal can be a complex and arduous task for images taken in field conditions, with complex and varying background. When comparing the performance of ML models based on manual feature selection with models based on deep learning, the latter has shown better performance in studies that compared both approaches using the same input data.

Highlighting the region of interest of the leaves and reducing the background noise can increase the model's performance. This is valid for plant disease identification as well as for insect classification.

Machine Learning Models:
The studies presented along this paper have mostly used SVM, RF, or deep learning-based ML models. All of these have shown promising results, highlighting the potential of using ML techniques for disease and pest classification, detection, and prediction. SVMs are robust and useful in high dimensional spaces due to their use of kernel trick. RF can avoid overfitting due to the high number of trees trained in different subsets of data. Deep learning usually achieves the best classification results due to its ability to create and extract hierarchical features from the inputs. Deep learning beats other ML models, particularly in image classification domains, especially when using pre-existing CNN architectures such as Inception and ResNet.

Despite deep learning models achieving higher accuracy values, SVM and RT can also achieve high values with accuracy above 94%, especially in disease classification on laboratory images. SVM also achieves high accuracy, with values above 90%, in the detection of tomato diseases.

RNNs are capable of establishing relationships between weather data and pest occurrence, surpassing other models such as RF and SVM.

In scenarios where data is difficult to obtain, models trained with a lower amount of data can benefit from the use of TF, rather than having the models trained from scratch. Most studies have their models pre-trained on large data-sets for image classification such as ImageNet or COCO. The inclusion of the Plant Village data-set with ImageNet for pre-training helps to improve the accuracy of models for disease classification on images acquired in the field. TF is typically applied by training some of the top layers of the pre-trained model jointly with the new classifier.

An alternative would be to address lack-of-data problems using few-shot learning approaches, as suggested in [3].

IV. PUBLICATION PRINCIPLES

Freely available data-sets can also be utilized for the development of ML-based applications. This enables researchers to directly compare the performance of different ML techniques and approaches. The data conditions a significant impact on the performance of ML models. Data-sets should be representative and include enough records for the model to perform an effective generalization. Machine learning combines three main components: model, data and loss. Machine learning methods implement the scientist cprinciple of \rial and error". These methods continuously validate and re ne a model based on the loss incurred by its predictions about a phenomenon that generates data.

VII. CONCLUSION

Compared with traditional image processing methods, which deal with plant diseases and pests detection tasks in several steps and links, plant diseases and pests detection methods based on deep learning unify them into end-to-end feature extraction, which has a broad development prospects and great potential. Although plant diseases and pests' detection technology is developing rapidly, it has been moving from academic research to agricultural application, there is still a certain distance from the mature application in the real natural environment, and there are still some problems to be solved.

Deep learning technology has made some achievements in the identification of plant diseases and pests. Various image recognition algorithms have also been further developed and extended, which provides a theoretical basis for the identification of specific diseases and pests. However, the collection of image samples in previous studies mostly come from the characterization of disease spots, insect appearance characteristics or the characterization of insect pests and leaves. Most of the research results are limited to the laboratory environment and are applicable only to the plant diseases and pests images obtained at the time. The main reason for this is that the growth of plants is cyclical, continuous, seasonal and regional. Similarly, the characteristics of the same disease or pest at different growing stages of crops are different. Images of different plant species vary from region to region. As a result, most of the existing research results are not universal. Even with a high recognition rate in a single trial, the validity of the data obtained at other times cannot be guaranteed.

REFERENCE

[1] Sabrol, H., and K. Satish. "Tomato plant disease classification in digital images using classification tree." 2016 International Conference on Communication and Signal Processing (ICCSP). IEEE, 2016.

[2] Sardogan, Melike, Adem Tuncer, and Yunus Ozen. "Plant leaf disease detection and classification based on CNN with LVQ algorithm." 2018 3rd International Conference on Computer Science and Engineering (UBMK). IEEE, 2018.

[3] Sharada.P Mohanty, David.P. Hughes, MarcelSalathe (Front. Plant Sci., 22 September **2016),** Using Deep Learning for Image-Based Plant Disease Detection.

[4] Shijie, Jia, Jia Peiyi, and Hu Siping. "Automatic detection of tomato diseases and pests based on leaf images." 2017 Chinese Automation Congress (CAC). IEEE, 2017.

[5] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv. arXiv:1409.1556. 2014.

[6] Sladojevic, Srdjan, et al. "Deep neural networks based recognition of plant diseases by leaf image classification." Computational intelligence and neuroscience 2016 (2016).

[7] Suryawati, Endang, et al. "Deep structured convolutional neural network for tomato diseases detection." 2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS). IEEE, 2018.

[8] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the AAAI conference on artificial intelligence. 2016.

[9] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proceedings of the 2015 IEEE conference on computer vision and pattern recognition, Boston, MA, USA, 7–12 June; 2015. p. 1–9.

[10] The Support Vector Machine (SVM) Based Near-Infrared Spectrum Recognition of Leaves Infected by the Leafminers, Wu Dake; Ma Chengwei

[11] Tm, Prajwala, et al. "Tomato leaf disease detection using convolutional neural networks." 2018 Eleventh International Conference on Contemporary Computing (IC3). IEEE, 2018.

[12] Xiao, Q.; Li, W.; Kai, Y.; Chen, P.; Zhang, J.; Wang, B. Occurrence prediction of pests and diseases in cotton on the basis of weather factors by long short-term memory network. BMC Bioinform. 2019, 20, 688. [Google Scholar] [CrossRef]

[13] Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. arXiv. arXiv:1611.05431. 2017.