

Red neuronal convolucional para la clasificación de piezas mecánicas usando un sistema de visión artificial



Colaboración

Ibarra Bonilla Mariana Natalia; Eusebio Grande Raul; López Velázquez Julio Ivan, Tecnológico Nacional de México / Instituto Tecnológico Superior de Atlixco

Fecha de recepción: 17 de noviembre de 2023

Fecha de aceptación: 18 de diciembre de 2023

RESUMEN: El presente artículo describe el proceso de entrenamiento de una red neuronal convolucional (CNN) para la identificación de piezas mecánicas, que se utilizan comúnmente en los procesos industriales. El entrenamiento de la CNN se realizó con imágenes capturadas de una cámara Intel RealSense utilizando las herramientas y librerías de Python, Tensor Flow y Keras completando 5 épocas de entrenamiento. La prueba de evaluación se ejecutó sobre un prototipo a escala, con fines didácticos, mediante una prueba controlada que simula un ambiente industrial típico, obteniendo un porcentaje general de precisión de 90.62%. Este trabajo responde a las aplicaciones de inteligencia artificial realizadas para su aplicación en un brazo robótico colaborativo educativo.

PALABRAS CLAVE: clasificación, red neuronal convolucional, visión por computadora.

ABSTRACT: This paper describes the convolutional neural network (CNN) training process for the identification of mechanical pieces, which are commonly used in industrial processes. The CNN training was carried out with images captured from an Intel RealSense camera using the toolboxes and libraries of Python, TensorFlow and Keras completing 5 training epochs. The evaluation test was executed on a scale prototype, for didactic purposes, through a controlled test that simulates a typical industrial environment, obtaining a general accuracy percentage of 90.62%. This work responds to the artificial intelligence applications made for its future application in an educational collaborative robotic arm.

KEYWORDS: classification, convolutional neural network, computer vision.

INTRODUCCIÓN

Los robots colaborativos (CoBots) son robots que pueden trabajar de forma segura junto con trabajadores humanos. Los CoBots representan una forma de tecnología de automatización avanzada en la fabricación y se espera que se conviertan en estándar en los sistemas de producción. Tienen el potencial de transformar los procesos de fabricación y ensamblaje. Los primeros robots colaborativos

se instalaron en la industria hace poco más de diez años, pero cada vez son más los robots instalados en la industria, y hoy en día estos robots son pieza clave para el desarrollo de la industria 4.0 [1].

Hasta hace poco, los Cobots no estaban contemplados en la educación de ingeniería, sin embargo, a medida que el uso de estos aumenta en la industrial se ha presentado la necesidad de incluir en los programas educativos aplicaciones con Cobots [2, 3]. Por esto, es importante acercar a los estudiantes a las tecnologías que se manejan para aplicaciones con Cobots e interesarlos en la investigación tecnológica, de tal manera que no solo sean operadores, sino que cumplan con las características de un diseñador tecnológico.

Por esta razón se ha vuelto muy importante que desde el aprendizaje se resalte el impacto que tienen las tecnologías de la información y comunicación (TIC) y aplicaciones de inteligencia y visión artificial para estudiantes y docentes. En la búsqueda de ese objetivo, el presente trabajo presenta el desarrollo del entrenamiento de una red neuronal convolucional, conocida como CNN por sus siglas en inglés de Convolutional Neural Network, para la clasificación de piezas mecánicas: tornillos, tuercas y una banda transportadora, que corresponde a una de las típicas actividades industriales [4, 5].

Cabe mencionar que este trabajo corresponde a una de las etapas de desarrollo de un proyecto de investigación que pretende implementar aplicaciones típicas de los robots colaborativos educativos.

MATERIAL Y MÉTODOS

Red Neuronal Convolucional

Una red neuronal artificial, conocida como ANN por sus siglas en inglés, Artificial Neural Network, es un modelo computacional, inspirado en la función de las redes neuronales biológicas, el cual está compuesto por elementos de procesamiento (neuronas) y múltiples conexiones entre ellos. Las redes multicapa son capaces de aprender patrones complejos y de gran dimensión a partir de un gran número de datos, por lo que son una buena opción para el reconocimiento de imágenes, en especial las redes neuronales convolucionales, CNN, que son una red multicapa que tiene gran aplicación práctica [6, 7, 8, 9]. La arquitectura de una CNN incluye multicapas compuestas principalmente por un banco de filtros llamados kernels, una capa de convolución, una función de activación no lineal y capa de agrupación o submuestreo (pooling). Cada etapa tiene como objetivo representar las características como conjuntos de matrices llamadas mapas de características. El proceso de la operación de la convolución, la cual se incluye en al menos una de las capas de una CNN, se visualiza en la Figura 1.

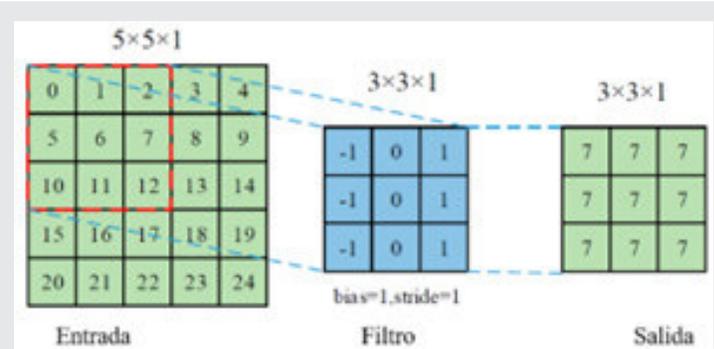


Figura 1. Proceso de la operación de la convolución.

Fuente: Recuperado de [10].

Un kernel es posicionado en la esquina superior-izquierda de la imagen, los valores de los pixeles cubiertos son multiplicados por los valores correspondientes del kernel, y los productos son sumados y el bias es sumado al final. El kernel es desplazado un pixel y el proceso se repite hasta que se filtren todas las posibles posiciones en la imagen [10].

Una vez que el banco de filtros produzca la salida se aplica una función de activación para producir los mapas de activación, donde solo las características activadas se transfieren a la siguiente capa. La función de activación más utilizada para la capa de convolución es ReLU, abreviación del inglés Rectified Linear Unit function. Esta es una función semi-rectificada cuya curva característica se presenta en la Figura 2 y se define matemáticamente como:

$$f(x) = \max(0, x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \quad \text{Ec. (1)}$$

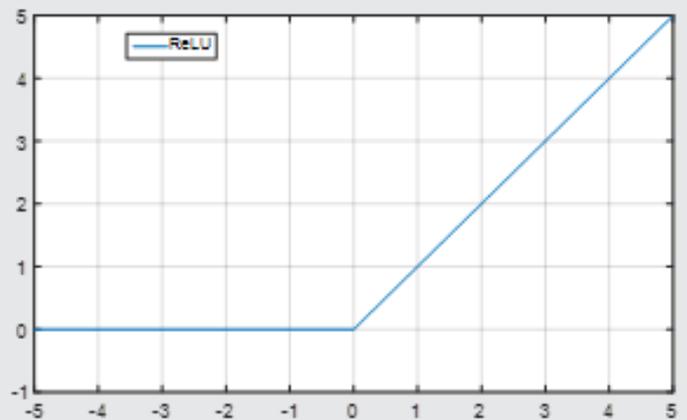


Figura 2. Representación gráfica de la función de activación ReLU.

Fuente: Elaboración propia.

La capa pooling es utilizada para reducir las dimensiones, con el objetivo de disminuir los tiempos de procesado, pero reteniendo la información más importante. La más usada es el máximo (maxpooling) o el promedio (average pooling). El maxpooling con-

siste en dividir la imagen de entrada en varias regiones rectangulares, según el tamaño del filtro, y generar el valor máximo por cada región. El average pooling realiza un proceso similar, pero la salida es el promedio de cada región. El proceso de la operación pooling se representa en la Figura 3.

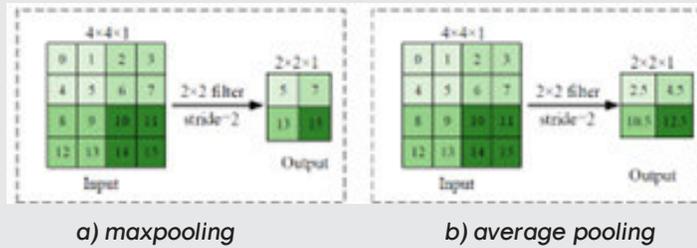


Figura 3. Proceso de la operación pooling.

Fuente: Recuperado de [10].

Arquitectura propuesta de la CNN

La arquitectura propuesta se presenta en la Figura 4 y consiste en una CNN con 4 capas convolucionales con función de activación ReLU y maxpooling entre cada capa de convolución. Las imágenes de entrada son 120x120 píxeles. Durante la primera convolución se aplican 32 filtros y con maxpooling los datos de cada filtro se dividen a la mitad quedando de 60x60, en la segunda capa de convolución se extraen 64 características y al aplicar la capa maxpooling los 64 filtros se reducen en tamaños de 30x30 cada uno, en la tercera capa de convolución se extraen 128 filtros y al aplicar maxpooling los datos de cada filtro se reducen a 15 cada uno, la última capa de convolución se aplican 256 filtros y al aplicar la capa maxpooling los datos se reducen a 7x7 en cada uno de los 256 filtros resultantes. Posteriormente, los datos se aplanan dando lugar a un arreglo unidimensional con los datos provenientes de la capa antecesora, en la siguiente capa se configuran 384 neuronas completamente conectadas con función de activación ReLU y, por último, en la última capa se configuraron 3 neuronas conectadas a una función de activación Softmax, las cuales entregarán las probabilidades de las tres clases a identificar: banda transportadora, tuerca y tornillo.

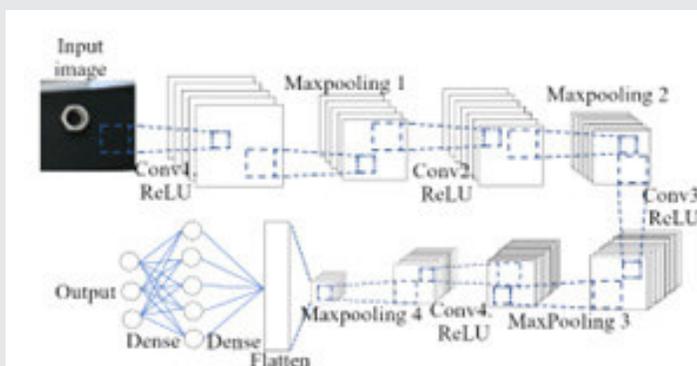


Figura 4. Arquitectura de la CNN propuesta.

Fuente: Elaboración propia.

Las imágenes de entrada con las que se entrenó el modelo de la red neuronal fueron adquiridas por medio de una cámara Intel RealSense, la cual entrega imágenes RGB 1920x1080 píxeles e imágenes de profundidad estéreo de 1280x720 píxeles, sin embargo, únicamente se utilizó un segmento de 120x120 píxeles de las imágenes RGB, el cual está centrado en la imagen original. La Figura 5 presenta un ejemplo de una imagen adquirida para el entrenamiento de la CNN, la cual se resalta en con un recuadro en color azul, en la imagen también se observa una sección de la banda transportadora y la tuerca sobre la misma. Una vez adquiridas las imágenes, estas son normalizadas a valores entre 0 y 1, dependiendo del valor del pixel original en cuyo caso podría estar entre 0 y 255, este último paso simplemente se realiza para un mejor manejo de los datos durante el entrenamiento de la red neuronal. En total se adquirieron 534 imágenes entre tornillos, tuercas y la banda transportadora, de las cuales el 80% se usaron para el entrenamiento de la red neuronal y el 20% para la validación.

Durante la etapa de entrenamiento de la CNN, de manera adicional se aplicó una función dropout o dilución para omitir aleatoriamente neuronas. La función se configuró al 50%, lo que significa que de manera aleatoria se desconectarán el 50% de las neuronas configuradas en cada batch, esto con el objetivo de evitar que la red neuronal memorice algunos caminos al reconocer los objetos y en su lugar se obliga a crear un modelo robusto que evite el sobreajuste.



Figura 5. Ejemplo de la adquisición de una imagen para el entrenamiento.

Fuente: Elaboración propia.

El entrenamiento y validación de la CNN se realiza off-line usando las herramientas y librerías de Python, Tensor Flow y Keras. Las Figuras 6 y 7 presentan las gráficas de precisión y la función de

pérdida obtenidas durante cada época de entrenamiento y validación. La precisión promedio durante las 5 épocas es de 0.889.

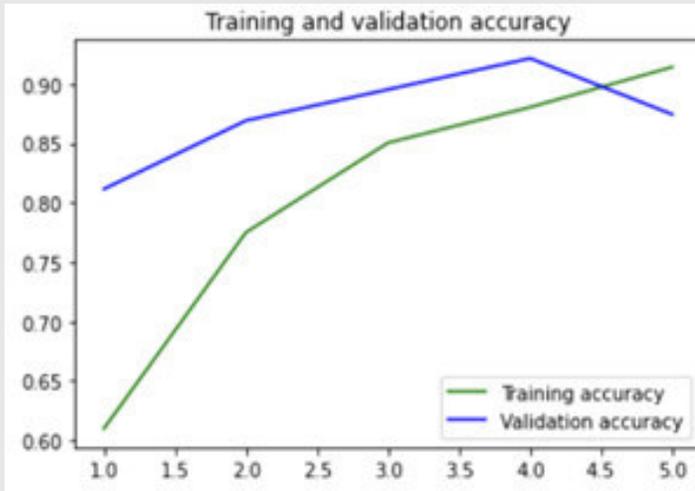


Figura 6. Gráfico de precisión obtenido durante el entrenamiento y validación en 5 épocas.

Fuente: Elaboración propia.

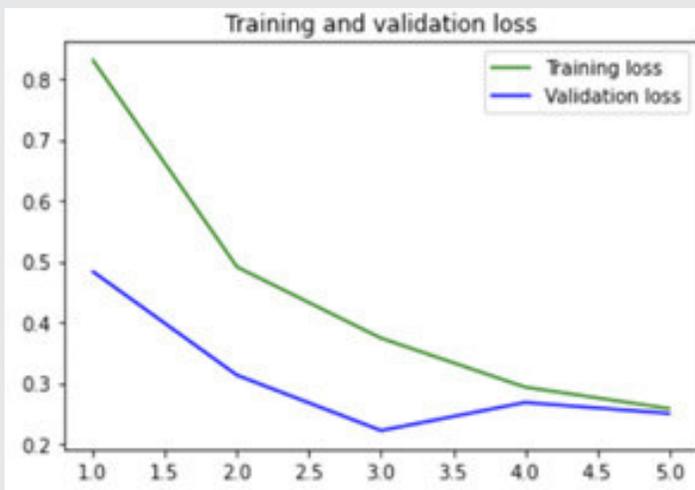


Figura 7. Gráfico de la función de pérdida obtenida durante el entrenamiento y validación en 5 épocas.

Fuente: Elaboración propia.

RESULTADOS

Para la validación de resultados se diseñó una prueba on-line dentro de un ambiente controlado del laboratorio de Mecatrónica del Instituto Tecnológico Superior de Atlixco. La prueba consistió en montar el prototipo a escala de una banda transportadora y una base metálica para el montaje fijo de la cámara Intel RealSense. La cámara se encuentra a 30 cm de altura del suelo y a 20cm de la banda transportadora, tal como muestra el esquema de diseño asistido por computadora (CAD) de la Figura 8. En la Figura 9 se visualiza el montaje real en el laboratorio del prototipo de la banda transportadora, cuyo movimiento será ejecutado por un motorreductor de corriente directa controlado por un microcontrolador PIC18F4550 embebido en una tarjeta de desarrollo Miuva [11].

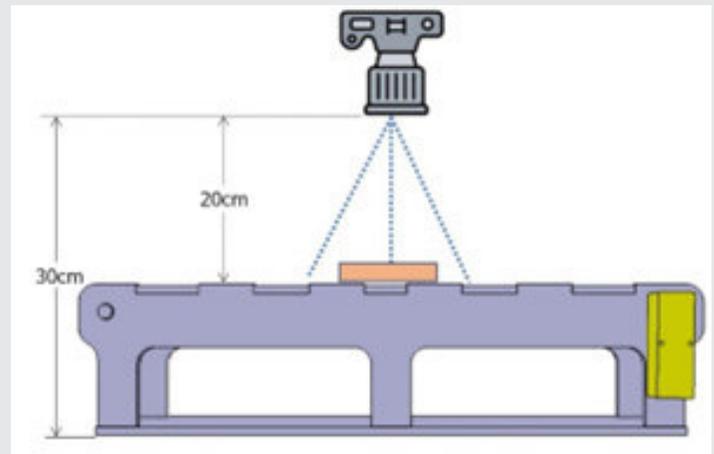


Figura 8. Esquema del prototipo de la banda transportadora y montaje de la cámara.

Fuente: Elaboración propia.

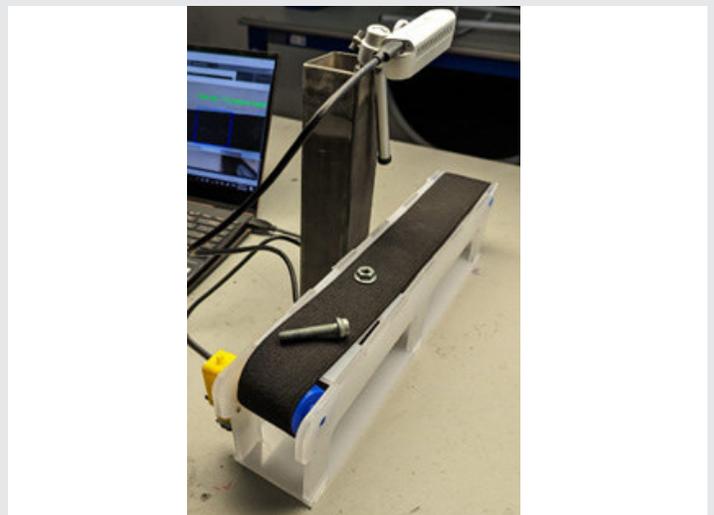


Figura 9. Montaje real del prototipo.

Fuente: Elaboración propia.

La cámara está conectada por USB a una PC que ejecuta Python, la cual a su vez se comunica por Bluetooth al microcontrolador. El microcontrolador controla el encendido de dos leds, los cuales indican el tipo de pieza identificada, luz azul para los tornillos, luz roja para tuerca y en caso de la banda transportadora ninguna luz enciende; a su vez también se conecta una pantalla LCD para visualización del resultado de la identificación de la pieza.

Entonces, la prueba para medir la capacidad de la CNN para identificar las 3 clases consistió en colocar sobre la banda los tornillos y tuercas, completando un total de 191 imágenes capturadas, con un muestreo promedio de 1.5 s. Del total de imágenes, 61 corresponden a ausencia de piezas, es decir a la banda transportadora, 69 a tornillos y 61 a tuercas. Los resultados de la clasificación se presentan en la matriz de confusión de la Tabla 1, donde la clase 0 corresponde a la banda transportadora, la 1 los tornillos y la 2 las tuercas. El porcentaje general de

precisión durante esta prueba fue del 90.62%. Las Figuras de la 10 a la 13 presentan imágenes capturadas durante la ejecución y se observa que desde la ventana de visualización en Python se imprime el tipo de pieza a la que corresponde.

Tabla 1. Matriz de confusión del porcentaje de precisión.

| Comparación de resistencias a la tracción | |
|---|--|
| Resistencia a la tracción original | Resistencia a la tracción mínimo detectado |
| Tubería: 325MPa (47ksi) | 290 (42ksi) |
| Domos: 620MPa (90ksi) | 580 (84ksi) |

Fuente. Elaboración propia.



Figura 12. Detección de un tornillo sobre la banda transportadora. Fuente: Elaboración propia.

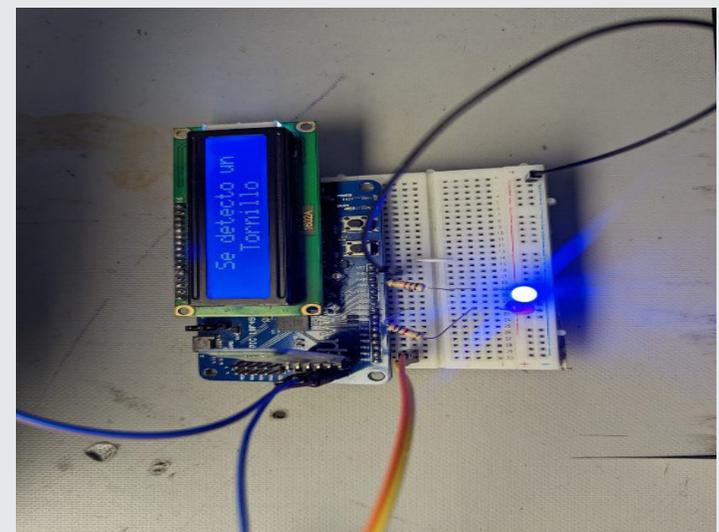


Figura 13. Tarjeta de control de la banda transportadora que muestra por medio de un display y una luz indicadora que se ha detectado un tornillo.

Fuente: Elaboración propia.



Figura 10. Detección de banda transportadora sin piezas. Fuente: Elaboración propia.



Figura 11. Detección de una tuerca sobre la banda transportadora. Fuente: Elaboración propia.

Discusión de los resultados

Los resultados obtenidos en la prueba de clasificación demuestran que la red neuronal fue capaz de identificar correctamente 53 de las 69 imágenes que presentaban un tornillo y 58 de las 61 imágenes con tuercas; mientras que las 61 imágenes con ausencia de piezas fueron identificadas en su totalidad. De acuerdo con esto, podemos calificar como satisfactorio el desempeño de la CNN. Sin embargo, como se mencionó anteriormente, esta prueba se realizó bajo un ambiente controlado con las condiciones óptimas de iluminación y con un prototipo a escala, por lo que consideramos necesario llevar estas pruebas a un entorno real, donde tanto las condiciones de iluminación como la posición de la cámara sean variables.

Con la finalidad de mejorar el porcentaje de clasificación de la CNN, actualmente se está trabajando en crear una

base de datos que contenga como mínimo 1000 imágenes que integren diferentes modelos de tornillos y tuercas, fotografiadas con diferentes fondos y luminosidad. Esto es importante, pues se pretende que la cámara se encuentre instalada en el gripper de un robot colaborativo móvil, cuyo diseño está basado en el MyCobot-Robot de la empresa Elephant Robotic [12], pero que incorpora un sistema de engranaje planetarios en las articulaciones. El esquema CAD se visualiza en la Figura 14 y la construcción del robot se encuentra actualmente en progreso.



Figura 14. Modelo CAD del robot colaborativo en SolidWorks con engranajes planetarios en las articulaciones.

Fuente: Elaboración propia.

CONCLUSIONES

Se presentó el proceso de entrenamiento y evaluación del desempeño de una red neuronal convolucional (CNN) para clasificación de piezas mecánicas. Los resultados presentados son satisfactorios, alcanzando un porcentaje del 90.62% de precisión durante la prueba de evaluación, ya que se puede considerar que el número de imágenes y épocas durante el entrenamiento fue reducido.

Si bien el proyecto aún está en etapa de desarrollo, y la CNN corresponde solo a una la etapa del sistema de visión de un robot colaborativo, nuestro objetivo es crear un ecosistema de robots en que los estudiantes puedan ejecutar aplicaciones didácticas de inteligencia artificial, que los acerque a la investigación tecnológica, y así cumplan con las características de un diseñador tecnológico, que no solo aplique, sino que seleccione la tecnología, la implemente y resuelva errores.

BIBLIOGRAFÍA

[1]. Escobar Gómez, D. A., Ibarra Jácome, O.A. (2020). "Diseño e implementación de una celda colaborativa robotizada mediante robots móviles y humanoides para clasificación de objetos". Tesis Licenciatura, Universidad de las Fuerzas Armadas, Ecuador.

[2]. P. Poór, T. Broum and J. Basl, "Role of Collaborative Robots in Industry 4.0 with Target on Education in Industrial Engineering," 2019 4th International Conference on Control, Robotics and Cybernetics (CRC), Tokyo, Japan, 2019, pp. 42-46, doi: 10.1109/CRC.2019.00018.

[3]. Ananias, Estevão; Gaspar, Pedro Dinis. A Low-Cost Collaborative Robot for Science and Education Purposes to Foster the Industry 4.0 Implementation. *Applied System Innovation*, 2022, vol. 5, no 4, p. 72.

[4]. Mosquera, Roberto Encarnación. Diseño de un algoritmo de enjambre para el trabajo colaborativo de mini robots para recoger y clasificar piezas de diferentes formas y colores. Propuesta de semillero de investigación SYNAP. En *Memorias de Congresos UTP*. 2017. p. 218-223.

[5]. Ortega Sanz, D., Gómez Muñoz, C.Q., García Márquez, F.P. (2022). Convolutional Neural Networks as a Quality Control in 4.0 Industry for Screws and Nuts. In: Suma, V., Baig, Z., Kolandapalayam Shanmugam, S., Lorenz, P. (eds) *Inventive Systems and Control. Lecture Notes in Networks and Systems*, vol 436. Springer, Singapore. https://doi.org/10.1007/978-981-19-1012-8_2.

[6]. Naranjo-Torres, José, Marco Mora, Ruber Hernández-García, Ricardo J. Barrientos, Claudio Frides, and Andres Valenzuela. 2020. "A Review of Convolutional Neural Network Applied to Fruit Image Processing", *Applied Sciences*, vol. 10, no. 10: 3443. <https://doi.org/10.3390/app10103443>

[7]. T. Guo, J. Dong, H. Li and Y. Gao, "Simple convolutional neural network on image classification," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, China, 2017, pp. 721-724, doi: 10.1109/ICBDA.2017.8078730.

[8]. Wagle, Shivali Amit, R. Harikrishnan, Sawal Hamid Md Ali, and Mohammad Faseehuddin. 2022. "Classification of Plant Leaves Using New Compact Convolutional Neural Network Models", *Plants*, vol. 11, no. 1: 24. <https://doi.org/10.3390/plants11010024>.

[9]. Gowrishankar J., et al. "Convolutional neural network classification on 2d craniofacial images". *International Journal of Grid and Distributed Computing*, 2020, vol. 13, no 1, p. 1026-1032.

[10]. LU, Jinzhu; TAN, Lijuan; JIANG, Huanyu. Review on convolutional neural network (CNN) applied to plant leaf disease classification. *Agriculture*, 2021, vol. 11, no 8, p. 707.

[11]. INTESC Electrónica y embebidos (1 de agosto de 2023). Tarjeta de desarrollo Miuva: manual de usuario. [On line] Obtenida de: <https://intesc.mx/productos/tarjeta-de-desarrollo-miuva/>.

[12]. Elephant Robotics (1 de agosto de 2023). MyCobot280 [On line] Obtenida de: <https://www.elephantrobotics.com/en/mycobot-en/>.