CONESCAPANHONDURAS2025paper53.pdf

Institute of Electrical and Electronics Engineers (IEEE)

Document Details

Submission ID

trn:oid:::14348:477773261

Submission Date

Jul 31, 2025, 11:28 PM CST

Download Date

Aug 12, 2025, 2:38 PM CST

File Name

CONESCAPANHONDURAS2025paper53.pdf

File Size

1.8 MB

5 Pages

3,207 Words

18,259 Characters

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

- **16** Not Cited or Quoted 10%
  Matches with neither in-text citation nor quotation marks
- **1** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

9% Internet sources
8% Publications
0% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **16** Not Cited or Quoted 10%
  Matches with neither in-text citation nor quotation marks
- **1** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

9% Internet sources
8% Publications
0% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	www.ijritcc.org	1%
2	Internet	www.research.lancs.ac.uk	1%
3	Internet	www.frontiersin.org	1%
4	Publication	Yanqiu Yang, Priyanka Mali, Lawrence Arthur, Faezeh Molaei, Sena Atsyo, Jiarui G...	1%
5	Internet	www.scirp.org	<1%
6	Internet	eee-conference.com	<1%
7	Internet	repositorio.upn.edu.pe	<1%
8	Internet	www.intechopen.com	<1%
9	Internet	www.sidalc.net	<1%
10	Publication	Shruti Sharma, Ashutosh Sharma, Trinh Van Chien. "The Intersection of 6G, AI/Ma...	<1%

11 Publication

William G Coon, Mattson Ogg. "Foundation Models Reveal Untapped Health Infor... <1%

12 Internet

repository.unimus.ac.id <1%

13 Internet

www.esp.org <1%

14 Internet

www.jatit.org <1%

15 Publication

"Intelligent Informatics", Springer Science and Business Media LLC, 2025 <1%

16 Publication

Anuj Kumar Singh, Sachin Kumar. "Security, Privacy, and Trust in WBANs and E-H... <1%

Detection of Anthracnose, Virosis and Citrus Tristeza Virus in Persian Lime using a Convolutional Neural Network

15 ***Abstract*—This research implements a convolutional neural network (CNN), using deep learning (DL), for disease detection in persian lime plants, a key crop for the economy of Honduras. The diseases analyzed such as anthracnose, huanglongbing and 12 citrus tristeza virus, constitute significant threats to agricultural production. Early identification of these pathologies is crucial 6 to minimize losses. Using the Roboflow 3.0 platform and an incremental methodology, the CNN was trained using a dataset of 1,500 images, applying augmentation techniques to simulate real field conditions. The model achieved a median accuracy (mAP) of 99.1 %, excelling especially in anthracnose due to its visual characteristics. Despite the similarities between huanglongbing and citrus tristeza virus, the network was able to accurately identify all thee diseases after several training cycles. The proposed system not only increases farmers ability to detect diseases in real time, but also optimizes decision making by improving precision agriculture and resource use in large-scale plantations. This contributes to protect production and reduce economic losses in the Honduran agricultural sector.**

***Index Terms*—Convolutional Neural Network, Deep Learning, Precision Agriculture.**

I. Introduction

The agricultural sector is currently facing significant challenges due to the emergence and spread of various diseases affecting crops, including Persian lemons. Diseases such as anthracnose, huanglongbing and tristeza virus can cause devastating losses in terms of both production and economic resources. Early and accurate detection of these diseases is crucial to mitigate their impact and ensure plantation sustainability. In response to this need, emerging technologies in data processing 8 and deep learning, especially convolutional neural networks (CNNs), have proven to be promising tools in the automation and optimization of crop disease diagnosis. The integration of artificial intelligence in agriculture not only facilitates more informed decision making, but also reduces maintenance costs and thus prevents the loss of productive hectares. This research 14 work proposes the use of a convolutional neural network for automatic disease detection in Persian lemon plants. Using 6 the Roboflow platform and an incremental methodology, it seeks to create a deep learning model capable of accurately identifying anthracnose, huanglongbing and tristeza virus. The objective of this research is to provide farmers with an innovative tool to help improve productivity and reduce losses caused by these diseases, thus promoting a more efficient and sustainable management of their crops.

II. Methods

A. Artificial intelligence

Jaboob et al., (2024) explains that, the term Artificial Intelligence (AI) is a research area within computer science that is dedicated to the development of systems capable of carrying out activities that usually require human intelligence. These activities include pattern recognition, decision making and natural language processing. AI ranges from basic pattern recognition systems to deep neural networks used in machine learning, which allows machines to hone their skills through experience and data fed to it. (2022) mentions that, in developing countries such as Honduras, artificial intelligence is very limited, due to several factors that hinder the advancement and modernization of the country. This is clearly perceptible when examining several regions of the country and observing that modernization through the implementation of new technologies is almost null.

B. Machine Learning

(Mahadevkar et al., 2022) mentions that Machine Learning represents a category of AI, which enables computers to reason in a manner similar to humans, based on past experiences and their subsequent expansion. This process requires reduced human intervention to examine data and recognize patterns. (Simeone, 2018) explains that, the large part of current machine learning applications is classified within supervised learning, whose purpose is to identify a pre-existing pattern between inputs and outputs. The impact of machine learning in society is wide and varied, affecting sectors such as industrial production, health, education, transportation and food, where ML and computer vision is able to locate objects, segment and classify them to follow a particular action desired by the human. (Huang et al., 2019) mentions that, Deep Learning (DL) or deep learning contemplates a set of advanced methods within machine learning, which employs deep neural networks to identify and learn complex patterns in extensive amounts of data. This technique is especially useful in activities such as image recognition, natural language processing and other applications that require automatic feature extraction from raw data. Unlike traditional machine learning (ML) methods, DL does not require manual creation of features for model input, since neural networks have the ability to learn them automatically during the training process using large data sets.

C. Convolutional neural networks and YOLO-NAS

(Yamashita et al., 2018) mentions that, the adjustment of parameters, such as kernels, is known as training. This is carried out with the aim of reducing the difference between the generated outputs and the labels, using an optimization algorithm called back propagation and gradient descent, among others. Training a network is a process that consists of identifying the kernels of the convolution layers and the weights in the fully connected layers, with the objective of minimizing the discrepancies between the generated predictions and the actual labels in a training dataset. Convolution constitutes an essential element in convolutional neural networks, in which a filter matrix or kernel is employed that moves through the input image, thus extracting the most relevant features by performing mathematical operations on various areas of the image. This procedure allows CNNs to identify spatial patterns, such as edges, textures or colors efficiently and automatically, which is crucial for classification and detection tasks(Michelucci, 2019). Terven et al., (2023) mentions that, Yolo-NAS (You Only Look Once - Neural Architecture Search) was launched in May 2023 by Deci, a company dedicated to the development of models and production tools for the creation, optimization and implementation of deep learning models. This model is oriented to the detection of small objects, optimizing localization accuracy and improving the relationship between performance and computational capacity, which makes it ideal for applications in real-time edge devices. In addition, its open source architecture is available for research purposes.

III. Experimentation

The main objective of this research is the detection and classification of diseases in Persian lemon leaves, concentrating specifically on : Anthracnose, Huanglongbing and tristeza virus. Using the quantitative incremental methodology, the ”Diseases database” is created for the neural network training process. This process extend into three stages.

A. Database images Labelling

For the creation of the database, we traveled to a specialized farm for a carefully chosen hour with near-optimal lighting conditions, although the rugged terrain made image capture an arduous task. Over the course of the visit, 1,500 images were collected to create a diverse and comprehensive database. This collection process was intended to ensure a wide variety of conditions and scenarios for further analysis, but what followed proved to be more intricate than anticipated. Although every effort was made to capture high quality images, the variability of each tree’s environment made the analysis much more complex than we originally expected. Factors such as the angle of the sun, variations in tree shape, and the different ways in which diseases manifested themselves created an unpredictable and often chaotic environment. These factors significantly affected the labeling, pattern detection and machine learning processes. For example, differentiating between diseases such as huanglongbing and tristeza virus required paying special attention to subtle visual cues that were not always obvious, especially when dealing with variations in natural lighting. Similarly, distinguishing the tiny black dots characteristic of anthracnose from shadows or dark spots caused by inconsistencies in lighting proved to be a formidable challenge. In shaded areas, these distinctions became even more difficult. The dynamic nature of the outdoor environment, combined with the inherent complexity of symptom

B. Training and test

After collecting the images and annotations for the database, the initial phase of the training and testing process was carried out using Roboflow. The steps of this process are shown in Figure 1.

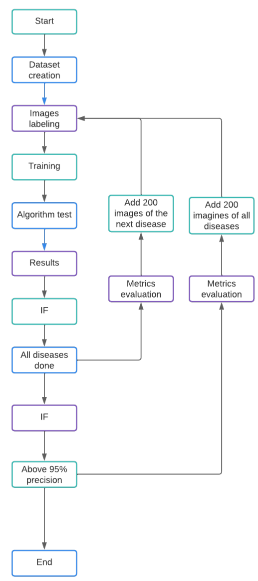

Fig. 1. Flow chart

C. Field test

To ensure optimal results in both theory and real-world applications, we conducted real-time field tests to evaluate detection performance in non-ideal environments. These tests were designed to confirm the reliability and effectiveness of the system under imperfect conditions, while also validating its optimal functionality in less controlled practical scenarios. To do this, we visited a different farm to ensure that none of the trees with which the system had been trained were used in the test phase. This ensured the impartiality of the evaluation and assessed the performance of the system with completely new and unknown data.We only used an iPhone and the raw system deployed through Roboflow for testing, which allowed us to evaluate the real-time sensing capability of the model in a practical field environment, without the need for additional equipment or post-processing.

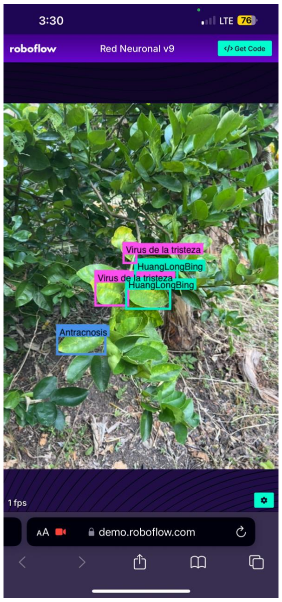

Fig. 2. Example of detections

IV. Methodology

A. Variables

The mAP measures a model´s precision in the overall system, It provides a mesure by assessing how well the system identifies and classifies objects across the dataset. To do this, its calculate the average precision values across differente recalls levels for all and each class in the database, is considered correct if the predicted box is close to the real box, the Higher the mAP, the better the model is at making acurrate preditions across all clases

B. Augmentations and data aumentations

Two system were trained in different settings,the first one was trained with 200 images in each increment of the dataset, to measure how the system´s perfomance improves as more images were added for training, this allow the system to learn more each time from a larger database.The other system incorporated 200 images in each database increment, but in addition also used data augmentations. This means that this system was trained not only with the original images, but also with those same images with various digital alterations. One of the augmentations applied was:

1) Saturation: ± 15: This adjustment generates saturated versions of the images, which improves the system's ability to tolerate and accurately identify real images that may have such imperfections. By diversifying the data set with these alterations, the model becomes more resilient to variations in image quality.

2) Blur: 0.3 pixels: This artificially generates artificially blurred images to make the system have a more difficult time trying to identify and correctly make the predictions while also helping us to diversify the dataset.

3) Noise: 1.5% pixels: This generates images with noise, these noise being with and black dots randomly generated, considering that anthracnose has this same effect on the leave, It was a great opportunity to test the system´s ability to identify real anthracnose instead of artificial imperfections.

C. Tools and Techniques

1) Excel: It is a program that is used extensively in various disciplines for the structuring, analysis and representation of data. It facilitates the users to execute mathematical calculations, the elaboration of graphics and the management of large volumes of information in an efficient way"(Rayat, 2018).

2) Roboflow: is a platform created to simplify the management and processing of image data in computer vision applications.

3) Microsoft Project: It is a software created for project management. It facilitates users in planning, scheduling, allocating resources to tasks and tracking project progress. This tool is especially valuable in the organization and supervision of complex projects as it allows the identification of potential drawbacks" (Muthusamy et al., 2023).

V. Result and Analisys

A. With augmentations

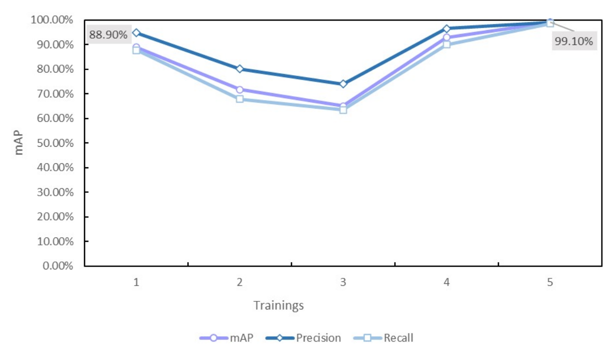

Fig. 3. Process of training

As previously mentioned, two neural networks were trained with different configurations: one with augmentations and one without. Figure X shows the behavior of the network without augmentations over five training sessions.

In the first training, the network started with promising performance, reaching high initial metrics. However, during the second increment, a decrease in performance was observed, a trend that continued in the third training. Although this decline might seem paradoxical or discouraging, it is expected behavior. The network was exposed to two new diseases, each with the same number of images, which increased the size of the total dataset but proportionally reduced the number of samples per class (disease). This phenomenon may cause the network to have difficulty generalizing effectively to the new classes, given the smaller number of examples per category.

However, this drop in performance was temporary. From the fourth training, the network began to fit the data better, achieving a substantial improvement in the fifth and final training, where the metrics exceeded 95%. These results indicate that, although the increase in data variety initially negatively affected performance, the network managed to recover and learn to differentiate between the different conditions in subsequent training. Importantly, the augmented network consistently achieved metrics above 95%, highlighting the importance of data augmentation techniques to improve model generalization.

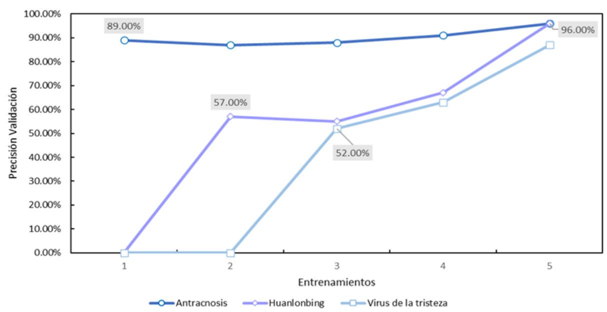

Fig. 4. performance of the accuracy of validation annotations

Figure 4 shows the behavior of the accuracy in the validation annotations during the training of the network with data augmentation techniques. It is important to highlight that, due to the methodology applied, where the diseases were introduced incrementally, both Huanglongbing and tristeza virus start with an accuracy of 0% in the first training, due to the incremental methodology used. On the other hand, the network shows a high initial performance in the identification of Anthracnose, with 89% accuracy in the first training.

The case of Anthracnose is particularly remarkable, as its accuracy remains high and relatively stable throughout all training, despite the gradual increase of the dataset, without necessarily increasing the number of images for this disease. This suggests that the network finds distinctive features more easily in Anthracnose, which facilitates its identification even when confronted with new disease classes.

In the case of Huanglongbing, accuracy starts lower, at 57% in the second training, but stabilizes and improves progressively throughout the process, reaching levels close to 96% in the last training. Finally, tristeza virus and Huanglongbing share several visual similarities, making it more difficult for the network to distinguish between the two. However, despite starting from 52%, the accuracy of tristeza virus improves significantly, reaching 96%, very close to the performance of Huanglongbing. This is remarkable, considering the challenge of differentiating these two diseases, especially since one of the main visual characteristics of tristeza virus can easily be mistaken for noise or dirt in the images. The fact that the network managed to achieve such a high level of accuracy, even at magnifications that included noise, highlights the great performance potential of this network, even in its unmagnified version.

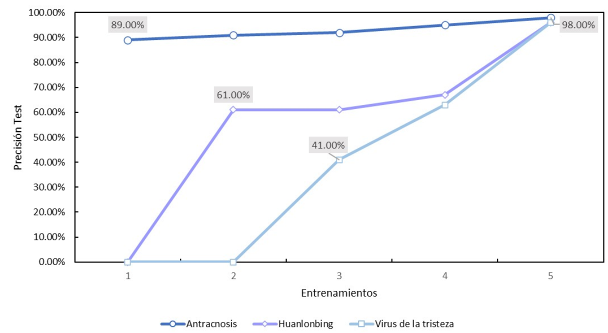

Fig. 5. performance of the accuracy of Test annotations

In Figure 5, the accuracy in the testing stage of the network with augmentations is shown. Anthracnose starts with a high accuracy (89%) and maintains a constant performance, reaching 98% in the last training, suggesting that the net easily identifies this disease. Huanglongbing starts with 61% and progressively improves to reach 98% as well, showing a rapid adaptation to increasing data. tristeza virus, which presents more visual challenges, starts with a low accuracy (41%) but improves significantly to 98%.

This demonstrates a clear trend: the network with augmentations has rapid adaptability in the incremental methodology, achieving remarkable improvements once training begins to be refined.

B. Without augmentations

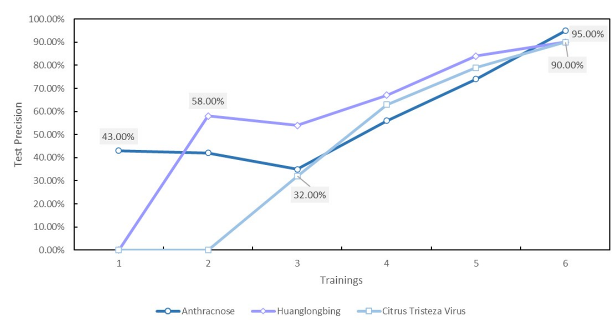

Fig. 6. Process of training Without augmentations

the augmented network had a promising start with a mAP of 88.90%, which gave us a good initial ability of the model to identify patterns in the dataset. In this network we see a huge drop, by half, suggesting that the model had difficulty learning without the additional diversity provided by the augmentations. nevertheless, the network did not maintain almost constant growth and achieved results that, although considerably lower, are still impressive.

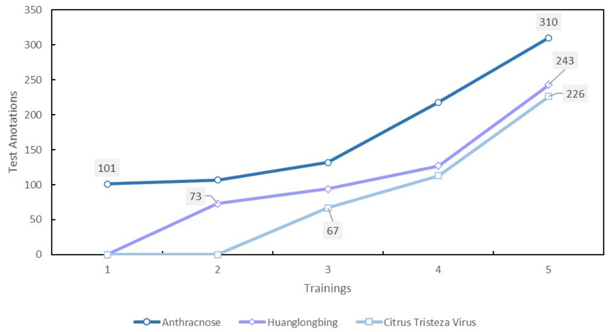

Fig. 7. Accuracy behavior in validation annotations

In the previous graph we can notice a somewhat erratic behavior in the way the network learns, anthracnose starts again with a result much lower than expected and goes down, it is not until the fourth increment, where we start again to add new images where we see a total inflection, where the three categories reach an almost homogeneous precision, anthracnose takes again the precision that was expected or that was seen in the network with augmentations and they are very close to each other. This shows that the proposed incremental methodology is having a positive influence on the final result of the network.

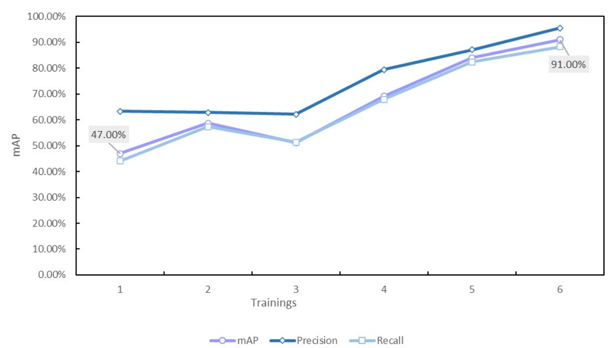

Fig. 8. Accuracy behavior in validation annotations

This last graph continues the trend mentioned above, we can see that the graphs follow a very similar rhythm among them, starting all of them below the expected and having a considerable jump from the fourth increment, which again tells us that the proposed methodology works particularly well for this type of networks.

C. Conclusions

In conclusion, the use of Roboflow 3.0 in the development of a convolutional neural network for Persian lemon disease detection has proven highly effective. The application of augmentations accelerated the learning process of the network, achieving outstanding results, with a mean Average Precision (mAP) of 99.1%, an accuracy of 98.1%, and a recall of 98.5%. This robust detection system will empower farmers by enabling early identification of critical diseases such as anthracnose, huanglongbing, and tristeza virus, ensuring better tree health and increased production. While anthracnose is detected with the highest precision due to its distinctive black spots, huanglongbing and tristeza virus initially posed challenges due to their similar visual patterns. Nonetheless, the system now detects all three diseases with high accuracy, providing a significant advantage for disease management in large-scale lemon cultivation.

References

[1] Ajra, Z., Xu, B., Dray, G., Montmain, J., & Perrey, S. (2022). Mental arithmetic task classification with convolutional neural network based on spectral-temporal features from EEG. 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 52-55. https://doi.org/10.1109/EMBC48229.2022.9870887

[2] Casaca, Á. D. (2014). El Cultivo del Limón Persa.

[3] Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. IEEE Access, 8, 75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510

[4] Donkersley, P., Silva, F. W. S., Carvalho, C. M., Al-Sadi, A. M., & Elliot, S. L. (2018). Biological, environmental and socioeconomic threats to citrus lime production. Journal of Plant Diseases and Protection, 125(4), 339-356. https://doi.org/10.1007/s41348-018-0160-x

[5] Espinal, E. (2021). Estudio de Factibilidad para la Producción de Plantas Sanas de Limón Tahití. 2021.

[6] FHIA. (2024). Informe Técnico 2023: Programa de Diversificación.

[7] Gallagher, J. (2022). What is a Confusion Matrix? A Beginner's Guide. 2022. https://blog.roboflow.com/what-is-a-confusion-matrix

[8] Gomez-Flores, W., Garza-Saldana, J. J., & Varela-Fuentes, S. E. (2022). A Huanglongbing Detection Method for Orange Trees Based on Deep Neural Networks and Transfer Learning. IEEE Access, 10, 116686-116696. https://doi.org/10.1109/ACCESS.2022.3219481

[9] Hazarika, S., Choudhury, R., Montazer, B., Medhi, S., Goswami, M. P., & Sarma, U. (2020). Detection of Citrus Tristeza Virus in Mandarin Orange Using a Custom-Developed Electronic Nose System. IEEE Transactions on Instrumentation and Measurement, 69(11), 9010-9018. https://doi.org/10.1109/TIM.2020.2997064

[10] Heras, D. (2017). Clasificador de imágenes de frutas basado en inteligencia artificial. Killkana Técnica, 1(2), 21. https://doi.org/10.26871/killkana_tecnica.v1i2.79