

**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**

**Facultad de Ingeniería en Electricidad y Computación**

Diseño e Implementación de un Sistema de Visión Artificial usando un Brazo Robótico Niryo One, Cámara RGB-D y Redes Neuronales Artificiales aplicado a la Robótica Colaborativa

**PROYECTO INTEGRADOR**

Previo la obtención del Título de:

**Ingeniera en Electrónica y Automatización Industrial**

Presentado por:

Mélany Verónica León Cuzco

GUAYAQUIL - ECUADOR

2024

## DEDICATORIA

A mi abuelo, Víctor Antonio Cuzco Sagbay, luz de mi alma e inspiración. A mi mamá mi apoyo incondicional y ejemplo de valentía.

## **AGRADECIMIENTOS**

Agradezco infinitamente a mi tutor Ph.D. Edwin Valarezo Añazco por su orientación y respaldo durante el desarrollo de este proyecto. Un especial agradecimiento al Ph.D. Ángel Domingo Sappa por inspirarme con su vocación y e invaluable contribución a la ciencia. Al Ing. Carlos Jiménez Farfán por ser mi amigo y gran apoyo en este proceso. A todos los profesores y los compañeros que se volvieron familia con el paso de los años.

## DECLARACIÓN EXPRESA

“Los derechos de titularidad y explotación, me corresponde conforme al reglamento de propiedad intelectual de la institución; Mélang Verónica León Cuzco doy mi consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual”

*Mélang León Cuzco*

---

Mélang Verónica León  
Cuzco

# EVALUADORES

.....  
**Ph.D. Efrén Herrera Muentes**

Profesor De La Materia

.....  
**Ph.D. Edwin Valarezo Añezco**

Profesor Tutor

## RESUMEN

La robótica colaborativa aplica la robótica en entornos compartidos con la finalidad de asistir a usuarios en tareas particulares. Estos robots colaboran directamente con usuarios garantizando la integridad de ambas partes con sistemas de seguridad que integran Inteligencia Artificial. La inclusión de los cobots en las industrias es limitada por las grandes inversiones que su instalación, puesta en marcha y mantenimiento implican.

En este trabajo de tesis se diseñó e implementó un sistema colaborativo que incorpora dos modelos de inteligencia artificial (IA) a un brazo robótico Niryo One, un gripper con 2 DoF, cámara RGB-D Intel RealSense y una tarjeta Jetson Nano para procesar los datos de los sensores y generar el control del hardware. El primer modelo basado en Procesamiento del Lenguaje Natural (PLN) reconoce órdenes ingresadas por voz. El segundo, es un modelo de visión artificial basado en redes neuronales artificiales que identifica y localiza los cuatro objetos incluyendo un cubo, cilindro, estrella y pirámide.

Los experimentos realizados determinan que el módulo de voz se inicializa en 0,256 segundos e infiere los comandos con una tasa de éxito del 96% en 0,53 ms. El módulo de visión es inicializado en 47 s e identifica y localiza los objetos de un frame en 3 segundos con exactitud promedio del 94%. Finalmente, el módulo control presenta 1cm de error de navegación, la tasa de éxito para las tareas de agarre y entrega al usuario son de 98,5% 90% respectivamente.

**Palabras Clave:** Robótica Colaborativa, Inteligencia artificial, Visión Artificial, Deep Learning, Niryo One.

## ABSTRACT

Collaborative robots (Cobots) focuses on implementing robots in common environments with human users and other robots. Cobots must avoid potential accidents between human and themselves by implementing safety systems based on Artificial Intelligence (AI). Also, the AI enhances the autonomy of the robots, allowing them to be deployed in less controlled scenarios. The inclusion of cobots in industries is limited due to the investments involved in buying, installation, and maintenance.

In this thesis work, a collaborative system was designed and implemented, integrating two models of Artificial Intelligence (AI) along with a Niryo One robotic arm, a gripper with 2 DoF, an Intel RealSense RGB-D camera, and a Jetson Nano. The first model, based on Natural Language Processing (NLP) recognizes voice-input commands. The second model is an artificial vision-based on Deep Learning to identify and locate four objects including a cube, cylinder, star, and pyramid.

The conducted experiments determined that the voice module initializes in 0.256 seconds and the inference time is 0.53 ms, the success rate of recognition is 96%. The vision module initializes in 47 seconds and identifies the objects within a frame in 3 seconds with an average accuracy of 94%. Finally, the control module demonstrates a navigation error of 1cm, with success rates of 98.5% for gripping and 90% for user delivery tasks.

**Keywords:** Collaborative Robotics, Artificial Intelligence, Computer Vision, Deep Learning, Niryo One.

# ÍNDICE GENERAL

EVALUADORES.....	5
RESUMEN.....	I
ABSTRACT.....	II
ÍNDICE GENERAL.....	III
ABREVIATURAS.....	V
SIMBOLOGÍA.....	VI
ÍNDICE DE FIGURAS.....	VII
ÍNDICE DE TABLAS.....	VIII
CAPÍTULO 1.....	9
1. INTRODUCCIÓN.....	10
1.1 Descripción del problema.....	11
1.2 Justificación del problema.....	12
1.3 Objetivos.....	13
1.3.1 Objetivo General.....	13
1.3.2 Objetivos Específicos.....	13
1.4 Marco teórico.....	14
1.4.1 Robótica Colaborativa.....	14
1.4.2 Inteligencia Artificial.....	15
1.4.3 Matriz de transformación.....	17
1.4.4 Hardware.....	17
1.5 Contribución.....	20
CAPÍTULO 2.....	22
2. METODOLOGÍA.....	23
2.1 Planteamiento del sistema.....	23
2.2 Diseño de solución.....	26



2.2.1	Módulo de Voz .....	28
2.2.2	Módulo de Visión.....	29
2.2.3	Procesamiento de Datos .....	32
2.2.4	Módulo Control.....	35
CAPÍTULO 3.....		37
3.	RESULTADOS Y ANÁLISIS .....	38
3.1	Módulo de Voz.....	38
3.2	Módulo de Visión .....	39
3.3	Módulo Motor.....	42
CAPÍTULO 4.....		46
4.	CONCLUSIONES Y RECOMENDACIONES.....	47
	Conclusiones .....	47
	Recomendaciones .....	49
BIBLIOGRAFÍA.....		50
5.	Bibliografía .....	51

## ABREVIATURAS

<b>ESPOL</b>	Escuela Superior Politécnica del Litoral
<b>DoF</b>	Degrees of freedom
<b>PLN</b>	Procesamiento del Lenguaje Natural
<b>NLP</b>	Natural Language Processing
<b>IA</b>	Inteligencia Artificial
<b>AI</b>	Artificial Intelligence
<b>IFR</b>	Federación Internacional de Robótica
<b>RNA</b>	Redes Neuronales Artificiales
<b>YOLOV8</b>	You Only Look Once Versión 8
<b>HRC</b>	Colaboración Humano-Robot
<b>TAM</b>	Modelo de Aceptación Tecnológica
<b>TPB</b>	Teoría del Comportamiento Planeado
<b>UTAUT</b>	Teoría Unificada de Aceptación y Uso de la Tecnología
<b>ASR</b>	Reconocimiento Automático de Voz
<b>CLN</b>	Comprensión del Lenguaje Natural

## SIMBOLOGÍA

<b>m</b>	Metro
<b>cm</b>	Centímetro
<b>mm</b>	Milímetro
<b>mm<sup>2</sup></b>	Milímetro cuadrado
<b>kg</b>	Kilogramo
<b>gr</b>	Gramo
<b>GHz</b>	Gigahercio
<b>MHz</b>	Megahercio
<b>GB/s</b>	Gigabit por segundo

# ÍNDICE DE FIGURAS

Figura 1.1 Brazo Robótico Niryo One .....	18
Figura 1.2 Jetson Nano Developer Kit [43] .....	18
Figura 1.3 Intel® RealSense™ Depth Camera D435i [49].....	19
Figura 2.4 Diagrama de flujo de la metodología .....	23
Figura 2.5 Máxima extensión del brazo robótico.....	24
Figura 2.6. Actuadores para el brazo robótico Niryo One [41] .....	26
Figura 2.7 Diagrama de dimensiones del Adaptive Gripper [41].....	27
Figura 2.8 Intents - Consola Rhino – Picovoice [53] .....	29
Figura 2.9 Arreglo de inferencia - Consola Rhino – Picovoice [53].....	29
Figura 2.10 Etiquetado de las figuras en Roboflow [54].....	30
Figura 2.11 Ejemplo del funcionamiento del modelo de visión artificial [54] .....	32
Figura 2.12 Referencia de la cámara [55].....	33
Figura 2.13 Referencia del brazo robótico Niryo One .....	33
Figura 2.14 Adaptive gripper con sticker rojo para su localización .....	34
Figura 2.15 Ubicación del hardware .....	35
Figura 2.16 Integración de los módulos propuestos .....	36
Figura 3.17 Resultado – Modelo de Voz.....	39
Figura 3.18 Matriz de confusión - Modelo de Visión Artificial.....	40
Figura 3.19 Resultados – Modulo de Visión Artificial .....	42
Figura 3.20 Resultados – Agarre – Modulo Control .....	44
Figura 3.21. Reubicación – Modulo Control.....	45

## ÍNDICE DE TABLAS

Tabla 1.1: Características Técnicas del Procesamiento de la Jetson Nano [44] .....	18
Tabla 1.2 Costos de los componentes.....	21
Tabla 2.3 Criterios de diseño del sistema .....	25
Tabla 2.4 Especificaciones de los actuadores [41] .....	26
Tabla 2.5 Posiciones fijas propuestas.....	27
Tabla 2.6 Posiciones predeterminadas desde la base del robot.....	28
Tabla 2.7 Especificaciones del dataset para el modelo de voz.....	30
Tabla 2.8 Detalles del dataset [54].....	31
Tabla 3.9 Resultado del modelo de visión artificial .....	41
Tabla 3.10 Error promedio en centímetros de la cinemática inversa .....	43

# **CAPÍTULO 1**

## 1. INTRODUCCIÓN

En la actualidad, la colaboración entre máquinas y seres humanos ha asumido un papel central en el ámbito de la industria 4.0. Los avances tecnológicos han permitido que los robots desempeñen un papel cada vez más importante en la ejecución de tareas repetitivas con una precisión que supera a las capacidades humanas.

Este contexto de colaboración máquina-humano plantea la necesidad imperante de desarrollar sistemas de robótica colaborativa que no solo sean altamente eficientes en sus funciones, sino que también sean seguros y capaces de interactuar con los humanos de una manera natural y efectiva.

Para facilitar la interacción los robots deben tener la capacidad de recibir órdenes de los usuarios a través de módulos Procesamiento del Lenguaje Natural (PLN), e.g., reconocimiento de voz. Además, debe reconocer, detectar y localizar objetos que los usuarios requieran usando visión artificial. Esto permite que el robot sea capaz de identificar y manipular objetos de manera autónoma, lo que amplía considerablemente su versatilidad y utilidad en una variedad de aplicaciones.

La aplicación de esta tecnología no se limita únicamente al ámbito industrial, sino que se extiende a la mejora de la calidad de vida de las personas en una amplia gama de contextos. Desde la rehabilitación física de pacientes a través de robots terapéuticos, hasta sistemas de asistencia y monitorización para personas de la tercera edad, la automatización de procesos impulsa soluciones innovadoras y efectivas en diversos campos.

En este proyecto integrador se busca implementar un sistema colaborativo robot-humano por medio de la integración de modelos de inteligencia artificial para visión y procesamiento de voz para sujetar objetos y entregarlos al usuario. La interacción entre el brazo robótico y el usuario inicia a través de un micrófono y en respuesta el robot entrega el objeto solicitado.

## 1.1 Descripción del problema

La interacción máquina-humano, se realiza en entornos diseñados para aprovechar las capacidades de las máquinas y los colaboradores, los humanos se encargan de tareas que requieren un nivel cognitivo más complejo como la toma de decisiones éticas, empáticas y creativas mientras que los robots realizan tareas repetitivas con alta precisión, optimizando la capacidad y desempeño de un trabajo realizado solo por humanos.

Dentro de los entornos colaborativos la seguridad del usuario es primordial, especialmente cuando se desarrollan tareas en las que los robots realizan manipulaciones aplicando una fuerza que ocasione un daño físico a los colaboradores, por esto, se procuran máquinas con adaptabilidad y flexibilidad de instalación.

Para dotar a los robots de habilidades como la capacidad de respuesta en diferentes entornos, se combina la Robótica con la Inteligencia Artificial (IA). De forma que al impulsar robots colaborativos con IA estos se pueden adaptar a situaciones cambiantes, mejorando significativamente la escalabilidad en entornos de trabajo variados, la precisión de los procedimientos y optimización de recursos.

Según el informe de la Federación Internacional de Robótica (IFR) del 2017 [1], para la implementación de cobots impulsados por IA las industrias realizan un análisis detallado de costos y beneficios antes de incurrir en una inversión de este tipo. Puesto que, dependiendo del tipo de industria, robot y alcance del proyecto, los costos pueden variar significativamente. Además, ya que incluyen la compra de hardware y software, parches de integración a sistemas ya implementados, capacitación de personal, mantenimiento y otros gastos asociados [1]. Los Robots Colaborativos (i.e., cobots) están demostrando ser eficientes en todo tipo de entornos de fabricación y logística, con precios que oscilan entre \$15 000 y \$35 000 [2].

La sinergia de la Robótica y la IA aporta grandes beneficios y a medida que la tecnología avance se espera que los costos de implementación de la



automatización de las máquinas sean más asequibles. Mientras tanto, se aprovecharán los recursos existentes y se asumirán los desafíos tecnológicos presentes en el desarrollo de sistemas embebidos que mantengan una relación costo-beneficio equilibrada, evitando necesitar grandes inversiones.

## **1.2 Justificación del problema**

Los algoritmos basados en Redes Neuronales Artificiales (RNA) han demostrado resultados excepcionales en investigaciones relacionadas con PLN y visión robótica o de máquina [3] [4] [5].

PLN es una técnica que permite convertir el lenguaje natural en una representación interna adecuada para ser procesada por un computador [6]. El aporte de estos algoritmos se evidencia en áreas como la medicina en el campo de la salud mental [7], asesoramiento remoto, predicción de epidemias y en el mejoramiento de los esquemas terapéuticos [8]. Mientras que dentro de la robótica colaborativa impulsa robots colaborativos basados en visión y sistemas de interpretación de voz [9].

En visión robótica las RNA permiten identificar y localizar una gran variedad de objetos que pueden estar distribuidos de forma aleatoria en una mesa de trabajo, facilitando su escalamiento en diferentes ambientes y aplicaciones como robots terapéuticos que sirven para la rehabilitación física [10] y cognitiva [11] de pacientes con aplicaciones como el desarrollo de sistemas de asistencia y monitorización para personas de la tercera edad y en general en la automatización de procesos que pueden abordar desafíos en diferentes campos y mejorar la calidad de vida de las personas.

## 1.3 Objetivos

### 1.3.1 Objetivo General

Desarrollar un sistema colaborativo robot-humano que contenga un módulo de reconocimiento de voz basado en un micrófono y PLN usando el modelo Rhino de Picovoice, un módulo de visión artificial utilizando una cámara RGB-D y redes neuronales artificiales; además de aplicar Cinemática Inversa para el control de los movimientos del robot. Los modelos de IA se programarán en Python y en general con herramientas *opensource*, estarán desplegados dentro de un sistema embebido (Jetson Nano) lo que obliga a que se optimicen los recursos de cómputo.

### 1.3.2 Objetivos Específicos

- Construir una base de datos para el entrenamiento de los algoritmos de inteligencia artificial utilizados en el módulo de voz y visión mediante la recopilación de datos directamente del entorno de trabajo y de internet procurando el entrenamiento de modelos de inteligencia artificial con exactitud superior a 90%.
- Desarrollar un módulo de reconocimiento de voz impulsado por un modelo basado en PLN que es compatible con las características de cómputo de la Jetson Nano y que servirá como información para el control del robot.
- Desarrollar un módulo de visión artificial usando una cámara RGB-D y un modelo de visión entrenado con YOLOV8 para la detección y localización de objetos ubicados aleatoriamente.
- Realizar el control del brazo robótico por medio de la integración del hardware y software propuesto para la manipulación de los objetos, ejecutando todos los algoritmos y módulos localmente en una Jetson Nano.

## 1.4 Marco teórico

### 1.4.1 Robótica Colaborativa

La robótica colaborativa plantea la cooperación directa y segura entre operadores humanos y robots como solución a los rápidos cambios tecnológicos de la industria 4.0 [12]. Los robots colaborativos (cobots) integran dispositivos de seguridad y tecnologías de interacción intuitiva [13] que permiten su operación en entornos sin barreras físicas. Los sistemas de Colaboración Humano-Robot (HRC) aumentan la productividad de los procesos reduciendo los riesgos relacionados con el trabajo que se realiza [14].

La producción académica referente a la exploración de la interacción humano-robot muestra un incremento significativo desde el 2016 hasta en el 2022, año en el que se establecen temáticas emergentes como la háptica (simulación del sentido del tacto), la inteligencia artificial industrial, realidad virtual, entre otras [15].

La tendencia en la automatización muestra que las empresas manufactureras buscan mejorar su fabricación y capacidad competitiva. En la salud se usa la tecnología y los cobots para asistir a personas con limitaciones psicocognitivas o de movilidad para mejorar su independencia y mantenerse activos [2].

El desarrollo tecnológico indudablemente afecta la dinámica laboral tradicional, la favorece creando nuevas fuentes de trabajo especializado e inevitablemente la reducción de trabajos repetitivos (i.e., trabajos tradicionales). Una de las barreras sociales de los cobots es que reducen las plazas de trabajos tradicionales de las personas en líneas de fabricación provocando 'desempleo tecnológico' lo que podría acentuar la desigualdad de oportunidades; sin embargo, la evolución de la tecnología se asocia también con nuevas formas y oportunidades de trabajo que compensarían esta posible amenaza de impacto negativo a la sociedad [16].

En 'Vivir con robots. Reflexiones éticas, jurídicas, sociales y culturales', Mario Toboso Martín et al. [17] se mencionan reflexiones éticas, jurídicas, sociales y culturales en las que se abordan las responsabilidades y consideraciones que se deben tener en cada ámbito mencionado, para la correcta regulación de privacidad,

protección de datos y posibles violaciones del ordenamiento jurídico en la integración real y efectiva de los ciudadanos en la robotización y el empleo.

### **1.4.2 Inteligencia Artificial**

La Inteligencia Artificial (IA), es una rama de la informática que a través de algoritmos permite a las máquinas aprender a partir de datos a operar con cierto nivel de autonomía en tareas específicas [6]. En la robótica colaborativa surge la necesidad de construir sistemas autónomos capaces de tener cierto nivel cognitivo para interpretar el ambiente de trabajo [18], por ejemplo, módulos de PLN y Visión Robótica.

PLN es la rama de la Inteligencia Artificial que experimenta con algoritmos para procesar e interpretar información verbal o escrita, i.e., letras o sonidos del lenguaje [19]. Además, permite enlazar el lenguaje humano y las computadoras para su interacción.

La Visión Robótica refiere a la Visión por Computador [20] aplicada en la Robótica, desarrolla algoritmos para que los robots puedan a través de cámaras percibir e interpretar su ambiente circundante; permitiéndoles efectuar tareas de mayor complejidad como la navegación autónoma, la clasificación, detección y seguimiento de objetos [21].

La IA con la ayuda de las Teorías de Aceptación de Tecnología como la utilidad del Modelo de Aceptación Tecnológica (TAM), la Teoría del Comportamiento Planeado (TPB), la Teoría Unificada de Aceptación y Uso de la Tecnología (UTAUT) [22] ha ganado una amplia aceptación en una variedad de campos. Los cuales incluyen asistencia virtual, atención médica y seguridad [23], aplicaciones de PLN como reconocimiento de voz y el reconocimiento de gestos faciales [24], el ingreso de comandos por voz y la detección de objetos [25].

#### **1.4.2.1 Módulo de Voz**

Picovoice es una plataforma orientada al desarrollo de modelos de voz personalizados, se especializa en el uso de motores de reconocimiento de voz

que son entrenados y adaptados a tareas específicas. Desde el navegador web se pueden entrenar modelos usando Python y JavaScript [26].

Para realizar el ingreso de comandos por voz (*speech to intent*), Picovoice cuenta con Rhino que fusiona el Reconocimiento Automático de Voz (por sus siglas en inglés ASR) y la Comprensión del Lenguaje Natural (CLN) prometiendo una eficiencia mayor al 97%, inmediatas y funcionar directamente en el dispositivo en que se instala la librería, sin tener la necesidad de conectarse a una nube [27].

Los modelos desarrollados en Picovoice deben ser compilados con conexión a internet al menos la primera vez, para que los pesos y librerías se descarguen, permitiendo su uso offline.

Para garantizar la precisión del modelo se debe considerar variabilidad en la base de datos, por ejemplo, pronunciación de personas nativas, no nativas, diferente género (tono de voz) y con diferentes acentos [28]. Picovoice-Rhino previamente se ha usado en la automatización de una estación de autoservicio [29] y de viviendas [30].

#### **1.4.2.2 Módulo de Visión**

Roboflow es una plataforma que facilita conjuntos de datos y herramientas para la creación de modelos de aprendizaje autónomo. Se utiliza para entrenar modelos de detección, segmentación o clasificación de objetos como Yolov8 [31]. En el contexto de este proyecto, se utiliza Roboflow para la recopilación de datos de imágenes (i.e., crear la base de datos) de bloques de uso didáctico y anotar sus características por medio de bounding boxes, con lo que se entrenó un modelo de detección de bloques (cilindro, estrella, cubo y pirámide).

YOLO (You Only Look Once), es un sistema de detección de objetos en tiempo real basado en redes neuronales convolucionales, con aplicaciones como monitoreo por video y conducción autónoma [32].

YOLOv8, la última versión de la serie YOLO, destaca por su alta velocidad de procesamiento, precisión y capacidad para detectar múltiples objetos en una sola

imagen [33] durante la realización de tareas relacionadas con la Visión Artificial como la detección, segmentación, estimación de pose, clasificación y rastreo [34]. Este sistema de visión artificial junto con la Robótica Colaborativa incursiona en áreas como la Biomedicina [35], Agricultura [36] y la Automatización de Vehículos [37].

### 1.4.3 Matriz de transformación

Dentro de la Robótica, las matrices de transformación son la herramienta con la que se describen los cambios de posición y orientación de objetos en el espacio [38]. Una matriz de transformación homogénea es de dimensión 4x4 y se conforma de 4 sub-matrices: una matriz de rotación 3x3, el vector de translación 3x1, la transformación de perspectiva 1x3 y el escalado global 1x1 [39], tal como se muestra en la Ecuación 1.1

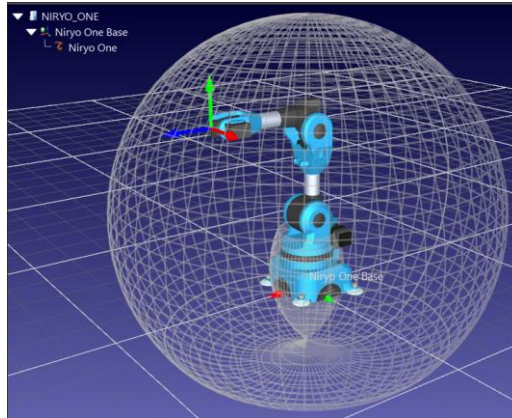
$$T = \begin{bmatrix} R_{3 \times 3} & P_{3 \times 1} \\ f_{1 \times 3} & w_{1 \times 1} \end{bmatrix} = \begin{bmatrix} \text{Rotación} & \text{Translación} \\ \text{Perspectiva} & \text{Escalado} \end{bmatrix} \quad (1.1)$$

Dependiendo de la naturaleza del ejercicio, las matrices de transformación pueden post-multiplicarse o pre-multiplicarse, representar únicamente rotaciones o translaciones, así como todo un algoritmo que constituya varias rotaciones y translaciones [40].

### 1.4.4 Hardware

#### 1.4.4.1 Niryo One

El Niryo One es un brazo robótico educativo, controlado por una Raspberry Pi 3B, de 6 grados de libertad que pesa 3,3 kg y tiene un alcance de 440 mm [41]. Figura 1.1 muestra el robot Niryo y su área de alcance en el software RobotDK.



**Figura 1.1 Brazo Robótico Niryo One**  
*Realizado con RobotDK*

Este manipulador robótico se ha implementado en múltiples trabajos de investigación debido a que es de código abierto. El Niryo cuenta con su propio software de simulación asequible para usarse en entornos educativos [42] .

#### 1.4.4.2 Jetson Nano

La Jetson Nano de NVIDIA® es una minicomputadora diseñada para su aplicación en IA [43]. La Figura 1.2 muestra el sistema embebido que compone el Jetson Nano Developer Kit.



**Figura 1.2 Jetson Nano Developer Kit [43]**

**Tabla 1.1: Características Técnicas del Procesamiento de la Jetson Nano [44]**

CARACTERÍSTICA	DESCRIPCIÓN
<b>CPU</b>	64-bit Quad-core ARM A57 @ 1.43GHz
<b>GPU</b>	128-core NVIDIA Maxwell @ 921MHz
<b>Memoria</b>	4GB 64-bit LPDDR4 @ 1600MHz   25.6 GB/s

La Tabla 1.1 muestra las principales características de la minicomputadora [44]. Su compatibilidad con softwares como TensorFlow, PyTorch y Caffe la convierten en una herramienta ideal para el desarrollo de proyectos de IA de baja potencia y bajo costo [45].

Este sistema embebido muestra gran precisión en la evaluación y rastreo de imágenes en su implementación en proyectos de identificación para detectar gestos faciales de somnolencia [46], postura del cuerpo [47], especies de plantas [48], entre otros.

#### 1.4.4.3 Cámara Intel Realsense D435i



**Figura 1.3 Intel® RealSense™ Depth Camera D435i [49]**

La cámara Intel Realsense D435i mostrada en la Figura 1.3 es una cámara de profundidad diseñada para además de capturar imágenes también capturar información por medio de sus sensores infrarrojos y RGB [49].

Esta herramienta se usa para múltiples propósitos según sus funciones, en la navegación robótica [50] gracias a su amplio rango de visión, así como para la captura de entornos de trabajo gracias a que puede realizar tareas de mapeo, sincronización y calibración básica de las imágenes de la cámara [51].



## 1.5 Contribución

El desarrollo de este proyecto de titulación se centra en implementar un robot colaborativo impulsado por inteligencia artificial usando un módulo de PLN para interactuar con el usuario, un módulo de visión robótica para percibir las posiciones aleatorias de los objetos y generar los parámetros para sujetar y mover el objeto requerido por el usuario. Finalmente, un módulo control que planifica la ruta de acuerdo con el brazo robótico usando cinemática inversa.

Tomando en cuenta la limitación que representan las grandes inversiones en las que se incurren con este tipo de proyectos. Este proyecto se enfoca en aprovechar las capacidades de componentes de bajo costo, sin comprometer la calidad ni la inteligencia que el sistema requiere. Así pues, la correcta elección tanto del hardware como del software que componen este proyecto permiten superar la barrera económica.

El módulo de reconocimiento de voz se compone de un modelo de reconocimiento de voz entrenado bajo un concepto de PLN llamado Speech to Intent. Permite al sistema, mediante un micrófono recibir órdenes específicas de un usuario, tales como “dame el cubo”.

El módulo de visión artificial consta de una cámara RGB-D y de un modelo entrenado mediante algoritmos de IA basados en redes neuronales artificiales. Esto sirve para: reconocer, detectar y localizar los objetos que el usuario requiera. La efectividad del modelo de visión artificial se constata con la matriz de confusión y las gráficas del entrenamiento. Además, los datos obtenidos por medio de la cámara y el modelo se usan en la generación de los parámetros de manipulación del objeto demandado por el usuario.

El módulo control está conformado por un brazo robótico Niryo One y una Jetson Nano. El brazo robótico Niryo One. El brazo robótico recibe comandos desde una Jetson Nano, un sistema embebido específico para IA y compatible con la cámara RGB-D.

La Jetson Nano además de ejecutar los módulos de voz y visión también aloja el procesamiento del módulo de cinemática inversa. La cinemática inversa se realiza entregando al robot las coordenadas específicas a las que se desea el actuador del robot llegue. Esta terna antes de ser enviada al robot pasará primero por una multiplicación de matrices de transformación con el fin de pasar del sistema de referencia de la cámara a la referencia del actuador del brazo robótico.

Los costos del sistema colaborativo implementado se detallan en la Tabla 1.2 La inversión aproximada es de \$ 3 110. Esta cifra, en comparación con el rango de inversión señalado en la problemática, es significativamente menor, lo que sugiere mayor probabilidad de adquisición por parte de los usuarios.

**Tabla 1.2 Costos de los componentes**

<b>COMPONENTE</b>	<b>COSTO</b>
Niryo One	\$ 2.500
Intel Realsense RGB-D 435i	\$ 350
Jetson Nano	\$ 150
Micrófono	\$ 10
Propiedad intelectual	\$ 100
<b>TOTAL</b>	<b>\$ 3.110</b>

## **CAPÍTULO 2**

## 2. METODOLOGÍA

En este capítulo se detallan los criterios en base a los que se ha diseñado el sistema de Visión Artificial aplicado a la Robótica Colaborativa como solución a la problemática expuesta en el Capítulo 1. La propuesta se diseña e implementa por medio de módulos; cada uno de ellos implica algoritmos, funciones y hardware específico para la función que cumple y su interacción con el cliente. En la Figura 2.4 se muestra un diagrama el flujo con el que este capítulo se desarrolla.

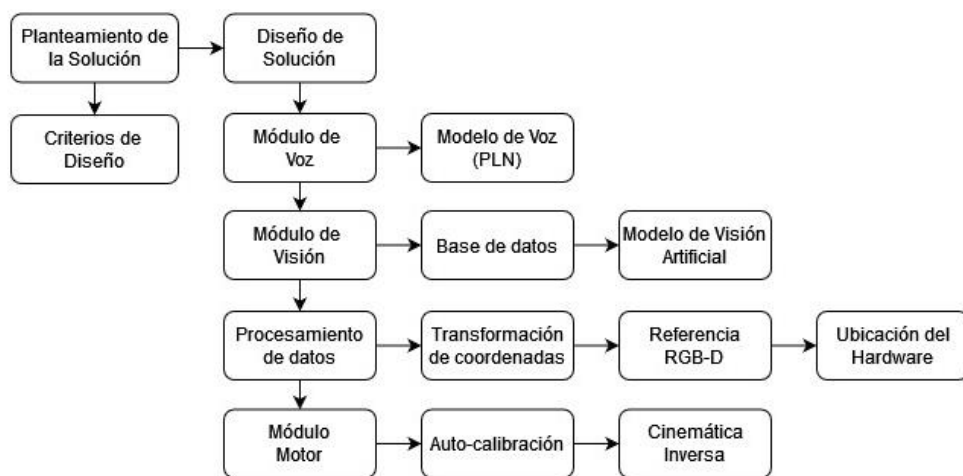


Figura 2.4 Diagrama de flujo de la metodología

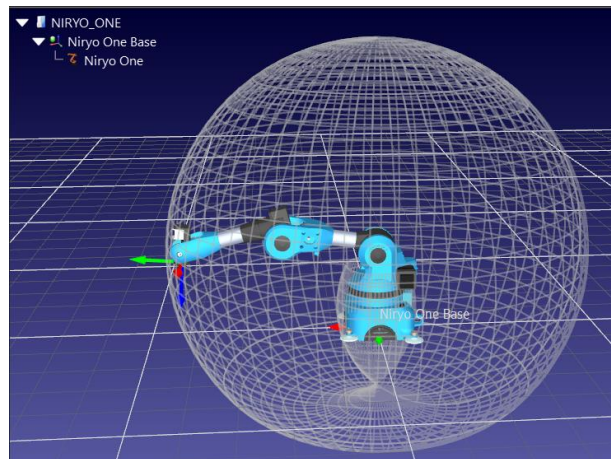
### 2.1 Planteamiento del sistema

Para satisfacer los requerimientos del cliente se enlistaron varios criterios que deberán ser tomados en consideración durante el desarrollo de la solución a la problemática planteada en el primer capítulo.

Tal como muestra la Tabla 2.3 el sistema será desarrollado para su escala según las necesidades del ámbito del mercado o industria en el que se quiera aplicar. Su escalabilidad depende directamente de la base de datos con el que los diferentes modelos de inteligencia artificial aplicados al proyecto se entrenen.

Para asegurar su adaptabilidad, se deberá seleccionar un actuador adecuado de forma que se pruebe que el brazo robótico tiene la capacidad de adaptarse a las

diferentes estructuras y pesos de los objetos. El proyecto se diseñará en torno al uso de figuras básicas, de color llano y con superficies lisas, como lo son: un cubo, un cilindro, una pirámide y una estrella. Estas cuatro formas representan utensilios que el brazo robótico podría manipular, por ejemplo: una taza (cilindro) y una caja de medicamento (cubo).



**Figura 2.5 Máxima extensión del brazo robótico**

*Realizado con RobotDK*

El brazo robótico presenta una zona de trabajo esférica, con un alcance máximo de 440 mm desde la base, tal como se muestra en la Figura 2.5 El área de trabajo total del robot brinda la oportunidad de establecer una superficie de recolección de mínimo  $200 \times 200 \text{ mm}^2$ . La posición home/inicio del robot y la posición de entrega serán predeterminadas en una coordenada que resulte cómoda y accesible para el cliente.

Mientras el objetivo se encuentre dentro de la zona de trabajo, el sistema una vez lo haya identificado y localizado tendrá la capacidad de crear una ruta óptima para desplazarse hasta su ubicación. Se espera que el brazo robótico realice todos sus movimientos de forma controlada y a una velocidad prudente, con la finalidad de procurar la estabilidad del sistema y la seguridad tanto de los usuarios como de los componentes involucrados. Por las características del Niryo One, detalladas en el marco teórico, este sistema colaborativo podrá manipular objetos de un peso menor o igual a 227 gr.

Dentro de los criterios de diseño también se señala el mantenimiento del hardware que será responsabilidad del desarrollador y del usuario según las especificaciones y sugerencias que se detallan en la solución. Estas acciones permitirán el funcionamiento integral del sistema.

**Tabla 2.3 Criterios de diseño del sistema**

<b>Criterio</b>	<b>Procedimientos</b>
Escalabilidad	El sistema robótico será capaz de alcanzar los diferentes objetos y atender a ordenes dependiendo de los modelos de visión artificial y reconocimiento de voz respectivamente.
Adaptabilidad	El actuador del brazo robótico tendrá la capacidad de adaptarse eficientemente a las formas de los objetos.
Área de trabajo	El brazo robótico deberá tener un área de recolección de mínimo 200x200 mm <sup>2</sup> . El área de entrega deberá ser pensado para la comodidad del usuario y podrá ser personalizado.
Desplazamiento	El brazo robótico tendrá la capacidad de crear su propia ruta para llegar al punto de recolección indicado.
Estabilidad	En todas las manipulaciones que el sistema realice, el brazo robótico se moverá con cadencia, control y a una velocidad prudente para la posterior interacción con los usuarios.
Robustez	El robot tendrá la capacidad de alzar hasta 227 gr.
Mantenimiento	El sistema contará con procesos que mantendrán calibrados los procesos de visión y transformación de coordenadas, además de calibrar el brazo robótico de forma automática cada vez que sea necesario.
Interacción con el usuario	El sistema colaborativo interactuará con el usuario a través de un micrófono para el ingreso de las órdenes y la posición de entrega para la asistencia del cliente.

## 2.2 Diseño de solución

Considerando los criterios de diseño expuestos en la Tabla 2.3 Para construir la propuesta final, se revisarán las opciones disponibles de actuadores del brazo robótico, de forma que se elija el que mayor adaptabilidad le otorgue al sistema. También se inspeccionará cuáles y cuantas posiciones predeterminadas se necesitarán y sus razones.



**Figura 2.6. Actuadores para el brazo robótico Niryo One [41]**

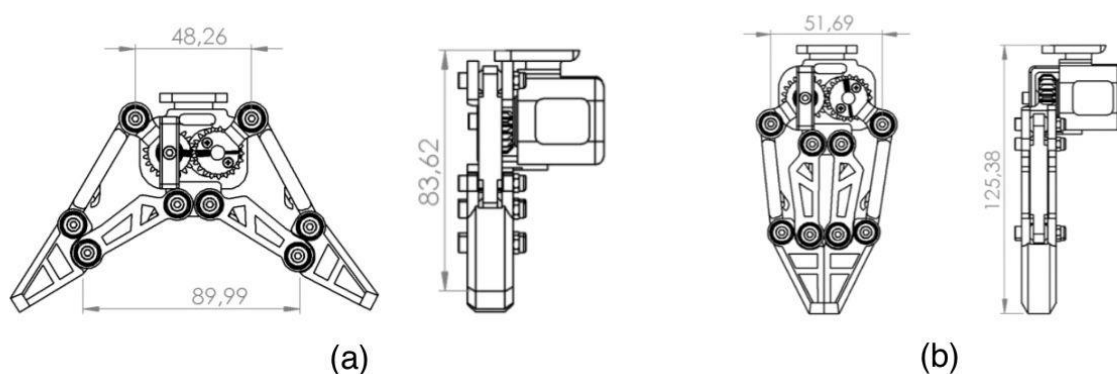
En la Figura 2.6 se presentan todos los actuadores de agarre que se pueden adaptar al brazo robótico Niryo One y en la Tabla 2.4 sus especificaciones. El Gripper 1 y Gripper 2 realizan la función de abrir y cerrar su pinza logrando un ancho de apertura máxima de 27 mm y 60 mm respectivamente. Con esto, se descarta al primero porque limitaría el agarre del sistema a objetos de menos de 27 mm de ancho.

El Gripper 3 se conforma por varias bisagras y una división en cada una de sus extensiones que le otorgan adaptabilidad en el agarre, replicando el funcionamiento de unas falanges. El también llamado Adaptive Gripper, por ser el que mayor apertura y distancia de agarre logra, será el actuador integrado al sistema propuesto.

**Tabla 2.4 Especificaciones de los actuadores [41]**

<b>Especificación / Actuador</b>	<b>Gripper 1</b>	<b>Gripper 2</b>	<b>Gripper 3</b>
<b>Longitud (gripper cerrado)</b>	80 mm	120 mm	125 mm
<b>Ancho de apertura máxima</b>	27 mm	60 mm	90 mm
<b>Distancia de agarre desde la base del actuador</b>	60 mm	80 mm	85 mm

En la Figura 2.7 se presentan los diagramas dimensionales del gripper seleccionado y serán tomados en cuenta para la correcta determinación de las posiciones fijas del sistema que se detallan en la Tabla 2.5.



**Figura 2.7 Diagrama de dimensiones del Adaptive Gripper [41]**

*Nota: Vista frontal y lateral del gripper cuando está abierto (a) y cuando está cerrado (b).*

Ya con el Gripper definido, se proponen cuatro posiciones fijas para el proceso de agarre y entrega realizados por el brazo robótico. Las posiciones Inicio/Home y puntos intermedios conceden estabilidad al sistema, puesto que permitirán al robot moverse con precisión y cadencia. Los puntos intermedios sirven como ‘descanso’ durante las manipulaciones del brazo robótico, procurando la seguridad tanto del sistema como de los usuarios.

**Tabla 2.5 Posiciones fijas propuestas**

<b>Nombre</b>	<b>Razón</b>
<b>Inicio/Home</b>	Mantener una posición fija durante la inicialización del sistema y mientras se realizan otros procesos que no implican el movimiento del brazo robótico.
<b>Puntos Intermedios</b>	Como parte del mantenimiento preventivo del sistema, se evita que le brazo robótico realice movimientos bruscos que lo desestabilicen o descalibren.
<b>Entrega</b>	Servirá para determinar una posición adecuada y accesible para diferentes, estaturas, edades y condiciones que limiten el movimiento del usuario.



La posición de entrega del sistema debe ser apto para su interacción con usuarios de diferentes estaturas, edades y sobre todo accesible para individuos con movilidad reducida. Basándonos en la Norma Ecuatoriana de la Construcción – Accesibilidad Universal (AU) del 2019 [52] la altura máxima para el accionamiento, alcance o sujeción de objetos es de 1 200 mm. Respetando los requerimientos del cliente y las normas de construcción del país, se determina que la distancia desde el piso hasta la posición de entrega del sistema sea de 1 150 mm.

En la Tabla 2.6 se detallan las coordenadas de las posiciones predeterminadas para el sistema.

**Tabla 2.6 Posiciones predeterminadas desde la base del robot**

<b>Nombre</b>	<b>Posición Propuesta (m)</b>
<b>Inicio/Home</b>	(0,075; -0,003; 0,164)
<b>Punto Intermedio 1</b>	(0,197; -0,009; 0,234)
<b>Punto Intermedio 2</b>	(0,200; 0,009; 0,234)
<b>Entrega</b>	(0,027; -0,232; 0,427)

Nota: Las coordenadas se referencian en el punto medio de la base del robot.

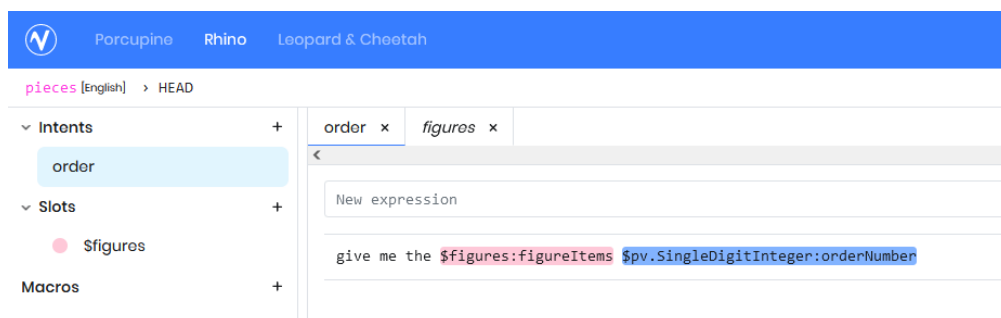
De acuerdo con la Figura 2.4, la solución propuesta se construirá a partir de los diferentes módulos explicados a continuación.

### **2.2.1 Módulo de Voz**

Desarrollado bajo el concepto de Speech to Intent de PLN, el módulo de voz se conforma por un modelo de IA de reconocimiento de voz entrenado en la consola en línea Rhino de Picovoice. La consola permite el ingreso del contexto en el que el modelo se usará por medio de sus apartados llamados ‘Intents’, ‘Slots’ y ‘Macros’ que permiten definir las acciones específicas a las que la orden hablada se asocia, los objetos o palabras que el modelo debe identificar y ordenes globales respectivamente.

Tal como se distingue en la Figura 2.8, se ha definido la orden ‘give me the’, luego se especifica que grupo de objetos se desea el modelo identifique con:

\$figures:figureItems y finalmente se pide que a cada figura se le otorgue un número con: \$pv.SingleDigitInteger:orderNumber. Con esta programación nos aseguramos de que el modelo entregue un arreglo con los datos correspondientes a cada forma tal como se muestra en la Figura 2.9.



**Figura 2.8 Intents - Consola Rhino – Picovoice [53]**

*Nota: En el apartado de Slots se define un grupo llamado 'figures' y en él se incluyen los nombres en inglés de cada una de las figuras.*

```
Inference clear  
  
{  
  intent: "order",  
  slots: {  
    figureItems: "star"  
    orderNumber: "2"  
  }  
}
```

**Figura 2.9 Arreglo de inferencia - Consola Rhino – Picovoice [53]**

## 2.2.2 Módulo de Visión

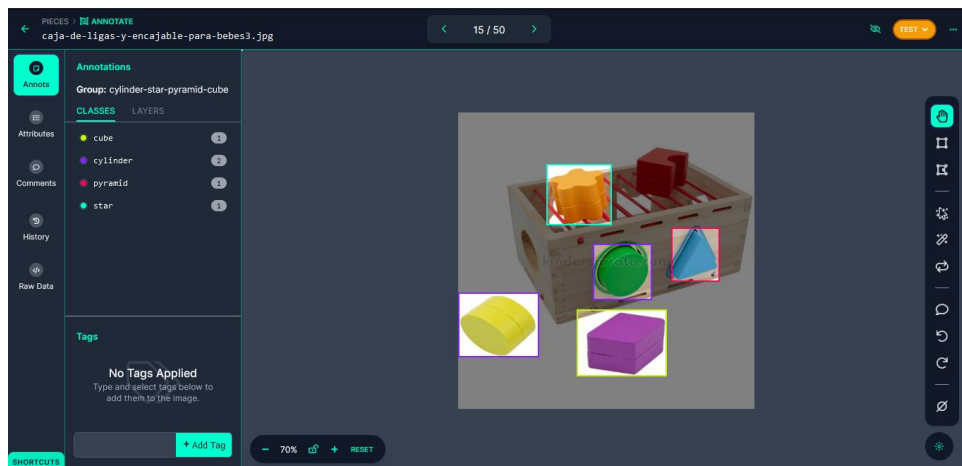
El módulo de visión está conformado por un modelo de visión artificial generado a partir de un dataset propio y del retorno de profundidad que la cámara RGB-D entrega.

El dataset fue creado exclusivamente para este proyecto y contiene imágenes en diferentes resoluciones, tamaños y colores de las cuatro figuras representativas del sistema. Como se puede corroborar en la Tabla 2.7, el conjunto de datos está conformado por 735 imágenes en las que se pudo identificar cada figura.

**Tabla 2.7 Especificaciones del dataset para el modelo de voz**

	Dataset puro	Dataset con augmentation
<b>Cubo</b>	735	1,170
<b>Cilindro</b>	735	1,170
<b>Pirámide</b>	735	1,170
<b>Estrella</b>	735	1,170
<b>TOTAL</b>	1,687	4,680

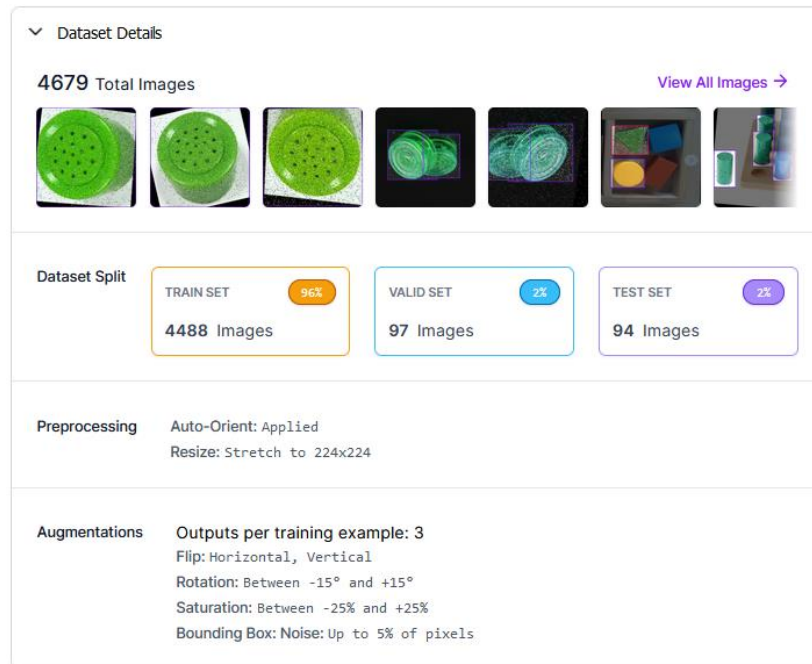
Usando las herramientas de Roboflow, a cada imagen, como la mostrada en la Figura 2.10, se le incluyeron los bounding boxes necesarios para la identificación de cada figura.



**Figura 2.10 Etiquetado de las figuras en Roboflow [54]**

Además de las 1 687 imágenes que se insertaron y etiquetaron manualmente, en la Tabla 2.8 se puede muestran las herramientas usadas para el aumento de los datos. Con variaciones como voltear las imágenes en diferentes orientaciones, rotaciones, modificaciones de los colores y el aumento del ruido dentro de los recuadros de etiquetado el data base llegó a las 4 679 imágenes tal como se muestra en la Tabla 2.7.

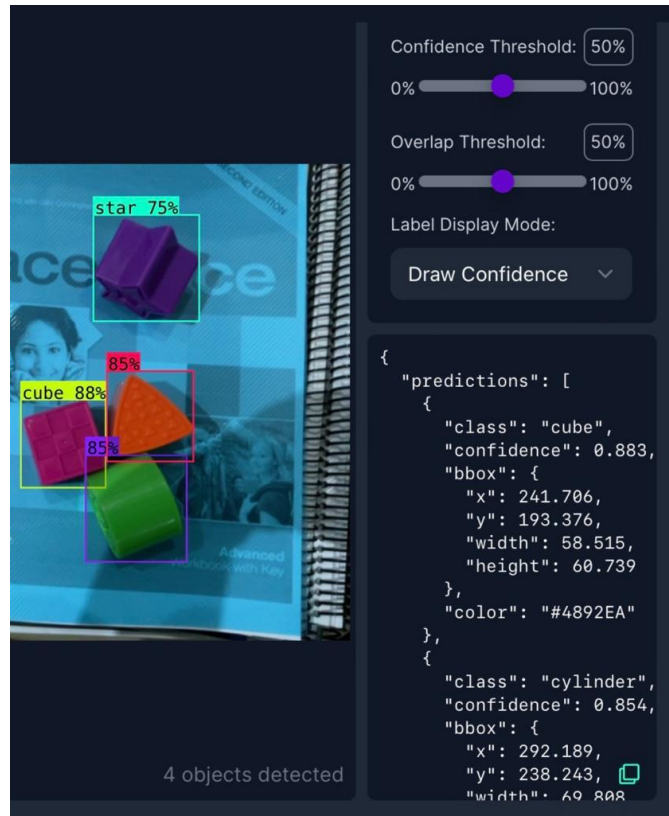
**Tabla 2.8 Detalles del dataset [54]**



Se creó un database balanceado con un 96% de datos para el entrenamiento un 2% para la validación y 2% para el conjunto de prueba. La división de los datos en los conjuntos de entrenamiento, validación y prueba sirvieron para la generación del modelo de visión artificial.

El modelo de visión artificial se generó con Yolov8 con el dataset antes explicado y en un total de 40 épocas. Se exporta el modelo en formato .pt tomando en cuenta que será implementado en una Jetson Nano que trabaja con Ubuntu 16.4, Python 3.6.9 y un Nvidia-Jetpack 4.6.1.

Yolov8 para identificar la figura genera un bounding box que se crea a partir de información en dos dimensiones (x, y) y el ancho y el alto del cuadrado. En la Figura 2.11 se muestra como el modelo es capaz de identificar con precisión las formas y en la parte inferior derecha se puede distinguir el arreglo que devuelve con la información que se usó para la localización en 2D de los objetos en el espacio de trabajo.

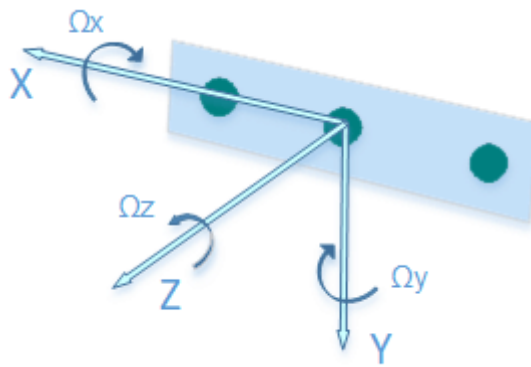


**Figura 2.11 Ejemplo del funcionamiento del modelo de visión artificial [54]**

Apoyándonos de las cualidades de la cámara Intel Realsense RGB-D 435i seleccionada para la implementación de esta solución, por medio de programación se obtiene la profundidad a la que los objetos identificados se encuentran. La profundidad calculada por la cámara también llamada ‘depth’ es entregada como un arreglo de pixeles sin procesar, sin embargo; usando la librería ‘pyrealsense2’ propia del sistema de la cámara el depth se convierte a milímetros internamente y nosotros podemos trabajar directamente con el valor.

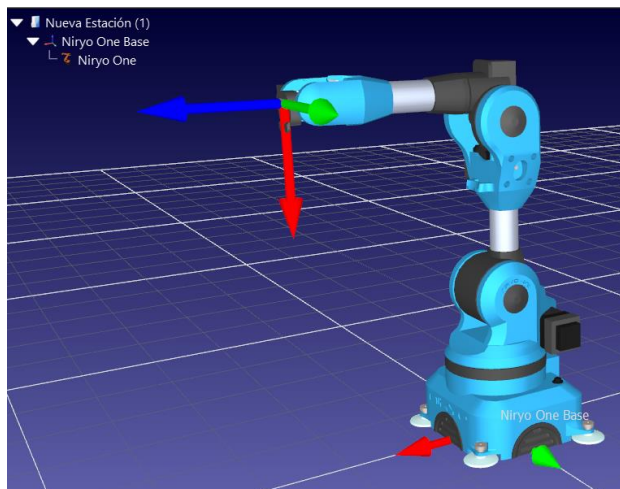
### 2.2.3 Procesamiento de Datos

El procesamiento de datos inicia con la integración del módulo de voz y el módulo de visión. De los arreglos entregados por cada uno de los modelos de inteligencia artificial se recopilan los datos necesarios para la localización en 3D de los objetos y además se realiza la alineación de datos en este caso por figura. Con esta conjunción se tiene a disposición la triada para la localización en 3D del objeto que el cliente demande a través del micrófono. Esta coordenada toma como punto de referencia el centro de la cámara como se muestra en la Figura 2.12.



**Figura 2.12 Referencia de la cámara [55]**

Si bien la cámara localiza el objeto, el brazo robótico realiza la recolección y entrega de este; por esta razón la ubicación espacial del objetivo debe referenciarse en el robot. El origen o punto de referencia del brazo robótico Niryo One en este proyecto se ubica en el gripper del brazo robótico como se constata en la Figura 2.13.



**Figura 2.13 Referencia del brazo robótico Niryo One**

*Realizado con RobotDK*

*Nota: La flecha roja representa el eje x, la verde el eje y y la azul el eje z.*

Para convertir el vector ubicación  $u$  de una referencia a otra se usó la multiplicación de matrices de transformación. La matriz de transformación global  $T_{C-R}$  tal como se muestra en la Ecuación 2.2 es el resultado de una translación y

dos rotaciones. Para construir las matrices se deben definir las variables  $l_1, l_2, l_3, \alpha$  y  $\beta$ .

$$T_{C-R} = T_{(l_1, l_2, l_3)} T_{(z, \alpha)} T_{(x, \beta)} \quad (2.2)$$

$$t = (l_1, l_2, l_3) \quad (2.3)$$

El vector de traslación  $t$  mostrado en la Ecuación 2.3, se obtuvo a través de la aplicación de un filtro RGB aprovechando las cualidades de la cámara. Este filtro consigue identificar los píxeles que se encuentren dentro del rango del Modelo de Espacio de Color *Hue-Saturation-Value* (HSV) específico. Para su aplicación, se adaptó un sticker justo en la garra del brazo robótico, como se distingue en la Figura 2.14, de forma que podamos trasladar el sistema de referencia directamente al actuador del brazo robótico. En cada inicialización del sistema, las variables de traslación se redefinirán mientras el sticker rojo sea detectado por la cámara, con esta acción nos aseguramos de no perder precisión en el agarre.



**Figura 2.14 Adaptive gripper con sticker rojo para su localización**

*Autoría propia*

A continuación, se muestra la Ecuación 2.4 donde  $u_R$  es el vector ubicación con referencia en el brazo robótico que se obtiene luego de multiplicar la ecuación global por el vector con referencia en la cámara  $u_C$ .

$$u_R = T_{C-R} * u_C \quad (2.4)$$

### 2.2.3.1 Ubicación del Hardware.

Los ángulos de rotación  $\alpha$  y  $\beta$  mostrados en la Ecuación 2.2 para esta aplicación son  $90^\circ$  y  $-90^\circ$  respectivamente. Estos valores son fijos dentro del sistema y no influyen negativamente en el desarrollo general de los procesos siempre que se respete la correcta ubicación del hardware. Las posiciones del brazo robótico y de la cámara están pensadas para evitar oclusión y desperfectos.



**Figura 2.15 Ubicación del hardware**

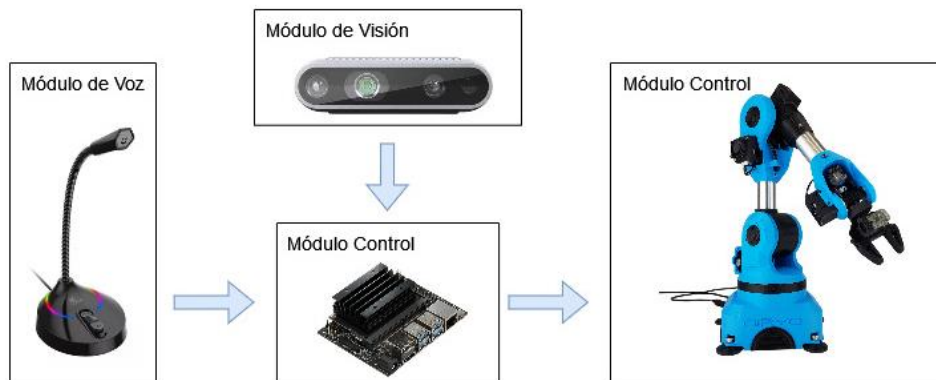
*Autoría propia*

Mientras que el micrófono y la Jetson Nano no necesitan tener una ubicación exacta, el brazo robótico y la cámara RGB-D deberán estar uno frente al otro de forma que el ángulo entre la referencia de la cámara y el gripper del robot sea casi cero. Con esta indicación se estará también aprovechando el alcance horizontal máximo del robot, incrementando la zona de trabajo.

### 2.2.4 Módulo Control

El módulo control está conformado principalmente por el brazo robótico Niryo One y la Jetson Nano que es la que le envía las órdenes. Se apoya del resto de módulos y hardware que constituyen todo el proyecto. En la Figura 2.16 se muestran todos los módulos y como interactúan entre ellos.





**Figura 2.16 Integración de los módulos propuestos**

*Autoría propia*

La comunicación entre la Jetson Nano y el robot se realizó por medio del protocolo SSH a través de una conexión física vía Ethernet. El proyecto que se desarrolló respeta completamente los protocolos, alarmas y procesos propios del software del brazo robótico. Se los usa como ventaja, ya que el sistema interno del brazo robótico mediante comandos permite su calibración automática, la definición de los puntos fijos de la Tabla 2.6 y el control de la velocidad con la que se realizan las manipulaciones.

## **CAPÍTULO 3**

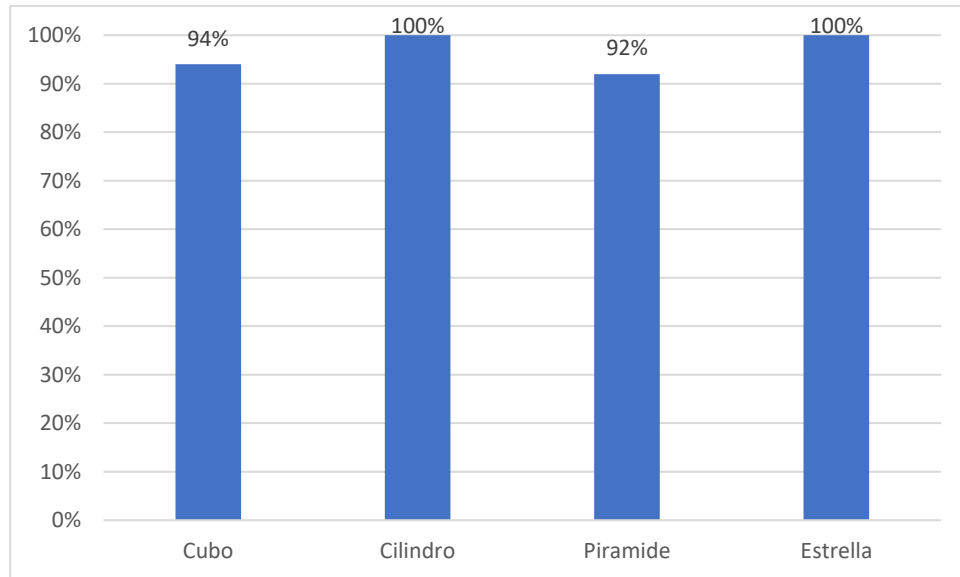
### 3. RESULTADOS Y ANÁLISIS

En este capítulo, siguiendo con la división de la solución por módulos, se mostrarán los resultados obtenidos con la implementación de la solución propuesta. Se mostrarán los tiempos en que cada tarea se realiza, su tasa de aciertos y demás resultados relevantes que se hayan encontrado durante los experimentos realizados.

#### 3.1 Módulo de Voz

El tiempo de inicialización del módulo de voz es **0,256 s**. Inmediatamente después de que se ha instanciado, el sistema de reconocimiento de voz está listo para recibir la orden del usuario. Habiendo recibido el comando, el tiempo de inferencia es de **0,53 ms**. La Figura 3.17 muestra el porcentaje de acierto en el reconocimiento de cada figura, el modelo reconoce las ordenes dirigidas la manipulación del cilindro y la estrella en un 100% mientras que para el cubo el comando se reconoce en un 94% y la pirámide en un 92%. Los porcentajes obtenidos por medio de la evaluación experimental del modelo demuestran su alta efectividad.

Por tratarse de un modelo desarrollado en base a herramientas de PLN y por los resultados obtenidos, se infiere, que la tasa del reconocimiento de las ordenes ingresadas por voz es directamente proporcional a la singularidad de la palabra. Así pues, las palabras '**cube**' y '**py**' definidas para el ingreso del comando para el cubo y la pirámide respectivamente presentan mayor error debido a que en el idioma inglés tienen varios homófonos y pocas sílabas. Según El Alfabeto Fonético Internacional [56] '**cube**' se pronuncian /**kju:b**/ cuya entonación se asimilan a una variada lista de palabras anglosajonas como: '**q**', '**cue**' y '**queue**'. Por otro lado '**py**' tiene una sola sílaba que al no pronunciarse con cadencia tiende a mezclarse con los ruidos del entorno.



**Figura 3.17 Resultado – Modelo de Voz**

*Autoría propia*

### 3.2 Módulo de Visión

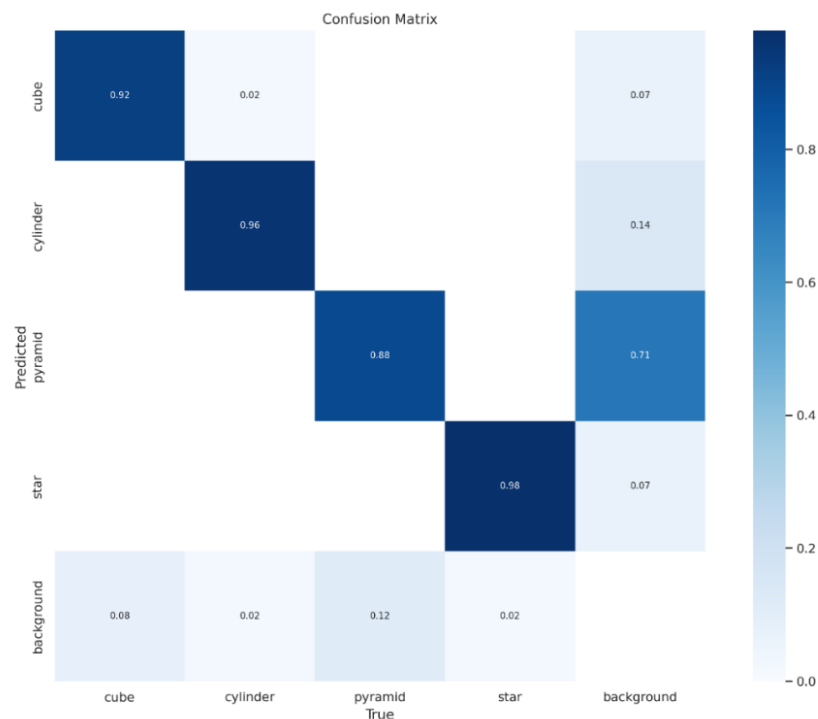
Esta sección incluye el análisis de los resultados del modelo de visión artificial considerando, tiempos de procesamiento y la precisión en el reconocimiento de los objetos.

Los resultados del modelo de detección de objetos se muestran a través de la matriz de confusión presente en la Figura 3.18. El eje horizontal de la matriz nombrado *'True'* representa las clases reales mientras que el eje vertical *'Predicted'* las clases predichas por el módulo. Se presentan cinco clases, las primeras cuatro corresponden a cada una de las formas que se espera el modelo de visión identifique y *'background'* que corresponde a todos los objetos no analizados en este estudio. La clase *'background'* se incluye con el fin de analizar cuan eficiente llega a ser el modelo en diferentes escenarios.

En la matriz de confusión, los elementos de la diagonal principal representan la tasa de acierto o precisión para cada clase y los elementos por fuera de la diagonal principal el error de clasificación. La Figura 3.18 muestra que el cubo se reconoce correctamente con una tasa de precisión de 92%, cilindro 96%, pirámide con 88% y estrella 98%. Además, los valores de la fila correspondientes a la clase

'background' muestran que el nivel de confusión entre las figuras y el entorno es mínimo en el caso del cilindro y la estrella con un 2%, mientras que el cubo y la pirámide presentan 8% y 12% respectivamente. Estos últimos porcentajes se confieren a que los objetos que usamos cotidianamente y en general todo lo que nos rodea tiende a tener formas simples como las de un cubo o una pirámide, en cambio las formas circulares o de múltiples ángulos como la estrella y el cilindro son menos comunes y por lo tanto fáciles de detectar.

La matriz también muestra un valor del 2% en la segunda columna de la primera fila. Lo que sugiere que esa porción de los datos correspondiente a la clase cubo está siendo erróneamente identificada como parte de la clase cilindro. El porcentaje de desacierto detallado se le atribuye a la similitud que llega a tener el cilindro con la forma del cubo cuando se lo observa desde una perspectiva diagonal.



**Figura 3.18 Matriz de confusión - Modelo de Visión Artificial**

Nota: El valor máximo al que cada elemento de la matriz puede llegar es 1.00

Los resultados del rendimiento general del modelo de visión artificial desarrollado se muestran en la Tabla 3.9. El modelo es capaz de predecir correctamente las instancias con un 94% de exactitud, su precisión del 98% sugiere la alta

probabilidad con la que las figuras están siendo reconocidas correctamente y el 90% de sensibilidad indica su efectividad. La tabla sugiere que el modelo es exacto, preciso y eficaz en las tareas de identificación y localización.

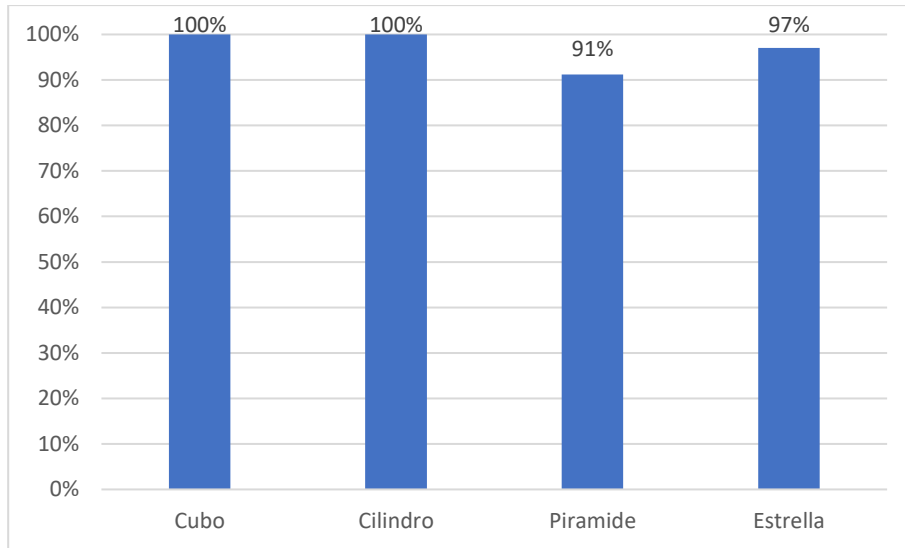
**Tabla 3.9 Resultado del modelo de visión artificial**

Exactitud	Precisión	Sensibilidad
94%	98%	90%

Este modelo de visión por computador integrado a la solución final tarda en promedio **45** segundos en cargarse a la memoria del sistema junto con sus ‘pesos’. Estos tiempos representan sobre todo la capacidad de procesamiento de la Jetson Nano, que, a pesar de haber optimizado la red neuronal artificial utilizada para su implementación en sus 4GB de RAM, limita los recursos a disposición para procesar todos los modelos de inteligencia artificial y cálculo matemático que el proyecto implica.

Ya en el proceso de identificación, este modelo tarda **3,38 s** en inferir, etiquetar y guardar en una lista la ubicación en el plano (x, y) de una imagen recibida en tiempo real. La solución está condicionada para enviar el promedio de la ubicación 2D de 10 frames capturados por la cámara, por esto, el tiempo de identificación de la solución es **33,82 s** aproximadamente. El tiempo de inferencia de este modelo es congruente con la cantidad de información que recibe, procesa y retorna; la cantidad de frames para la obtención del promedio de las coordenadas promueve el decrecimiento de la tasa de error en la identificación y localización de las figuras reflejado en los resultados de la Figura 3.19.

En los experimentos realizados, la capacidad de reconocimiento del modelo de visión fue ampliamente efectiva. En la Figura 3.19 se muestra como el porcentaje de aciertos para el cubo y el cilindro es del 100%, en la detección de la pirámide se obtuvo un 91% y con la estrella un 97%. Durante los ensayos realizados se observó que los desaciertos en la detección de la pirámide y la estrella corresponden a limitaciones del entorno como la cantidad de luz natural o artificial durante diferentes horas del día, la cantidad de ruido visual y la presencia de objetos con colores similares en los que las figuras pueden ‘perderse’.



**Figura 3.19 Resultados – Modulo de Visión Artificial**

### 3.3 Módulo Motor

El módulo control encierra todas las acciones que el brazo robótico realiza, entre ellas el agarre y reubicación de los objetos. Las órdenes y datos, que el robot necesita para efectuar las manipulaciones esperadas, son enviadas desde la Jetson Nano. En este proceso de comunicación y control el sistema se tarda **2,43 ms**, acto seguido el robot tomará la posición inicial detallada en la Tabla 2.6.

El brazo robótico recibe instrucciones para realizar acciones específicas: auto calibración, cinemática inversa, proceso de agarre, reubicación y finalmente la entrega. Los resultados de este módulo se rigen a una zona de trabajo de **300mm x 300mm** determinada por la intersección del alcance horizontal máximo del brazo robótico mostrado en la Figura 2.5 y el rango de visión de la cámara. La auto calibración corresponde completamente al sistema interno del brazo robótico y no representa porcentaje de error.

La fidelidad del proceso de cinemática inversa se obtiene de la comparación de las coordenadas (x, y) generadas por el robot desde su propio software versus las obtenidas durante la fase experimental del proyecto. La Tabla 3.10 detalla la diferencia promedio en centímetros en los ejes (x, y) de cada una de las figuras. Con un error en centímetros de (0,77; 0,70) para el cubo, (0,67; 0,90) el cilindro,

(0,92; 0,82) la pirámide y (0,81; 0,56) para la estrella, en este proceso, el error es menor a 0,92 centímetros en todos los casos. El error de la cinemática inversa se detalla en milímetros y se considera que está dentro de tolerancia puesto que las dimensiones del terminal seleccionado detalladas en la Figura 2.7 están en centímetros dando oportunidad de corregir estas fallas durante el proceso de agarre.

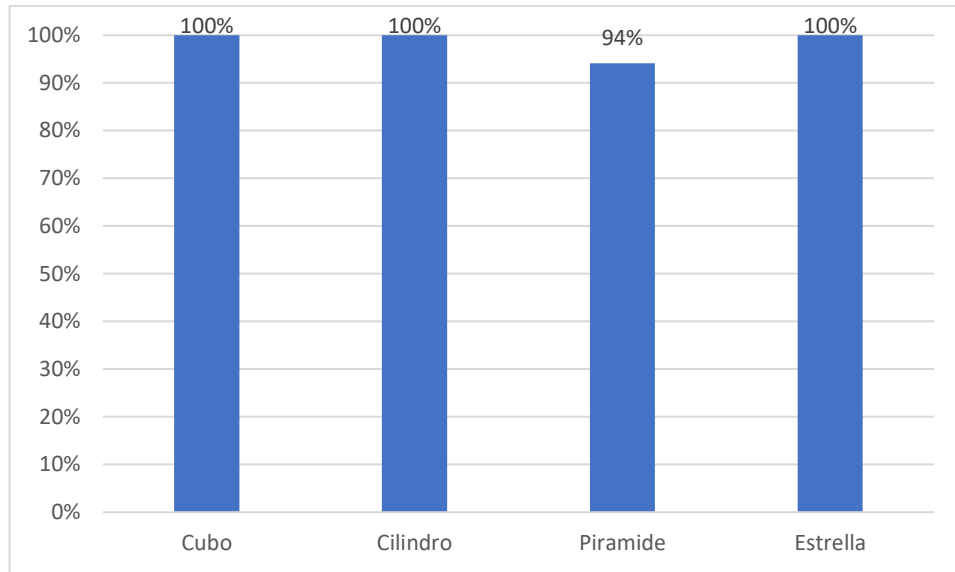
**Tabla 3.10 Error promedio en centímetros de la cinemática inversa**

	<b>Cubo</b>	<b>Cilindro</b>	<b>Pirámide</b>	<b>Estrella</b>
<b>X</b>	0,77	0,67	0,92	0,81
<b>Y</b>	0,70	0,90	0,82	0,56

*Nota: El brazo robótico recibe coordenadas en metros, sin embargo, estos resultados se muestran en centímetros para su mejor interpretación.*

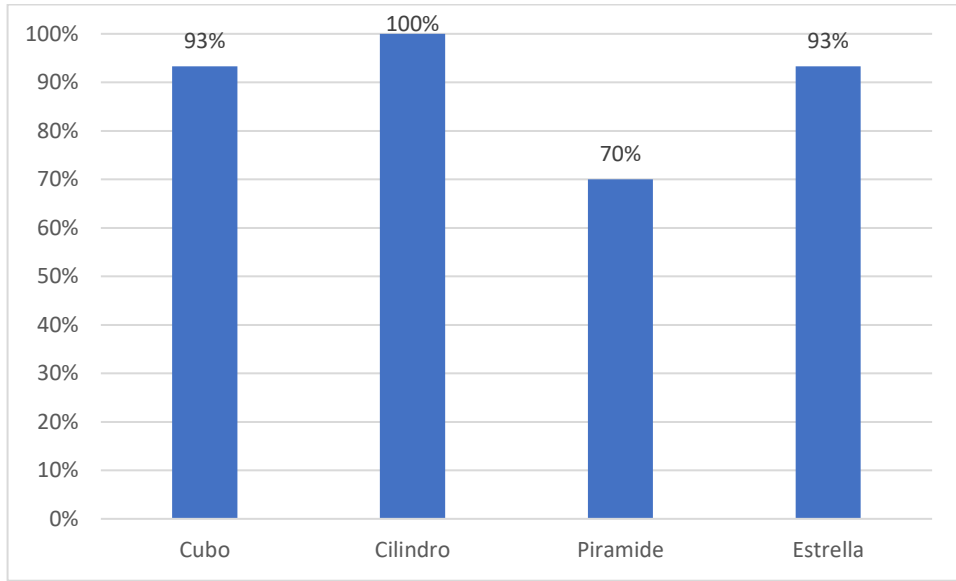
En el proceso de agarre se obtuvo el porcentaje de aciertos mostrado en la Figura 3.20. El brazo robótico es capaz de ejecutar perfectamente el agarre del cubo, el cilindro y la estrella. Con la pirámide, por su forma, el agarre se realiza con una efectividad del 94%. El agarre depende de varios factores como la rugosidad de las piezas, su peso y la capacidad de ajuste que tenga el terminal para adaptarse a las diferentes formas, de acuerdo con los resultados la pirámide por sus ángulos se convierte en un objeto que a diferencia del cilindro o el cubo presenta menor superficie de contacto para con el gripper provocando este pequeño porcentaje de error.





**Figura 3.20 Resultados – Agarre – Modulo Control**

La Figura 3.21 detalla el porcentaje de reubicación exitosa obtenida durante los experimentos realizados. El sistema robótico colaborativo logra llevar de una posición a otra cada una de las figuras con una tasa de acierto del 100% para el cilindro, 93% para el cubo y la estrella además de un 70% para la pirámide. Esta última tarea del brazo robótico implica mantener el control sobre la fuerza de agarre que imprime el terminal sobre el objetivo durante los movimientos continuos que sus articulaciones realizan. El driver del Niryo One mantiene estable el sistema de control antes mencionado siempre que los movimientos se realicen a una velocidad prudente, pero dependiendo de la calidad del agarre y de la rugosidad de las superficies el sistema puede fallar. En este estudio se observa que la pirámide tiene un porcentaje de reubicación de 70% debido a su forma angulada y la perspectiva desde la que el terminal la sujeta, *i.e.*, es necesario considerar que la orientación de agarre es perpendicular al plano X Y.



**Figura 3.21. Reubicación – Modulo Control**

# **CAPÍTULO 4**

## 4. CONCLUSIONES Y RECOMENDACIONES

Este proyecto constituido por un módulo de voz, uno de visión y uno de control se ha desarrollado con la intención de presentar una alternativa funcional, de baja potencia y bajo costo para la implementación de sistemas colaborativos en diferentes ámbitos de la industria. Pensado para mejorar la calidad de vida de los usuarios y de los servicios brindados, el sistema robótico se ha diseñado minuciosamente para que cumpla con las tareas de reconocimiento, localización, agarre y reubicación de cuatro figuras representativas.

### Conclusiones

Tras el análisis de los resultados obtenidos en los múltiples experimentos realizados se llega a las siguientes conclusiones:

- El módulo de voz conformado por un micrófono y un modelo de reconocimiento de voz obtenido a partir de las herramientas de la consola de Picovoice, presenta un promedio del 97% de acierto en la identificación de las órdenes ingresadas. El modelo demora un aproximado de 0,256 s en integrarse a la minicomputadora y 0,53 ms en inferir los comandos por voz. Los tiempos y el porcentaje de eficiencia del modelo permiten concluir que se ha logrado desarrollar un módulo de voz que funciona de forma continua, efectiva y sin retardos.
- El módulo de visión constituido por un modelo de visión artificial y una cámara RGB-D muestra gran eficiencia y efectividad en su unificación al sistema colaborativo implementado. Con un dataset balanceado de 4,680 imágenes luego de realizar el aumento de datos, se obtiene un modelo de visión artificial entrenado con YOLOV8 que presenta una exactitud del 94%, un 98% de precisión y un 90% de sensibilidad. Los porcentajes obtenidos indican que la capacidad de identificación y localización de las cuatro formas incluidas en este estudio es ampliamente eficaz. En referencia al tiempo de procesamiento del modelo de visión artificial se obtiene un promedio de 47 segundos para inicialización y 3 segundos para inferencia. Si bien, estos tiempos pueden representar un retraso en el funcionamiento general, en especial en su primera ejecución, se concluye que se trata de

un modelo robusto y optimizado para su ejecución con los recursos limitados de la Jetson Nano.

- Las tareas realizadas por el brazo robótico como la cinemática inversa, el agarre de los objetos y su posterior reubicación muestran un buen porcentaje de efectividad. Para que el robot se dirija hasta la ubicación del objetivo se deberá primero pasar las coordenadas referenciadas en la cámara a la referencia del robot. Esta etapa de procesamiento de datos es crucial para el funcionamiento integral del sistema, razón por la que a partir de un vector de ubicación y una matriz de transformación que se reescriben con cada inicialización del programa se mitigan errores e incentiva la adaptabilidad del proyecto a los diferentes entornos en los que se puede implementar. La exactitud de la cinemática inversa con la que la solución trabaja se mide en centímetros, considerando la ubicación deseada y la ubicación alcanzada, obtenido un error promedio de 0,92 centímetros en el eje X y 0,9 centímetros en el eje Y. Este error es compensado con la capacidad de apertura y agarre del terminal 'Gripper 3' seleccionado. Las tareas de agarre y reubicación se concluyen con una efectividad del 100% y del 93% en promedio respectivamente para el cubo, estrella y el cilindro. La pirámide por su geometría es sujeta correctamente el 94% de las veces y se relocaliza adecuadamente en el 70% de los casos.
- La integración de los modelos de inteligencia artificial tanto de reconocimiento de voz como de visión por computador para su funcionamiento offline dentro de la Jetson Nano es completamente posible. Los módulos desarrollados junto con el hardware robótico seleccionado que presenta un error promedio menor a 1cm, es la combinación ideal para el ensamblaje de un proyecto de robótica colaborativa impulsada por inteligencia artificial de bajo costo. El costo del producto final deducido en \$ 3 110, valor que representa el 15% del valor de productos afines comerciales. Esta dramática reducción en costos garantiza una diseminación de estos sistemas colaborativos, tanto en consumidores como en instituciones de investigación, lo que acorta la brecha tecnológica que afecta países con recursos limitados como Ecuador.

## Recomendaciones

- Para el entrenamiento de modelos de inteligencia artificial se recomienda construir una base de datos balanceada, variada y con la mayor similitud dable a las diferentes posibilidades de datos con el que el producto final tendrá que trabajar. Se sugiere procurar el balance de los datos según la cantidad de clases, de forma que se asegure la generalización y el rendimiento del modelo y facilitando la interpretación de los resultados.
- Tomar en cuenta que por las características de la Jetson Nano, para su implementación se deberán adaptar los modelos de inteligencia artificial a los requerimientos de software de la minicomputadora, además los procesos deben realizarse en orden y procurando no saturar la memoria RAM.
- Se recomienda también, en proyectos que incluyen manipulaciones con brazos robóticos como el Niryo One iniciar realizando ensayos a bajas velocidades y en ambientes controlados para garantizar la seguridad tanto de los desarrolladores como la de los implementos. Durante la experimentación con un brazo robótico se debe tener presente que para mantener su calibración y por recomendación general se aconseja evitar realizar movimientos bruscos, manipulaciones que impliquen la extensión máxima de las articulaciones y colisiones con superficies poco flexibles.
- De acuerdo con los resultados obtenidos, se alienta continuar con el desarrollo de proyectos de esta índole con el objetivo de promover la inclusión de robots colaborativos de bajo costo en procesos industriales. Se recomienda también, hacer un estudio enfocado a la investigación de la orientación del actuador final para la sujeción de objetos.

# **BIBLIOGRAFÍA**

## 5. BIBLIOGRAFÍA

- [1] International Federation of Robotics (IFR), «Executive Summary World Robotics 2017 Industrial Robots,» *World Robotics 2017 Report*, pp. 15-24, 2017.
- [2] G. Lefranc, I. López, R. Osorio-Comparán y M. Peña, «Cobots in automation and at home,» de *2022 IEEE International Conference on Automation/XXV Congress of the Chilean Association of Automatic Control (ICA-ACCA)*, 2022.
- [3] K. S. Amores Romero y K. G. Trelles Muñoz, «Implementación de un sistema para detectar la enfermedad de la Sigatoka Negra en una plantación de banano empleando técnicas de visión artificial,» Azay, Ecuador, 2022.
- [4] J. F. H. Patiño, B. L. V. Carrascal, D. R. Bautista y N. G. Díaz, «Impacto transformador de la inteligencia artificial y aprendizaje autónomo en la producción agropecuaria: un enfoque en la sostenibilidad y eficiencia,» *Formación Estratégica*, vol. 1, nº 7, pp. 40-55, 2023.
- [5] A. Martín Sanchís, «Desarrollo de un sistema de visión artificial basado en redes convolucionales para el conteo automatizado de microcápsulas en tejidos.,» Universidad Politécnica de Valencia, Valencia, 2023.
- [6] R. C. Schank, «What Is AI, Anyway?,» *AI Magazine*, vol. 8, nº 4, pp. 59-65, 1987.
- [7] J. V. Sancho Escrivá, C. Fanjul, M. de la Iglesia Vaya, J. A. Montell y M. J. Escarti, «Aplicación de la Inteligencia Artificial con Procesamiento del Lenguaje Natural para textos de investigación cualitativa en la relación médico-paciente con enfermedad mental mediante el uso de tecnologías móviles,» *Revista de Comunicación y Salud*, vol. 10, nº 1, pp. 19-41, 2020.
- [8] V. Arias, J. Salazar, C. Garicano, J. Contreras, G. Chacón, M. Chacín-González y V. Bermúdez-Pirela, «Una introducción a las aplicaciones de la inteligencia artificial en Medicina: Aspectos históricos.,» *Revista Latinoamericana de Hipertensión*, vol. 14, nº 5, pp. 590-600, 2019.
- [9] C. -C. J. Hsu, P. -J. Hwang, W. -Y. Wang, Y. -T. Wang y C. -K. Lu, «Vision-Based Mobile Collaborative Robot Incorporating a Multicamera Localization System,» *IEEE Sensors Journal*, vol. 23, nº 18, pp. 21853-21861, 2023.



- [10] C. Castellanos Ormeño, «Sistema de rehabilitación de miembro superior asistido mediante un robot colaborativo UR3,» Universidad de Alicante. Departamento de Física, Ingeniería de Sistemas y Teoría de la Señal, 2021.
- [11] S. S. José, «Ayuda a personas mayores con un robot social: estimulación cognitiva.,» Universidad Carlos III de Madrid, Madrid, 2019.
- [12] M. Paliga, «Las relaciones entre la fluidez en la interacción humano-cobot con el desempeño laboral y la satisfacción laboral entre los operadores de cobots: el papel moderador de la carga de trabajo.,» *MDPI Open Access Journals* , vol. 6, p. 20, 2023.
- [13] V. Di Pasquale, V. De Simone, V. Giubileo y S. Miranda, «A taxonomy of factors influencing worker's performance in human–robot collaboration,,» *IET Collaborative Intelligent Manufacturing*, vol. 1, nº 5, 2023.
- [14] G. Borges, A. Reis, R. Neto, D. de Mattos, A. Cardoso, H. Gonçalves, E. Merino, A. Colim, P. Carneiro y P. Arezes, «Decision-Making Framework for Implementing Safer Human-Robot Collaboration Workstations: System Dynamics Modeling.,» *Safety*, vol. 7, nº 4, p. 75, Oct 2021.
- [15] G. Vargas-Yara y J. A. Calderon, «ROBÓTICA COLABORATIVA: UNA PROSPECTIVA SOBRE LA TRANSFORMACIÓN DE LAS ACTIVIDADES HUMANAS,» *SENA Centro Metalmeccanico - Regional Distrito Capital -I Grupo GICEMET*, pp. 10-14, 2022.
- [16] J.-V. d. A. Dalprá, «La protección del trabajo digno mediante los impactos de las nuevas formas de robótica laboral, inteligencia artificial y nuevas tecnologías,» 2020.
- [17] M. Toboso Martín y M. A. Grau Ruiz, «Vivir con robots. Reflexiones éticas, jurídicas, sociales y culturales,» *Arbor-ciencia Pensamiento Y Cultura*, vol. 197, nº 802, p. a623, 2021.
- [18] A. G. Y. -S. Ong, «AIR5: Five Pillars of Artificial Intelligence Research,» *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 3, nº 5, pp. 411-415, Oct 2019.
- [19] A. Gelbukh, «Procesamiento de lenguaje natural y sus aplicaciones.,» *Komputer Sapiens*, vol. 1, pp. 6-11, 2010.

- [20] T. Domínguez Míngues, *Visión Artificial. Aplicaciones prácticas con OpenCV-Python*, MARCOMBO, 2021.
- [21] P. Rosenberger, A. Cosgun, R. Newbury, J. Kwan, V. Ortenzi, P. Corke y M. Grafinger, «Object-Independent Human-to-Robot Handovers Using Real Time Robotic Vision,» *IEEE Robotics and Automation Letters*, vol. 6, nº 1, pp. 17-23, 2021.
- [22] O. K. K. Sohn, «Technology acceptance theories and factors influencing artificial Intelligence-based intelligent products,» *Telemat. Inform*, vol. 47, pp. 1-14, 2019.
- [23] M. M. A. E. Y. M. M. M. Nazar, «A Systematic Review of Human-Computer Interaction and Explainable Artificial Intelligence in Healthcare With Artificial Intelligence Techniques,» *Acceso IEEE*, vol. 9, pp. 153316-153348, 2021.
- [24] A. Bhattacharjee, P. Das, D. Kundu, S. Ghosh y S. D. Gupta, «A Real-Time Face Motion Based Approach towards Modeling Socially Assistive Wireless Robot Control with Voice Recognition.,» *IJACSA Int. J. Adv. Comput. Sci. Appl*, vol. 6, pp. 205-220, 2015.
- [25] S. Pawar, P. Futane, N. Uke, S. Patil, R. Shah, H. Shah y O. Jain, «AI-Based Autonomous Voice-Enabled Robot with Real-Time Object Detection and Collision Avoidance Using Arduino. In: Neustein, A., Mahalle, P.N., Joshi, P., Shinde, G.R. (eds) AI, IoT, Big Data and Cloud Computing for Industry 4.0,» *Signals and Communication Technology*, 2023.
- [26] PICOVOICE, «PICOVOICE DOCS,» 2023.
- [27] Picovoice, «Picovoice — Voice commands & NLU,» 2022. [En línea]. Available: <https://picovoice.ai/platform/rhino/>.
- [28] P. Setiawan y R. Yusuf, «IoT Device Control with Offline Automatic Speech Recognition on Edge Device,» *2022 12th International Conference on System Engineering and Technology (ICSET)*, pp. 111-115, 2022.
- [29] E. Wu y D. Maslov, «Voice Interaction Drive-through Self-service Station,» *Raspberry Pi Retail Applications*, pp. 93-121, 2022.
- [30] R. Zhang, C. Xu y M. Xie, «Secure Decentralized IoT Service Platform Using Consortium Blockchain,» *Sensors*, vol. 22, nº 21, pp. 81-86, 2022.

- [31] D. Deepa, A. Sivasangari, R. Roonwal y R. Nayan, «Pothole Detection using Roboflow Convolutional Neural Networks,» de *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2023.
- [32] J. Terven y D. Cordova-Esparza, «A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond,» 2023.
- [33] Y. Zhang, Z. Wu, X. Wang, W. Fu, J. Ma y G. Wang, «Improved YOLOv8 Insulator Fault Detection Algorithm Based on BiFormer,» de *2023 IEEE 5th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, Shenyang, China, 2023.
- [34] Ultralytics, «Ultralytics YOLOv8 Docs,» 2023. [En línea]. Available: <https://docs.ultralytics.com/#yolo-a-brief-history>. [Último acceso: Noviembre 2023].
- [35] V. Cherkasov, L. Rybak, G. Carbone, D. Malyshev y D. Dyakonov, «Method of Localization of Racks with Biomaterial for Robot Grasp Based on Segmented Contour Processing.,» *Mechanisms and Machine Science*, vol. 148, 2023.
- [36] N. Wang, H. Liu, Y. Li, W. Zhou y M. Ding, «Segmentation and Phenotype Calculation of Rapeseed Pods Based on YOLO v8 and Mask R-Convolution Neural Networks,» *Plants*, vol. 12, nº 18, pp. 33-28, 2023.
- [37] D. Kumar y N. Muhammad, «Object Detection in Adverse Weather for Autonomous Driving through Data Merging and YOLOv8,» *Sensors*, vol. 23, nº 20, pp. 84-71, 2023.
- [38] A. Ollero Baturone, *ROBÓTICA Manipuladores y robots móviles*, Sevilla: MARCOMBO S.A, 2001.
- [39] R. F. Mendoza Garcia, *Fundamentos de Robótica Herramientas Matemáticas para la Localización Espacial, Matrices de transformación Homogenea.*, Arica, Chile: Escuela Universitaria de Ingeniería Mecánica, Universidad de Tarapacá, 2014.
- [40] W. Sanz, «Enseñanza y Aprendizaje de Robótica Industrial desde la Virtualidad,» *Revista Tecnológica-Educativa Docentes 2.0*, vol. 11, nº 2, pp. 19-27, 2021.
- [41] NIRYODocs, «Niryo One Mechanical Specifications,» 2018.
- [42] A. S. Ahtik, «Integracija robotskega manipulatorja Niryo One v sistem Manus,» *Doctoral dissertation, Univerza v Ljubljani, Fakulteta za računalništvo in informatiko*, 2022.

- [43] NVIDIA Developer, «Get started with Jetson Nano Developer Kit,» 2023. [En línea]. Available: <https://developer.nvidia.com/embedded/learn/get-started-jetson-nano-devkit>.
- [44] NVIDIA Developer, «NVIDIA Jetson Modules, support, Ecosystem, and lineup,» 2023. [En línea]. Available: <https://developer.nvidia.com/embedded/jetson-modules>.
- [45] A. Basulto-Lantsova, J. A. Padilla-Medina, F. J. Perez-Pinal y A. I. Barranco-Gutierrez, «Performance comparative of OpenCV Template Matching method on Jetson TX2 and Jetson Nano developer kits,» de *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, 2020.
- [46] P. Inthanon y S. Mungsing, «Detection of Drowsiness from Facial Images in Real-Time Video Media using Nvidia Jetson Nano,» de *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Phuket, Thailand, 2020.
- [47] T. T. Than, D. K. D. Danh, H. L. Nguyen y M. S. Nguyen, «Researching and Implementing the Posture Recognition Algorithm of the Elderly on Jetson Nano,» de *2022 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, Phu Quoc, Vietnam, 2022.
- [48] S. Chavan, J. Ford, X. Yu y J. Saniie, «Plant Species Image Recognition using Artificial Intelligence on Jetson Nano Computational Platform,» de *2021 IEEE International Conference on Electro Information Technology (EIT)*, Mt. Pleasant, MI, USA, 2021.
- [49] Intel RealSense, «Depth Camera D435i – Intel® RealSense™ depth and tracking cameras», Intel® RealSense™ Depth and Tracking Cameras,» Junio 2021. [En línea]. Available: <https://www.intelrealsense.com/depth-camera-d435i/>. [Último acceso: 2023].
- [50] J. Haviland, N. Sándershauf y P. Corke, «A Holistic Approach to Reactive Mobile Manipulation,» *IEEE Robotics and Automation Letters*, vol. 7, nº 2, pp. 3122-3129, 2022.
- [51] A. Newman, G. Yang, D. A. B. Wang y J. Saniie, «mbedded Mobile ROS Platform for SLAM Application with RGB-D Cameras,» de *2020 IEEE International Conference on Electro Information Technology (EIT)*, Chicago, IL, USA, 2020.

- [52] X. G. Torres Correa, NEC - Norma Ecuatoriana de la Construcción - ACCESIBILIDAD UNIVERSAL (AU), Ministerio de Desarrollo Urbano y Vivienda (MIDUVI), 2019.
- [53] PICOVOICE, «Rhino, Picovoice Console,» 2023. [En línea]. Available: <https://console.picovoice.ai/rhn/0f6cf2a9-2fb1-42a4-9eb3-6e4c13afaaa4>.
- [54] M. León Cuzco, «Roboflow,» 2023. [En línea]. Available: <https://universe.roboflow.com/avm1/pieces-gto4b>.
- [55] IntelRealSense, «Transformation Matrices, D435I - Issue #1449 - IntelRealSense/RealSense-ROS,» 16 Octubre 2020. [En línea]. Available: <https://github.com/IntelRealSense/realsense-ros/issues/1449#issuecomment-709889034>. [Último acceso: 2023].
- [56] International Phonetic Association, Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet, Cambridge University Press, 1999.
- [57] Z. Bi, C. Luo, Z. Miao, B. Zhang, W. Zhang y L. Wang, «Safety assurance mechanisms of collaborative robotic systems in manufacturing.,» *Robot. Comput. Integr. Manuf.*, p. 67, 2021.
- [58] L. M. y F. F. R. Gervasi, «A conceptual framework to evaluate human-robot collaboration,» *The International Journal of Advanced Manufacturing Technology*, vol. 108, nº 3, pp. 841-865, May 2020.
- [59] T. -. K. Kim, «Short Research on Voice Control System Based on Artificial Intelligence Assistant,» de *2020 International Conference on Electronics, Information, and Communication (ICEIC)*, Barcelona, España, 2020.
- [60] Y. Yan y Y. Jia, «A Review on Human Comfort Factors, Measurements, and Improvements in Human–Robot Collaboration,» *Sensors*, vol. 22, nº 19, p. 7431, Sep 2022.
- [61] S. Hopko, J. Wang y R. K. Mehta, «Human Factors Considerations and Metrics in Shared Space Human-Robot Collaboration: A Systematic review,» *Frontiers in Robotics and AI*, vol. 9, 9 feb 2022.