

# Modelado 3D de plantas de fresa por medio de segmentación semántica YOLOv7

## 3D Modeling of Strawberry Plants through Semantic Segmentation with YOLOv7

Emmanuel Bonilla Balbuena (1).  
Estudiante Tecnológico Nacional de México/I. T. de Tuxtla Gutiérrez.  
[119270548@tuxtla.tecnm.mx](mailto:119270548@tuxtla.tecnm.mx).

Angel Andrés López Espinosa (2)\*. Estudiante Tecnológico Nacional de México/I. T. de Tuxtla Gutiérrez.  
[119270568@tuxtla.tecnm.mx](mailto:119270568@tuxtla.tecnm.mx).

Néstor Antonio Morales Navarro (3), Tecnológico Nacional de México/Instituto Tecnológico de Tuxtla Gutiérrez,  
[nestor.mn@tuxtla.tecnm.mx](mailto:nestor.mn@tuxtla.tecnm.mx).

Madaín Pérez Patricio (4), Tecnológico Nacional de México/Instituto Tecnológico de Tuxtla Gutiérrez,  
[madain.pp@tuxtla.tecnm.mx](mailto:madain.pp@tuxtla.tecnm.mx).

Osvaldo Brindis Velazquez (5), Tecnológico Nacional de México/Instituto Tecnológico de Comitan,  
[osvaldo.bv@comitan.tecnm.mx](mailto:osvaldo.bv@comitan.tecnm.mx)

---

\*corresponding author.

Artículo recibido en julio 07, 2023; aceptado en agosto 16, 2023.

### Resumen.

*La aplicación de la inteligencia artificial en la agroindustria representa un campo de investigación en constante crecimiento. Se ha avanzado considerablemente en la detección de biomasa mediante algoritmos que procesan imágenes en 2D, con resultados considerablemente buenos. Sin embargo, persisten problemas en la detección de características clave, como la altura de las plantas y la densidad del follaje. En el marco de este proyecto, se implementaron técnicas de segmentación con el uso de YOLOv7 y los resultados se combinaron con datos de profundidad proporcionados por una cámara Intel RealSense D435. La propuesta se dividió en tres fases claramente definidas. En primer lugar, se inició el proceso con la captura de imágenes RGB y de profundidad. Posteriormente, el modelo YOLOv7 se encargó de segmentar las estructuras de las plantas en las imágenes RGB. Finalmente, se llevó a cabo la fusión de datos en 2D y de profundidad a través del modelo, lo que facilita la generación de una reconstrucción tridimensional integral. Este proyecto representa una aplicación técnica de gran relevancia en el ámbito de la visión por computadora y el desarrollo de modelos tridimensionales.*

**Palabras claves:** Agroindustria, Biomasa, Inteligencia Artificial, YOLOv7

### Abstract.

*The application of artificial intelligence in agroindustry represents a continually growing field of research. Considerable progress has been made in biomass detection through algorithms that process 2D images, with significantly positive results. However, challenges persist in detecting key features, such as plant height and foliage density. Within the framework of this project, segmentation techniques were implemented using YOLOv7, and the results were combined with depth data provided by an Intel RealSense D435 camera. The proposal was divided into three clearly defined phases. Firstly, the process began with the capture of RGB and depth images. Subsequently, the YOLOv7 model was responsible for segmenting plant structures in the RGB images. Finally, the fusion of 2D and depth data was performed through the model, facilitating the generation of a comprehensive three-dimensional reconstruction. This project represents a technically significant application in the field of computer vision and the development of three-dimensional models*

**Keywords:** Agroindustries, Artificial Intelligence, Biomass, YOLOv7

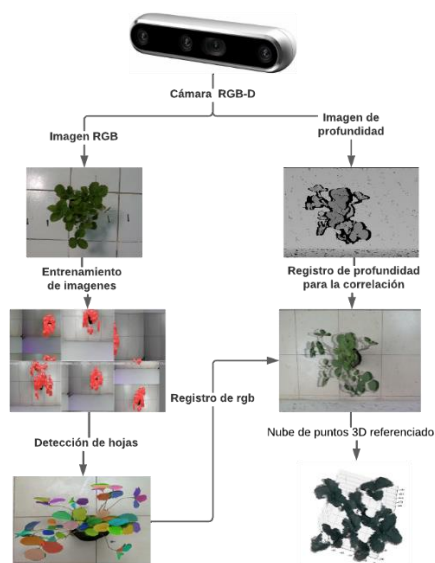
## Introducción.

La agroindustria es un componente esencial de la economía global y desempeña un papel crucial en la producción de alimentos para satisfacer las crecientes demandas de una población en constante crecimiento (Da Silva et al., s. f.). Sin embargo, esta industria se enfrenta a una serie de desafíos críticos que amenazan su sostenibilidad y la seguridad alimentaria a nivel mundial (FDA, 2021). En este contexto, el cambio climático emerge como una problemática central que afecta directamente a la agroindustria, impactando la producción agrícola y la disponibilidad de recursos naturales (Da Silva et al., s. f.).

Según las estadísticas proporcionadas por el Instituto Nacional de Estadística y Geografía (INEGI, 2022), se evidencia un aumento preocupante en la variabilidad y eventos climáticos extremos que afectan de manera significativa la producción agrícola en México y en todo el mundo (INEGI, 2019). Estos datos respaldan la urgente necesidad de abordar los problemas alimentarios relacionados con el cambio climático, como la pérdida de cultivos, la disminución de la calidad de los alimentos y la amenaza a la seguridad alimentaria (INEGI, 2019). En este contexto, la agricultura de precisión apoyada por tecnologías emergentes como la visión artificial y algoritmos de *Deep Learning*, como YOLOv7 (Y Wang, H Wang, 2022), se presenta como una solución prometedora para abordar los desafíos planteados por la agroindustria (Villota Neira, 2019). Estas tecnologías avanzadas permiten una gestión más eficiente de los recursos agrícolas, la detección temprana de problemas en los cultivos y la optimización de la producción, contribuyendo así a una mayor resiliencia y sostenibilidad de la agroindustria en un contexto de cambio climático (JPT Soto, JSS Suárez, 2019).

## Métodos.

La integración de técnicas como la visión artificial y los algoritmos de *Deep Learning* aplicados al sector de la agricultura desempeñan un papel crucial en la transformación de la agroindustria y en la búsqueda de soluciones innovadoras para los problemas alimentarios en un mundo afectado por el cambio climático (Isabel Cisternas, Ignacio Velásquez, 2020). Por lo tanto, lo que se propone en este proyecto una metodología (Figura 1) donde se utilizan como entradas las imágenes en formato RGB y de Profundidad de una planta. Para ello, se utiliza una cámara *RGB-D Intel RealSense D435*, la cual mediante las librerías de *OpenCV* y *realsense2* que ofrece *Python* se obtienen las imágenes. Posteriormente, se realiza el etiquetado de las imágenes a través de la plataforma *Roboflow*. Una vez que las imágenes están etiquetadas, se utiliza el modelo de segmentación *YOLOv7*, la cual se encarga de llevar a cabo la segmentación semántica de las plantas en las imágenes. A continuación, el resultado de la segmentación semántica la utilizamos para realizar el registro de las imágenes RGB y de Profundidad para conservar únicamente a la planta. Finalmente, obtenemos el modelo 3D.



**Figura 1.** Metodología propuesta

## Imagen RGB

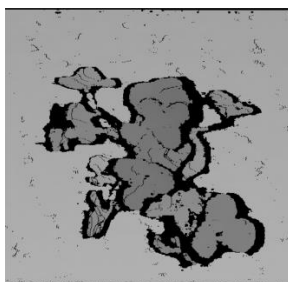
Una imagen RGB (Figura 2) por sus siglas de "*Red, Green, Blue*" (rojo, verde y azul en español), es una representación digital que se basa en la combinación de tres canales de color. Cada canal representa la intensidad de uno de estos tres colores primarios, lo que permite una amplia gama de colores al variar las intensidades en cada canal. La representación de imágenes en formato RGB es esencial en la fotografía. Cada píxel en una imagen RGB se define por sus valores en estos tres canales, lo que posibilita una representación fiel y detallada. (Pérez Barnuevo et al., 2009)



**Figura 2.** Imagen RGB

## Imagen de profundidad.

Una imagen de profundidad (Figura 3) es una representación visual que proporciona información sobre la distancia o la ubicación tridimensional de los objetos. Esta representación visual ofrece una visión de la profundidad espacial en una imagen, permitiendo a los observadores percibir la disposición de los objetos en relación con su distancia unos de otros y del punto de vista del espectador. La imagen de profundidad es una herramienta valiosa en campos como la visión por computadora, la realidad virtual y la robótica, donde la percepción de la profundidad es esencial para una comprensión precisa del entorno.



**Figura 3.** Imagen de Profundidad

## Roboflow.

Un proceso fundamental para un algoritmo de *Deep Learning* para que detecte las hojas de la planta de fresa es realizar el etiquetado manual de las imágenes. El proceso anterior, permitirá obtener los datos de entrada para el algoritmo de *YOLOv7*, y así realizar el entrenamiento para la detección automática de las hojas. En nuestro caso, utilizamos la herramienta *Roboflow* para el etiquetado, debido a que proporciona herramientas que resultan fundamentales para generar y perfeccionar el etiquetado en imágenes, un aspecto esencial en el proceso de creación de aplicaciones de visión por computadora.

## YOLOv7

*YOLO*, en sus diferentes versiones, es un enfoque ampliamente reconocido para la detección eficiente y precisa de objetos en imágenes y videos. Cada nueva versión de *YOLO*, como *YOLOv7*, representa un esfuerzo constante por mejorar el rendimiento y la precisión del modelo. La arquitectura de *YOLOv7* permite detectar objetos por medio de una imagen RGB, y en este trabajo lo utilizamos para la detección automática de las hojas de fresa. Un proceso esencial es el etiquetado para obtener información para el entrenamiento de la Red Neuronal Convolutiva que permite obtener como resultado final la segmentación semántica nos devuelve la información de la planta.

### Modelo Pinhole

El modelo del *pinhole*, o cámara estenopeica, es un enfoque básico utilizado en la reconstrucción 3D de imágenes. Funciona capturando la proyección de un objeto tridimensional en un plano bidimensional (estenopo). Para la reconstrucción 3D a partir de una serie de imágenes 2D capturadas desde diferentes posiciones, se utilizan técnicas de triangulación. Alejandro I Barranco Gutiérrez, José J Medel Juárez, (2020). Básicamente, se determina la posición tridimensional de un punto en el espacio identificando las líneas de proyección de ese punto desde dos o más vistas diferentes y encontrando el punto donde se cruzan esas líneas (David C. Lindberg, 1968).

La información en  $Z$  la obtenemos directamente de la imagen de profundidad  $I$ , donde las coordenadas  $(x, y)$  representan una coordenada en la imagen, y la constante  $k$  debe tener el valor de la distancia del objeto más lejano en la escena. Y ajustamos el modelo 3D mediante las ecuaciones siguientes:

$$Z = I(x, y) * k \text{ Ecuación 1.}$$

$$X = \left( (x - c_x) * \frac{Z}{f_x} \right) \text{ Ecuación 2.}$$

$$Y = \left( (y - c_y) * \frac{Z}{f_y} \right) \text{ Ecuación 3.}$$

Donde  $(X, Y, Z)$  son las coordenadas tridimensionales que representa al modelo 3D de la planta. El punto focal de la cámara es representado por  $f_x$  y  $f_y$ . Por último,  $c_x$  y  $c_y$  corresponde a las coordenadas del centro focal de la cámara. A continuación, definimos algunas constantes, como 'k', que utilizamos para calcular la distancia del objeto más lejano.

### Resultados

En el proceso de recrear un modelo 3D de una planta de fresa, se emplea la cámara *Intel RealSense D435* junto con un algoritmo que se realizó en *Python*. El proceso de adquisición de datos se realiza en dos etapas, la primera es la captura de imágenes RGB y la segunda se obtienen las imágenes de Profundidad.

#### Etiquetado de imágenes por medio de Roboflow

En el proceso de etiquetado en *Roboflow* se asignan etiquetas a las imágenes de las hojas de plantas de fresa (Figura 4) que hemos adquirido previamente. En este proceso, es necesario ser precisos al etiquetar cada hoja de la planta con el fin de obtener resultados que aseguren una mayor precisión en la detección automática de las hojas. Una vez que se han etiquetado todas las hojas de la planta, la herramienta de *Roboflow* generará automáticamente tres grupos de imágenes: entrenamiento, validación y prueba. El primer conjunto de imágenes proporciona los datos de entrada para entrenar el algoritmo de *YOLOv7*. El segundo grupo de imágenes permite ajustar los parámetros durante cada época del entrenamiento del algoritmo, de tal forma que se optimice la precisión del mismo. El tercer grupo de imágenes permite probar el algoritmo de *YOLOv7* una vez finalizado el entrenamiento, con esto obtenemos los datos de precisión de la detección automática. Este conjunto de datos se exportará para su uso en nuestro modelo de entrenamiento a través de *YOLO V7*.



**Figura 4.** Etiquetado de hojas

## Segmentación semántica YOLO V7

El conjunto de datos generado por *Roboflow* se importa en *Google Colab*, donde se integra con el modelo *YOLO V7*. Configuramos los parámetros de entrenamiento para un total de 600 iteraciones, lo que nos permite llevar a cabo el proceso de entrenamiento y validación del conjunto de datos que utilizaremos en esta etapa. *Google Colab* se encarga de organizar y almacenar los archivos en *Google Drive*.

Una vez que todo el proceso ha concluido, el modelo *YOLO V7* proporcionará un conjunto de datos que constituye estadísticas que validan la eficacia de nuestro entrenamiento. Los resultados de las pruebas realizadas en este trabajo muestran que la matriz de confusión (Figura 5) indica una precisión de detección del 95% con un margen de error del 5%. Este puntaje es altamente favorable y respalda nuestros objetivos de manera significativa.

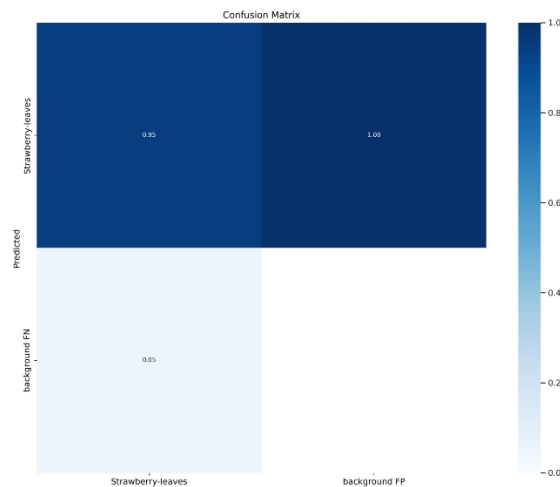


Figura 5. Matriz de confusión

De acuerdo con los resultados del *Google Colab*, se obtienen graficas (figura 6) que demuestran el comportamiento del entrenamiento estadístico de nuestras pruebas realizadas durante el proceso del entrenamiento de nuestro modelo de *Deep Learning* por medio del modelo *YOLO V7*

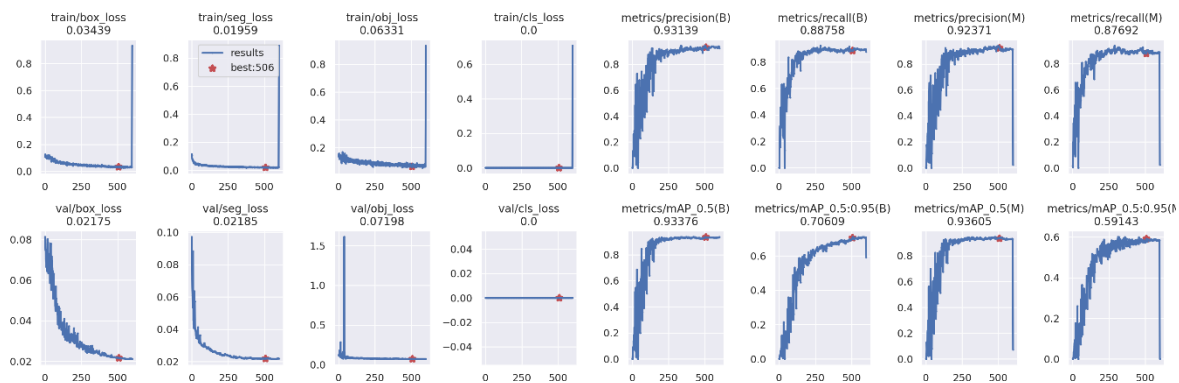


Figura 6. Graficas de Colab

### Generación del modelo 3D

El algoritmo que hemos realizado está implementado con *Python*, tiene como objetivo la visualización de datos de profundidad en un espacio tridimensional utilizando imágenes 2D. (Figura 7, Figura 9).

### Resultados 3D

En primer lugar, se procedió a tomar muestras de plantas de fresa utilizando la cámara RGB-D en combinación con el software desarrollado juntamente con la herramienta de *Python*. Estas muestras provienen de plantas cultivadas y mantenidas en un entorno controlado en un laboratorio, lo que ha permitido capturar imágenes fotográficas en RGB. Estas fotografías nos proporcionarán la información necesaria sobre las coordenadas de los objetos en los ejes X e Y.



**Figura 7.** Imagen RGB

Al incorporar imágenes de hojas de plantas de fresa al modelo *YOLOv7* y aplicar segmentación semántica, se logró la detección y delimitación precisa de las hojas. Esta combinación permitió asignar colores distintos a cada hoja, facilitando su identificación visual (Figura 8).

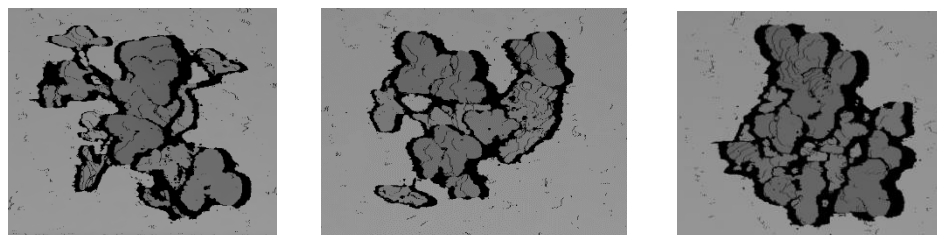
El uso de *YOLOv7* posibilitó la identificación rápida y precisa de los objetos en las imágenes, mientras que la segmentación semántica permitió asignar colores a cada región, en este caso, a las hojas de las plantas. Esto facilita un análisis detallado de la salud de las hojas, lo cual resulta relevante en aplicaciones agrícolas para el seguimiento y diagnóstico de la salud de las plantas de fresa.



**Figura 8.** Imagen YOLOv7

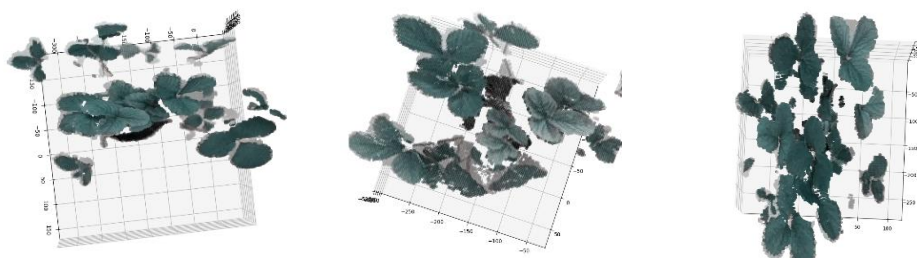


En una fase previa, se procedió a capturar muestras de imágenes de profundidad (Figura 9) de las plantas de fresa. El propósito central de esta acción era obtener información sobre la altura y la profundidad (eje Z). El objetivo detrás de esta recolección de datos era establecer una correlación entre las imágenes RGB y las imágenes de profundidad. Esta correlación se buscaba para recrear un modelo 3D, integrando la información proveniente de ambas fuentes de imágenes.



**Figura 9.** Imagen de profundidad

Esta representación tridimensional resultante puede ser valiosa en aplicaciones relacionadas con la visión por computadora y la percepción espacial, como en el campo de la robótica. Además, el algoritmo extrae los valores de color RGB del píxel correspondiente en la imagen en color y los almacena junto con las coordenadas tridimensionales. Al final de la ejecución del código, las listas 'X', 'Y', 'Z' contendrán las coordenadas tridimensionales y la lista 'rgb' contendrá los valores de color RGB de los píxeles correspondientes en la imagen de profundidad. Estos datos se pueden utilizar para visualizar el objeto en un espacio 3D, con colores que corresponden a la imagen en color original.



**Figura 10.** Modelado 3D

## Conclusiones

La obtención del modelo 3D de las plantas de fresa se llevó a cabo satisfactoriamente siguiendo la metodología propuesta. En primer lugar, se capturaron y etiquetaron con precisión las imágenes de las hojas de las plantas de fresa utilizando la cámara *Intel RealSense D435* y la plataforma *Roboflow*. Este proceso de etiquetado requirió precisión para garantizar la eficacia del modelo.

Posteriormente, se utilizó el conjunto de datos de las imágenes etiquetadas en *Google Colab* y se emplearon para entrenar un modelo *YOLO V7*, configurando parámetros de entrenamiento específicos que lograron una detección de alta precisión.

Además, se desarrolló una herramienta en *Python* que permitió la visualización de datos de profundidad en un espacio tridimensional. La segmentación semántica de las hojas de fresa, permite obtener un modelo 3D más preciso de la planta, lo que resulta de gran utilidad en campos como la visión por computadora y la robótica.

En resumen, se logró la obtención de un modelo 3D detallado de una planta de fresa, cumpliendo los objetivos de captura de datos precisos, entrenamiento de un modelo de alta calidad y representación visual efectiva en un espacio tridimensional. Estos logros respaldan diversas aplicaciones y prometen tener un impacto positivo en proyectos futuros relacionados con la percepción espacial, la representación visual de objetos, estimación de biomasa, medición de estrés, entre otras aplicaciones.

## Referencias bibliográficas.

- Barranco, A. I. (2006).** Ecuaciones de visión estereoscópica en visión por computadora. ResearchGate. [https://www.researchgate.net/publication/275044367\\_Ecuaciones\\_de\\_vision\\_estereoscopica\\_en\\_vision\\_por\\_computadora](https://www.researchgate.net/publication/275044367_Ecuaciones_de_vision_estereoscopica_en_vision_por_computadora)
- Da Silva, C. A., Baker, D., Shepherd, A. W., Jenane, C., & Da Cruz (Eds.). (2013).** Agroindustrias para el desarrollo. <http://www.lamolina.edu.pe/postgrado/pmdas/cursos/innovacion/lecturas/Adicional/20%20-%20da%20Silva%20et%20al.pdf>
- Eigen, D. (2014).** Depth map prediction from a single image using a Multi-Scale deep network. [https://proceedings.neurips.cc/paper\\_files/paper/2014/hash/7bccfde7714a1ebadf06c5f4cea752c1-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2014/hash/7bccfde7714a1ebadf06c5f4cea752c1-Abstract.html)
- Imagen depth David Eigen, Christian Puhrsch, Rob Fergus,(2014).** <https://proceedings.neurips.cc/paper/5539-depth-map-prediction-from-a-single-image->
- inegi. (2019).** El cambio climático en México [https://rde.inegi.org.mx/index.php/2015/05/10/politica-del-cambio-climatico-en-mexico-avances-obstaculos-y-retos/#:~:text=El%20cambio%20clim%C3%A1tico%20\(CC\)%20tendr%C3%A1,y%205%25%20en%20el%20verano.](https://rde.inegi.org.mx/index.php/2015/05/10/politica-del-cambio-climatico-en-mexico-avances-obstaculos-y-retos/#:~:text=El%20cambio%20clim%C3%A1tico%20(CC)%20tendr%C3%A1,y%205%25%20en%20el%20verano.)
- Juárez Ruanova, L. I., Linares Fleites, G., Sandoval Solis, M. L., & Cigarroa Alonso, K. M. (2021).** Cambio de uso de suelo y vegetación asociado a la carencia de servicios públicos y políticas públicas en Atoyatempan, Puebla. Nexa Revista Científica, 34(06), 1611–1622. <https://doi.org/10.5377/nexo.v34i06.13122>
- Lindberg, D. C. (1968).** The Theory of Pinhole Images from Antiquity to the Thirteenth Century. Archive for History of Exact Sciences, 5(2), 154–176. <http://www.jstor.org/stable/41133285>
- Pérez Barnuevo, L., Castroviejo Bolivar, R., & Berrezueta, E. (2009).** Análisis Digital de Imágenes RGB aplicado a la Geometalurgia. Puesta a punto del equipo y su potencial frente a métodos tradicionales. Archivo digital UPM. <https://oa.upm.es/11293/>

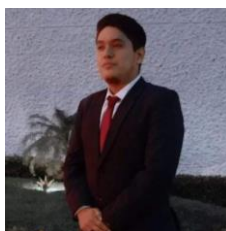


**Tovar Soto, J. P., Solórzano Suárez, J. de los S., Badillo Rodríguez, A., & Rodríguez Cainaba, G. O. (2019).** Internet de las cosas aplicado a la agricultura: estado actual. Lámpsakos (revista Descontinuada), (22), 86–105. <https://doi.org/10.21501/21454086.3253>

**Villota Neira, P. E. (2019).** Implementación de una estación prototipo con visión artificial, aplicado a la agricultura de precisión. Universidad de Cuenca. <https://core.ac.uk/download/pdf/288582212.pdf>

**Y Wang, H Wang. (2022).** Deep learning con YOLOv7 <https://ieeexplore.ieee.org/abstract/document/9994733/>

### Información de los autores.



**Emmanuel Bonilla Balbuena**, Estudiante del noveno semestre de la carrera de Ingeniería en sistemas computacionales del Instituto Tecnológico de Tuxtla Gutiérrez, trabaja en proyecto de residencia titulado “Estimación no invasiva de biomasa vegetal mediante el fenotipado de la Fragaria basado en la segmentación SuperVoxel a partir de una nube de puntos 3D”



**Angel Andrés López Espinosa**, Estudiante del noveno semestre de la carrera de Ingeniería en sistemas computacionales del Instituto Tecnológico de Tuxtla Gutiérrez, trabaja en proyecto de residencia titulado “Estimación no invasiva de biomasa vegetal mediante el fenotipado de la Fragaria basado en la segmentación SuperVoxel a partir de una nube de puntos 3D”



**Néstor Antonio Morales Navarro** es Doctor en Desarrollo Tecnológico. Docente del departamento de la carrera de Ingeniería en Sistemas Computacionales y Jefe del departamento del Centro de Cómputo del Tecnológico Nacional de México campus Tuxtla Gutiérrez. Miembro del Sistema Nacional de Investigadores, nivel Candidato. Se especializa en el área de Visión e Inteligencia Artificial.



**Osvaldo Brindis** es Ingeniero en Electrónica, Maestro en ciencias en Ingeniería Mecatrónica, por parte del Instituto Tecnológico de Tuxtla Gutiérrez. Maestro en Administración Empresarial egresado del Instituto de Estudios Superiores de Monterrey. Cuenta con la certificación de SOLIDWORKS Associate in Mechanical Design por parte de la empresa Francesa Dassault Systèmes. Reconocimiento como Microsoft Innovative Educator Expert for 2023-2024. Miembro del Colegio de Ingenieros Mecánicos y Electricistas del Estado de Chiapas. Participa en el Grupo de Investigación Turix Dynamics con presencia en México, Francia, España. Jefe de la División de Estudios de Posgrado e Investigación del Instituto Tecnológico de Comitán, donde se centra su investigación sobre el desarrollo de vehículos eléctricos para la agricultura de precisión, instrumentación y control automático.



**Madafín Pérez Patricio** es graduado como Doctor en Ciencias en Automatización y Computación Industrial en 2005, por la Universidad de Ciencias y Tecnologías de Lille, Francia. Investigador y Docente desde 1997 en el Departamento de Posgrado y la Carrera en Sistemas Computacionales del Instituto Tecnológico de Tuxtla Gutiérrez. Es miembro del Claustro Doctoral "Doctorado en Ciencias de la Ingeniería" perteneciente al Programa Nacional de Posgrados de Calidad, I.T. de Tuxtla Gutiérrez desde 2016. Miembro del Sistema Nacional de Investigadores, nivel II.