Smart Red Spider Pest Detection for Pesticide Optimization in Strawberry Crops a Short Literature Review

Inti Toalombo[1], **Inziman Ul Haq**[2]
Department of Electrical Engineering and Mechatronics, Faculty of Engineering, University of Debrecen, Hungary [1,2]

Abstract. Strawberry production is a vital economic activity for many families in Ecuador, yet it faces a significant challenge from red spider mite (Tetranychus urticae) infestations, which can devastate crops and lead to substantial financial losses. Current methods of detection rely on manual visual inspection, which is time-consuming and often inaccurate, leading to inefficient pesticide use and environmental damage. This study aims to address this issue by developing a machine learning-based solution to detect red spider mites more accurately and assist farmers in optimizing pesticide application. Using IBM Watson Studio, we will train deep learning models such as YOLO and PMML-based algorithms on a large dataset of strawberry leaf images collected from web-crawling, open-source platforms, and field data. The system aims to reduce pesticide overuse while ensuring timely pest control interventions, promoting sustainable farming practices in Ecuador's strawberry industry, while contributing to the preservation of our environment and health.

Keywords: Red spider mite, strawberry crops, Pest detection, Machine Learning, Smart Farming, IBM Watson.

1. Introduction

Strawberry production is a crucial economic activity for thousands of families in Ecuador. However, it faces a significant threat from the red spider mite, a tiny pest capable of causing severe financial losses. For example, a farmer with 30,000 strawberry plants typically invests around $15,000 initially, plus an additional $200 to manage pests. These investments are at risk due to the destructive potential of red spider mite infestations, which can be hard to detect and control, especially when the infestation reaches critical levels.

The red spider mite, *Tetranychus urticae*, is the most common pest in strawberry crops. Its nearly microscopic size makes early detection challenging, with female mites measuring 0.5mm and males 0.3mm. They are typically visible only with magnification, and their eggs, which are colorless and only 0.14mm in diameter, are difficult to see until they are close to hatching. Female mites lay two to six eggs daily, leading to around 15 generations per year. As a result, by the time a farmer notices the infestation, it may be too late to effectively control it. [1]

Pesticide application is the most common method used by Ecuadorian Farmers for red spider mite control. While this approach can be effective, it's not always applied in time, especially during the summer when infestations spread more rapidly. This often leads to farmers applying pesticides "just in case," which not only increases costs but also causes environmental damage. Excessive pesticide use contaminates soil and water, harms beneficial insects, and disrupts ecosystems. [2] Pesticides can accumulate in the food chain, and there is a risk of pests developing resistance, reducing the effectiveness of future treatments. [3]

This research demonstrates and explores the potential of Machine Learning (ML) for optimizing pesticide application by improving red spider mite detection. By collecting data on factors like leaf images, environmental conditions, and pesticide usage, ML models can be trained to assist farmers in pest management. This study aims to introduce AI-driven tools to Ecuadorian farmers, moving from traditional, environmentally harmful methods to more advanced, sustainable pest control solutions by introducing the Smart Farming concept.

2. Related Work

Recent advancements have been made in detecting red spider mites using AI technologies such as deep learning. For instance, Congliang et al. (2023) employed two different deep learning models, YOLOv4 and Faster R-CNN ResNet50, achieving detection accuracies of 93.5% and 86.3%, respectively. The researchers utilized three different smartphone cameras to collect over 3,000 images of the red spider mite and its natural predator, creating a comprehensive dataset. They trained the models to evaluate detection performance with three different image sizes (160 x 160, 320 x 320, and 640 x 640), finding a direct correlation between image size and detection accuracy. The 640 x 640 pixel size was determined to be ideal, yielding a 93.3% accuracy for the YOLOv4 model. Additionally, the researchers found that the YOLOv4 model could effectively detect the pest under varying illumination conditions.[4]

In the same way, Lee et al. (2023) have built a deep-learning model to detect strawberry diseases and pest infections in the very early stage. For this purpose, they have developed an automatic data acquisition system with which have collected a large data set of over 13,000 plant images. The data acquisition system actually is a rail-based mechatronics machine using a step motor that allows transversal movements which equipped with a 3625 x 2448 pixel camera covering three beds of strawberries in one scavenging and collecting large amounts of raw data. For Data Labeling they have created a gold standard data set with the support of plant biology experts, annotating images manually through VGG Image Annotator (VIA). They emphasize the model robustness through data augmentation in order to prevent overfitting to make the model applicable in not knowing situations that is to say out of data acquisition conditions. PyTorch package has been used to perform online data augmentation by using image transform functions. Regarding the Detection Model, the researchers have trained nine YOLO models with varying network sizes and training for what they built a training PC with GPU device NVIDIA Tesla V100 with driver and cuda 11.1. Training was done by randomly selecting 90% of the data for training and the rest of the data to test. The obtained results have shown that this model based on the YOLOv5 can detect pest signs in strawberries with decent performance represented by the Area Under Precision-Recall Curve (AUPRC), 0.819 of its average. [5]

Another selected approach has been developed by Kim et al. (2023) based on Deep-Learning classification model to detect multiple pest infections in strawberry crops. The researchers have collected over 10,000 thousand datasets of strawberry leaves from smart farms in Korea and also, and they have used open-source data from the Kaggle platform. They used data augmentation techniques to improve the quality of collected data which includes healthy and unhealthy leaves. The researchers have selected Regnet and Efficientnet models in combination with Pseudo-labeling to have the higher accuracy in pest detection. This model could be integrated into mobile applications for real-time assistance in pest detection in smart framings. This approach does not focus on detecting a specific pest but various infections by classifying five types of leaf images achieving an overall accuracy of 85.6%. [6]

Despite the study developed by Yamada et al. (2024) not focused on strawberry crops pest detection in cotton plants is a noteworthy approach since they have used hyperspectral imaging in combination with

Machine learning to detect and classify Red Spider Mite (*Tetranychus urticae*). By collecting and processing hyperspectral data from cotton leaves with different levels of pest infections they have trained various models such as Random Forest(RF), Support Vector Machine(SVM), and Feedforward Neural Network (FNN) is RF the most accurate model since it has achieved the highest accuracy ranging from 80% to 100%. This solution offers a promising approach for the early detection of red spider mite infestations enhancing significantly pest management and reducing pesticide usage.[7]

A promising study based on cloud service has been developed by Yang et al. (2020) which consists in a low-cost system to classify strawberry disease for this project a novel-supervised multinetwork fusion classification model has been developed involving the designing of easy-to-use cloud-based systems which consist of a Location network, a Feedback network and a classification network to identify categories of common strawberry diseases. This model can identify effectively diseased regions in strawberry crops without the need for annotation such as bounding boxes. In comparison with popular Convolutional Neural Networks (CNN) and other models such as Bilinear-CNN, PC-DenseNet-161, Autoaugmented, and LFC-Net (k=4) this Self-supervised multinetwork fusion classification model has shown promising results in the range of 88.45% to 92.48% accuracy on various strawberry detection pests including the red spider mite. The client (mini program) has been released on the WeChat platform reporting marvellous results since its performs is almost perfect in the actual test, which supports the effectiveness of the system in pest management in the strawberry production industry.[8]

Recent advancements have been made in detecting red spider mites using AI technologies such as deep learning. For instance, Congliang et al. (2023) employed two deep learning models, YOLOv4 and Faster R-CNN ResNet50, achieving detection accuracies of 93.5% and 86.3%, respectively. The researchers utilized three different smartphone cameras to collect over 3,000 images of red spider mites and their natural predators, creating a comprehensive dataset. They trained the models to evaluate detection performance with three image sizes (160 x 160, 320 x 320, and 640 x 640), finding a direct correlation between image size and detection accuracy. The 640 x 640 pixel size yielded the best results with 93.3% accuracy for the YOLOv4 model. Furthermore, the YOLOv4 model effectively detected pests under varying illumination conditions [4].

Similarly, Lee et al. (2023) developed a deep-learning model to detect strawberry diseases and pest infections in their early stages. They designed an automatic data acquisition system, using a rail-based mechatronic machine equipped with a 3625 x 2448 pixel camera, which collected over 13,000 plant images. A gold-standard dataset was manually annotated with the help of plant biology experts using the VGG Image Annotator (VIA). They emphasized model robustness through data augmentation to prevent overfitting, using PyTorch for online data augmentation. Their YOLOv5-based model demonstrated decent performance, achieving an average Area Under Precision-Recall Curve (AUPRC) score of 0.819 [5].

Another approach by Kim et al. (2023) involved developing a deep-learning classification model to detect multiple pest infections in strawberry crops. The researchers collected over 10,000 images from smart farms in Korea and open-source platforms like Kaggle. Data augmentation techniques were employed to enhance the quality of the dataset. The researchers used RegNet and EfficientNet models in combination with Pseudo-Labeling, achieving an overall detection accuracy of 85.6%. This model has the potential to be integrated into mobile applications for real-time pest detection in smart farming systems [6].

Yamada et al. (2024) focused on a different crop (cotton) but provided a noteworthy approach for red spider mite detection using hyperspectral imaging combined with machine learning. By collecting hyperspectral data from cotton leaves with various levels of mite infestation, they trained models like Random Forest (RF), Support Vector Machine (SVM), and Feedforward Neural Network (FNN). The Random Forest model proved to be the most accurate, achieving detection accuracies ranging from 80%

to 100%. This method offers a promising early detection tool for pest management, reducing pesticide use [7].

Finally, Yang et al. (2020) developed a cloud-based system for classifying strawberry diseases using a novel supervised multinetwork fusion classification model. This system includes a location network, a feedback network, and a classification network to identify common strawberry diseases. The model, which requires no image annotations such as bounding boxes, achieved accuracies ranging from 88.45% to 92.48% when compared to models like Bilinear-CNN and PC-DenseNet-161. The system, deployed via a WeChat mini-program, demonstrated near-perfect results in practical testing, making it highly effective for pest management in strawberry production [8].

3. Materials and Method

3.1. *Model Framework Overview*

As explored in the state-of-the-art section, several machine learning models are widely used for training models based on image datasets. One of the most popular models is the deep learning model *You Only Look Once* (YOLO), known for its high speed and relatively low computing cost, making it suitable for such projects. However, for this study, we will utilize IBM Watson Studio, a robust platform for building, training, and deploying machine learning applications across industries, including agriculture. The main reason for choosing IBM Watson Studio is its support for various open-source frameworks like PyTorch, TensorFlow, and scikit-learn, in addition to offering machine learning models such as neural networks, decision trees, regression, and YOLO [9].

Additionally, IBM Watson Studio supports the use of Predictive Model Markup Language (PMML), enabling models such as K-nearest neighbours, Naive Bayes, and general regression. It also offers an automated model selection feature, which analyses the training data and selects four models to train simultaneously, evaluating which algorithms perform best in predicting defaults [10]. As IBM Watson Studio is a cloud-based service, computing costs are managed by the platform, allowing us to train models on large datasets efficiently.

In our training process, 90% of the data will be used for training, and 10% will be reserved for testing. To assess the model's prediction performance, we will use the confusion matrix, a standard tool for evaluating classification models

3.2. *Data acquisition*

Data collection will take place across various strawberry fields in Ecuador, each with distinct features. The goal is to gather a large amount of data to create a robust system capable of detecting red spider mites in diverse fields. Additionally, strawberry leaf images will be collected using web crawling techniques to automatically gather a vast number of images [11]. We will also utilize open-source platforms like Kaggle for additional data. These images will vary in pixel size, and we will determine the ideal size for model training during the process.

3.3. *Data Labeling*

To enhance the accuracy of the model, careful attention will be given to the labelling process. Agricultural engineers will manually annotate the images using the VGG Image Annotator (VIA) [12]. For each leaf image, regions with a potential red spider mite presence will be annotated, including the infestation levels (eggs, young, adults). It is important to note that during this stage, we are not confirming the presence of the pest; rather, based on expert knowledge, we are labelling areas that are likely infected. Confirmation of infestation will occur in subsequent stages. Manual annotation is a laborious process and prone to errors, especially due to fatigue, so the annotators will work in multiple sessions to mitigate this risk.

3.4. *Data Augmentation and Future Work*

To develop a robust model, we will apply data augmentation, a commonly used technique in machine learning that generates new data from existing data through transformations [13]. Using the PyTorch library, we will apply techniques such as horizontal and vertical flips, stretching, zooming, cropping, and image rotations. This method will artificially increase the size of the dataset, helping the model detect a wider range of characteristics in strawberry leaves. The entire process—from data acquisition to model training and evaluation—is outlined in Fig. 1.

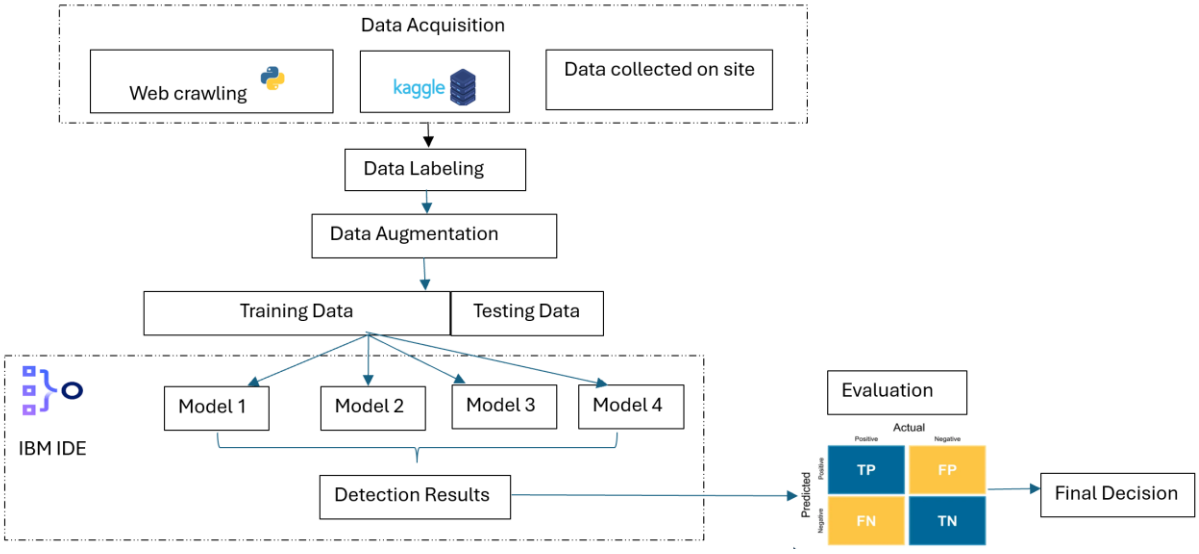

Figure 1. The overall process from data acquisition, training and evaluation with controlled data flow

Further enhancements may include refining the model with more sophisticated deep learning architectures and deploying it in real-time by integrating the model into mobile applications for immediate pest detection and timely intervention. The dataset could also be expanded by including images from other geographical regions and collected under varying conditions and seasons to improve the model's generalization capabilities. Finally, the system could be expanded to detect a range of pests and diseases beyond red spider mites, making it more versatile for broader agricultural applications.

4. Conclusion

In our study, we aim to address the critical challenge of detecting *Tetranychus urticae* in strawberry crops by applying deep learning techniques to assist farmers in pest control and pesticide application. Using IBM Watson Studio, we will train various models on a comprehensive dataset obtained from multiple sources, such as Kaggle, web crawling, and our collection of images, ultimately selecting the best-performing model. Our approach combines the expertise of agricultural professionals with advanced labelling methods to ensure accuracy in identifying red spider mite infestations across different crop fields and growth stages.

By leveraging data augmentation and manual annotations, we increase the likelihood of developing a robust model capable of adapting to diverse strawberry crop environments. Through the implementation of automated machine learning models like YOLO and PMML-based algorithms, our goal is to deliver an effective, scalable, and reliable solution for pest detection in strawberry crops. This system has the potential to assist farmers in reducing pesticide application, providing greater certainty about the presence of pests, and promoting sustainable farming practices in Ecuador.

Acknowledgements
Supported by the Department of Electrical Engineering and Mechatronics of the University of Debrecen, Faculty of Engineering. Special thanks to our supervisor, *Dr Istvan Balajti*, for permanent support, constructive criticism and continuous encouragement during our study and manuscript preparation.

5. References

[1] Agriculture, "Goverment of Western Australia," December 2022. [Online].

[2] Ali, S., Ullah, M.I., Sajjad, A., Shakeel, Q., Hussain, A. (2021). Environmental and Health Effects of Pesticide Residues. In: Inamuddin, Ahamed, M.I., Lichtfouse, E. (eds) Sustainable Agriculture Reviews 48. Sustainable Agriculture Reviews, vol 48. Springer, Cham.

[3] Bras, A., Roy, A., Heckel, D.G., Anderson, P. & Karlsson Green, K. (2022) Pesticide resistance in arthropods: Ecology matters too. *Ecology Letters*, 25, 1746–1759. Available from: https://doi.org/10.1111/ele.14030

[4] Zhou, C., Lee, W. S., Liburd, O. E., Aygun, I., Schueller, J. K., & Ampatzidis, Y. (2021). Smartphone-based Tool for Two-Spotted Spider Mite Detection in Strawberry. In *2021 ASABE Annual International Virtual Meeting* (p. 1). American Society of Agricultural and Biological Engineers.

[5] Lee, S., Arora, A. S., & Yun, C. M. (2022). Detecting strawberry diseases and pest infections in the very early stage with an ensemble deep-learning model. *Frontiers in Plant Science*, *13*, 991134.

[6] Kim, H., & Kim, D. (2023). Deep-Learning-Based Strawberry Leaf Pest Classification for Sustainable Smart Farms. *Sustainability*, *15*(10), 7931.

[7] Yamada, M., Thiesen, L. V., Iost Filho, F. H., & Yamamoto, P. T. (2024). Hyperspectral Imaging and Machine Learning: A Promising Tool for the Early Detection of Tetranychus urticae Koch Infestation in Cotton. *Agriculture*, *14*(9), 1573.

[8] Guo-Feng, Y., Yong, Y. A. N. G., Zi-Kang, H. E., Xin-Yu, Z., & Yong, H. E. (2022). A rapid, low-cost deep learning system to classify strawberry disease based on cloud service. *Journal of integrative agriculture*, *21*(2), 460-473.

[9] Magistretti, S., Dell'Era, C., & Petruzzelli, A. M. (2019). How intelligent is Watson? Enabling digital transformation through artificial intelligence. *Business Horizons*, *62*(6), 819-829.

[10] Zeckl, J., Wastian, M., Brunmeir, D., Rappelsberger, A., Arseniev, S. B., & Adlassnig, K. P. (2021). From machine learning to knowledge-based decision support—A predictive-model-markup-language-to-Arden-syntax transformer for decision trees. *Soft Computing for Biomedical Applications and Related Topics*, 89-99.

[11] Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, *13*(3).

[12] Ojha, U., Adhikari, U., & Singh, D. K. (2017, June). Image annotation using deep learning: A review. In *2017 International conference on intelligent computing and control (I2C2)* (pp. 1-5). IEEE.

[13] Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., & Shen, F. (2022). Image data augmentation for deep learning: A survey. *arXiv preprint arXiv:2204.08610*.