

Universidad Católica de Santa María  
Facultad de Ciencias e Ingenierías Físicas y Formales  
Escuela Profesional de Ingeniería de Sistemas



Desarrollo de un dispositivo de asistencia ergonómico y de bajo costo para reconocimiento de billetes basado en visión artificial e IoT para personas con discapacidad visual en la ciudad de Arequipa

Tesis presentada por los Bachilleres:

Caytairo Silva, Nicolás Esleyder

ORCID: 0000-0003-1656-396X

Peña Alejandro, Jackeline Melady

ORCID: 0000-0002-3586-1826

para optar el Título Profesional de Ingeniero de Sistemas

Asesor(a):

Dra. Castro Gutierrez, Eveling Gloria

ORCID: 0000-0002-0203-041X

Arequipa - Perú

2024

UCSM-ERP

# UNIVERSIDAD CATÓLICA DE SANTA MARÍA

## INGENIERIA DE SISTEMAS

### TITULACIÓN CON TESIS

#### DICTAMEN APROBACIÓN DE BORRADOR

Arequipa, 09 de Abril del 2024

**Dictamen: 011834-C-EPIS-2024**

Visto el borrador del expediente 011834, presentado por:

**2019118521 - CAYTUIRO SILVA NICOLAS ESLEYDER**

**2019119012 - PEÑA ALEJANDRO JACKELINE MELADY**

Titulado:

**DESARROLLO DE UN DISPOSITIVO DE ASISTENCIA ERGONÓMICO Y DE BAJO COSTO PARA RECONOCIMIENTO DE BILLETES BASADO EN VISIÓN ARTIFICIAL E IOT PARA PERSONAS CON DISCAPACIDAD VISUAL EN LA CIUDAD DE AREQUIPA**

Nuestro dictamen es:

**APROBADO**

**29302116 - DELGADO DELGADO FREDY RAMIRO  
DICTAMINADOR**



**29612305 - SULLA TORRES JOSE ALFREDO  
DICTAMINADOR**



**43635330 - ESQUICHA TEJADA JOSE DAVID  
DICTAMINADOR**



# Desarrollo de un dispositivo de asistencia ergonómico y de bajo costo para reconocimiento de billetes basado en visión artificial e IoT para personas con discapacidad visual en la ciudad de Arequipa

## ORIGINALITY REPORT

3%

SIMILARITY INDEX

5%

INTERNET SOURCES

2%

PUBLICATIONS

2%

STUDENT PAPERS

## PRIMARY SOURCES

1

[www.learndatasci.com](http://www.learndatasci.com)

Internet Source

1%

2

[sedici.unlp.edu.ar](http://sedici.unlp.edu.ar)

Internet Source

1%

3

[hdl.handle.net](http://hdl.handle.net)

Internet Source

1%

4

[cybertesis.unmsm.edu.pe](http://cybertesis.unmsm.edu.pe)

Internet Source

1%

5

[imt.mx](http://imt.mx)

Internet Source

1%

Exclude quotes Off

Exclude matches < 1%

Exclude bibliography Off

## Dedicatoria

*Dedicamos esta tesis a nuestras familias, en especial a nuestros padres, que siempre nos han apoyado para que podamos alcanzar nuestros sueños; gracias a su amor, paciencia y aliento constante podemos seguir superando retos.*



## Agradecimientos

*Agradecemos a Dios, por bendecir a nuestras familias y a nuestros seres queridos.*

*A la Universidad Católica de Santa María, por ser nuestra alma mater.*

*A nuestros docentes y mentores, por compartirnos sus conocimientos y sabiduría, especialmente a la Dra. Eveling Castro Gutierrez, cuya guía y apoyo incondicional fueron fundamentales en cada etapa de nuestra tesis.*

*A la Asociación Unión de Ciegos de Arequipa, por aceptar ser parte de este proyecto, pues sin su apoyo, no habría sido posible.*



## RESUMEN

La discapacidad visual es una condición que presenta desafíos significativos en la vida diaria de quienes la experimentan. En la ciudad de Arequipa, como en muchas otras regiones, las personas con discapacidad visual enfrentan dificultades adicionales al tratar de identificar billetes y objetos en su entorno cotidiano. Por ello, la presente investigación tiene como objetivo construir un dispositivo de asistencia ergonómico y de bajo costo para reconocimiento de billetes basado en visión artificial e IoT para personas con discapacidad visual en la ciudad de Arequipa. La metodología utilizada se basó en Design Thinking, en cuyas fases se detallan el proceso de desarrollo del dispositivo de asistencia propuesto que son: (1) empatía, (2) definición, (3) ideación, (4) prototipado y (5) evaluación. Para la evaluación del dispositivo se realizaron pruebas funcionales y no funcionales. Las pruebas funcionales se enfocaron en comprobar el reconocimiento de billetes con 44 miembros de la Asociación Unión de Ciegos de Arequipa, en diferentes condiciones ambientales. Por otro lado, las pruebas no funcionales se orientaron a evaluar aspectos como el rendimiento, seguridad, confiabilidad y usabilidad del dispositivo. Los resultados demuestran que el dispositivo cumple con las especificaciones requeridas, además que alcanza un nivel de precisión del 99.75 % con el modelo de visión artificial Xception. En conclusión, fue posible realizar la construcción de un prototipo de asistencia ergonómico y de bajo costo para el reconocimiento de billetes para personas con discapacidad visual, gracias a la integración de técnicas de visión artificial con la tecnología de IoT.

**Palabras Claves.** Dispositivo de Asistencia, Discapacidad Visual, Ergonómico, Bajo Costo, Reconocimiento de Billetes, Visión Artificial, IoT.

## ABSTRACT

Visual impairment is a condition that presents significant challenges in the daily lives of those who experience it. In the city of Arequipa, as in many other regions, people with visual impairments face additional difficulties when trying to identify bills and objects in their everyday environment. Therefore, the present research aims to build an ergonomic and low-cost assistance device for bill recognition based on computer vision and IoT for people with visual impairments in the city of Arequipa. The methodology used was based on Design Thinking, in whose phases the development process of the proposed assistance device is detailed: (1) empathy, (2) definition, (3) ideation, (4) prototyping, and (5) evaluation. For the evaluation of the device, functional and non-functional tests were conducted. Functional tests focused on verifying bill recognition with 44 members of the Arequipa Blind Union Association, under different environmental conditions. On the other hand, non-functional tests aimed to evaluate aspects such as performance, security, reliability, and usability of the device. The results demonstrate that the device meets the required specifications, and it achieves a precision level of 99.75% with the Xception artificial vision model. In conclusion, it was possible to construct an ergonomic and low-cost assistance prototype for bill recognition for people with visual impairments, thanks to the integration of computer vision techniques with IoT technology.

**Keywords.** Assistance Device, Visual Impairment, Ergonomic, Low Cost, Bill Recognition, Computer Vision, IoT.

# Índice general

DEDICATORIA

AGRADECIMIENTOS

RESUMEN

ABSTRACT

INTRODUCCIÓN

1

CAPITULO I

2

**1 PLANTEAMIENTO DE LA INVESTIGACIÓN**

**3**

1.1 Planteamiento del problema . . . . .	3
1.2 Objetivos de la Investigación . . . . .	5
1.2.1 Objetivo General . . . . .	5
1.2.2 Objetivos Específicos . . . . .	5
1.3 Preguntas de Investigación . . . . .	5
1.4 Hipótesis, Variables e Indicadores . . . . .	6
1.4.1 Hipótesis . . . . .	6
1.4.2 Variables de investigación . . . . .	6
1.4.3 Indicadores . . . . .	6
1.5 Línea y Sublínea de Investigación . . . . .	7
1.6 Palabras Clave . . . . .	7
1.7 Solución Propuesta . . . . .	7
1.7.1 Justificación e Importancia . . . . .	7
1.7.2 Descripción de la Solución . . . . .	8
1.7.2.1 Esquema general de la propuesta . . . . .	8
1.7.2.2 Esquema del funcionamiento (Software - Visión artificial) . . . . .	9
1.7.2.3 Esquema del funcionamiento (Hardware - IoT) . . . . .	10
1.7.3 Aporte . . . . .	11

CAPITULO II

13

**2 FUNDAMENTOS TEÓRICOS**

**14**

2.1 Antecedentes de investigación . . . . .	14
2.2 Estado del Arte . . . . .	16
2.3 Bases Teóricas de la Investigación . . . . .	29
2.3.1 Familias de billetes de 10, 20, 50 y 100 nuevos soles . . . . .	29
2.3.1.1 Nueva familia de billetes . . . . .	29
2.3.1.2 Familia de billetes 2011 . . . . .	31
2.3.2 Discapacidad Visual . . . . .	33
2.3.3 Tecnologías de asistencia . . . . .	34

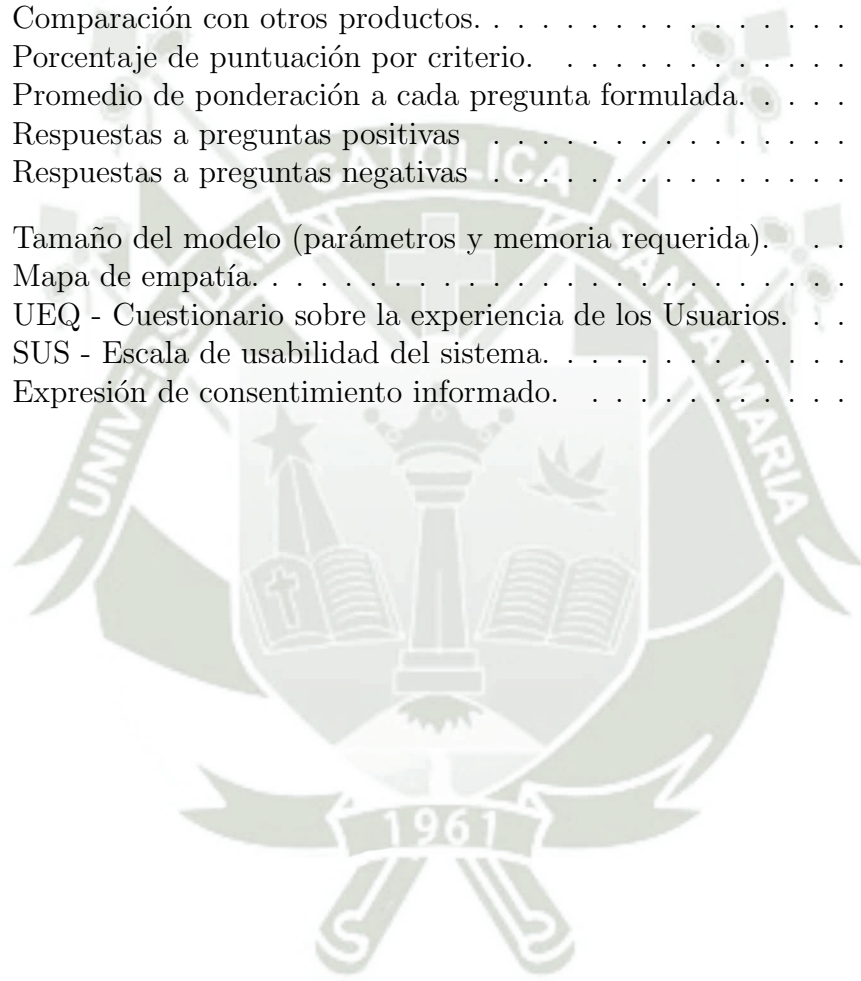
2.3.4	Tecnologías vestibles . . . . .	34
2.3.5	Tecnologías de asistencia e IoT . . . . .	34
2.3.6	Aprendizaje Supervisado . . . . .	35
2.3.7	Redes Neuronales Convolucionales . . . . .	36
2.3.7.1	Elementos básicos que componen una Red Neuronal . . . . .	36
2.3.7.2	Capas de convolución . . . . .	38
2.3.7.3	Capas de Pooling . . . . .	38
2.3.7.4	Capas fully-connected . . . . .	39
2.3.7.5	Capas de softmax y clasificación . . . . .	39
2.3.8	Visión Artificial . . . . .	40
2.3.9	Algoritmos de Visión Artificial . . . . .	40
2.3.10	Transferencia de aprendizaje y modelos de deep learning pre-entrenados . . . . .	41
2.3.11	Estándares de ergonomía para dispositivos IoT vestibles . . . . .	49
2.3.12	Diseño inclusivo vs diseño universal . . . . .	51
<b>CAPITULO III</b>		<b>59</b>
<b>3</b>	<b>MARCO METODOLÓGICO</b>	<b>60</b>
3.1	Alcances y Limitaciones . . . . .	60
3.2	Tipo y Nivel de Investigación . . . . .	60
3.2.1	Tipo de Investigación . . . . .	60
3.2.2	Nivel de Investigación . . . . .	60
3.3	Universo, Población y Muestra . . . . .	61
3.3.1	Universo . . . . .	61
3.3.2	Población . . . . .	61
3.3.3	Muestra . . . . .	61
3.4	Métodos, Técnicas e Instrumentos de Recolección de Datos . . . . .	62
3.4.1	Metodología de la investigación . . . . .	62
3.4.2	Técnicas de recolección de datos . . . . .	62
3.4.3	Instrumentos de recolección de datos . . . . .	62
3.4.4	Técnicas de procesamiento de datos . . . . .	62
3.4.5	Herramientas para el procesamiento de datos . . . . .	63
<b>CAPITULO IV</b>		<b>64</b>
<b>4</b>	<b>DESARROLLO DE LA PROPUESTA DE INVESTIGACIÓN</b>	<b>65</b>
4.1	Desarrollo en Visión Artificial . . . . .	65
4.1.1	Generación del Conjunto de Datos . . . . .	65
4.1.1.1	Configuración del Equipo de Captura de Imágenes . . . . .	65
4.1.1.2	Procedimiento para la Captura de Imágenes . . . . .	65
4.1.2	Modelos de Aprendizaje Profundo . . . . .	70
4.1.2.1	Entrenamiento de modelos de aprendizaje profundo preentrenados . . . . .	70
4.1.3	Resultados obtenidos . . . . .	77
4.1.3.1	Configuración de hiperparámetros . . . . .	77
4.1.3.2	Duración del entrenamiento y precisión de los modelos . . . . .	78
4.2	Desarrollo en IoT . . . . .	79
4.2.1	Diseño del dispositivo IoT . . . . .	79

4.2.1.1	Requisitos Funcionales . . . . .	79
4.2.1.2	Requisitos no Funcionales . . . . .	81
4.2.1.3	Flujo del funcionamiento del dispositivo de asistencia . . . . .	82
4.2.1.4	Diseño del Circuito . . . . .	83
4.2.2	Diseño e impresión en 3D . . . . .	84
4.2.2.1	Diseño de la carcasa . . . . .	84
4.2.2.2	Dispositivo final . . . . .	86
4.3	Integración de Visión Artificial e IoT . . . . .	87
4.4	Pruebas y validación del dispositivo . . . . .	91
4.4.1	UEQ - Cuestionario sobre la experiencia de los Usuarios . . . . .	91
4.4.2	SUS - Escala de usabilidad del sistema . . . . .	94
4.4.3	Tiempo de procesamiento y respuesta del dispositivo . . . . .	98
<b>CAPITULO V</b>		<b>99</b>
<b>5</b>	<b>ANÁLISIS Y VALIDACIÓN DE RESULTADOS</b>	<b>100</b>
5.1	Resultados . . . . .	100
5.1.1	Resultados en Visión Artificial . . . . .	100
5.1.1.1	Tamaño del modelo (parámetros y memoria requerida) . . . . .	100
5.1.1.2	Relación entre precisión y tiempo de predicción . . . . .	100
5.1.2	Resultados en IoT . . . . .	100
5.2	Discusión . . . . .	102
5.2.1	Principales hallazgos . . . . .	102
5.2.1.1	Software . . . . .	102
5.2.1.2	Hardware . . . . .	103
5.2.2	Comparación con trabajos previos . . . . .	103
5.2.2.1	Software . . . . .	103
5.2.2.2	Hardware . . . . .	104
5.2.3	Fortalezas y limitaciones de la propuesta . . . . .	105
5.2.3.1	Software . . . . .	105
5.2.3.2	Hardware . . . . .	106
5.2.4	Análisis Costo/Beneficio Del Proyecto . . . . .	106
<b>CONCLUSIONES</b>		<b>109</b>
<b>RECOMENDACIONES</b>		<b>110</b>
<b>TRABAJOS FUTUROS</b>		<b>111</b>
<b>REFERENCIAS</b>		<b>112</b>
<b>APÉNDICES</b>		<b>122</b>
5.1	Mapa de empatía . . . . .	122
5.2	Plan de Costos del Proyecto . . . . .	123
5.3	Código para el funcionamiento del dispositivo IoT . . . . .	124
5.4	Cuestionarios aplicados . . . . .	127
5.5	Comparación con otras placas electrónicas . . . . .	130
5.6	Artículo publicado en Scopus . . . . .	131

## Índice de figuras

1.1	Esquema general de la propuesta. . . . .	8
1.2	Esquema del funcionamiento (Software - Visión artificial). . . . .	10
1.3	Esquema del funcionamiento (Hardware - IoT). . . . .	11
2.1	Billete de S/ 10 Nueva Familia. . . . .	29
2.2	Billete de S/ 20 Nueva Familia. . . . .	30
2.3	Billete de S/ 50 Nueva Familia. . . . .	30
2.4	Billete de S/ 100 Nueva Familia. . . . .	30
2.5	Billete de S/ 10 Familia 2011. . . . .	31
2.6	Billete de S/ 20 Familia 2011. . . . .	31
2.7	Billete de S/ 50 Familia 2011. . . . .	32
2.8	Billete de S/ 100 Familia 2011. . . . .	32
2.9	Billete de S/ 200 Familia 2011. . . . .	32
2.10	Un conjunto de entrenamiento etiquetado para aprendizaje supervisado (por ejemplo, clasificación de spam). . . . .	35
2.11	Ejemplo de regresión. . . . .	36
2.12	Ejemplo de una red neuronal totalmente conectada. . . . .	37
2.13	Comparación entre neurona biológica y una neurona artificial. . . . .	37
2.14	Capas CNN con campos receptivos locales rectangulares. . . . .	38
2.15	Ejemplo de capa fully-connected. . . . .	39
2.16	Ejemplo del funcionamiento de una CNN. . . . .	42
2.17	Arquitectura de una VGGNet. . . . .	43
2.18	Arquitectura VGG-16 de un modelo VGG16. . . . .	52
2.19	Configuración de la CNN de VGGNet. . . . .	53
2.20	Arquitectura MobileNet. . . . .	54
2.21	Construcción de una arquitectura MobileNetV2. . . . .	55
2.22	Un Bloque Residual de una Red Neuronal Profunda Residual. . . . .	55
2.23	Representación intuitiva de la Función Residual. . . . .	56
2.24	Arquitectura ResNet50. . . . .	56
2.25	Convolución Separable por Profundidad Original. . . . .	57
2.26	Convolución Separable por Profundidad Modificada en Xception. . . . .	57
2.27	Arquitectura de una red Xception. . . . .	58
4.1	Captura de billetes completos (Anverso y Reverso). . . . .	66
4.2	Captura de detalles de billetes (Anverso y Reverso). . . . .	67
4.3	Captura de billetes en distintos entornos y condiciones de iluminación. . . . .	67
4.4	Porcentaje de cada denominación de billete en el conjunto de datos. . . . .	68
4.5	Comparación de la precisión entre transfer-learning y fine-tuning. . . . .	78
4.6	Diagrama de flujo del funcionamiento del dispositivo de asistencia. . . . .	82
4.7	Diseño de la carcasa del dispositivo. . . . .	84

4.8	Disposición de orificios en la parte superior. . . . .	85
4.9	Parte frontal del dispositivo. . . . .	85
4.10	Parte lateral del dispositivo. . . . .	86
4.11	Dispositivo final en funcionamiento. . . . .	87
4.12	Interfaz WinSCP para inicializar nueva conexión con Raspberry Pi 4B+. . . . .	88
4.13	Interfaz WinSCP para cargar archivos a Raspberry Pi 4B+ desde ordenador. . . . .	88
4.14	Valor medio por ítem. . . . .	93
4.15	Valoración por escala. . . . .	94
4.16	Comparación con otros productos. . . . .	94
4.17	Porcentaje de puntuación por criterio. . . . .	95
4.18	Promedio de ponderación a cada pregunta formulada. . . . .	96
4.19	Respuestas a preguntas positivas . . . . .	97
4.20	Respuestas a preguntas negativas . . . . .	97
5.1	Tamaño del modelo (parámetros y memoria requerida). . . . .	101
5.2	Mapa de empatía. . . . .	122
5.3	UEQ - Cuestionario sobre la experiencia de los Usuarios. . . . .	127
5.4	SUS - Escala de usabilidad del sistema. . . . .	128
5.5	Expresión de consentimiento informado. . . . .	129



## Índice de tablas

1.1	Indicadores . . . . .	6
4.1	Características del dispositivo de captura de imágenes billetes . . . . .	65
4.2	Tabla de Denominaciones Totales . . . . .	68
4.3	Configuración de hiperparámetros en los modelos . . . . .	77
4.4	Duración del entrenamiento y precisión de los modelos . . . . .	78
4.5	Requerimiento Funcional 1: Captura de Billetes . . . . .	79
4.6	Requerimiento Funcional 2: Identificación de Billetes . . . . .	80
4.7	Requerimiento Funcional 3: Salida de Audio . . . . .	80
4.8	Requerimientos no Funcionales . . . . .	81
5.1	Valor Actual Neto (VAN) . . . . .	106
5.2	Tasa Interna de Retorno (TIR) . . . . .	107
5.3	Comparación con otros productos del mercado. . . . .	108
5.4	Plan de Costos del Proyecto . . . . .	123
5.5	Comparación con otras placas electrónicas . . . . .	130

## Lista de Códigos

4.1	Importación de datos y procesamiento de imágenes . . . . .	70
4.2	Configuración del modelo . . . . .	73
4.3	Transferencia de Aprendizaje en el Entrenamiento . . . . .	74
4.4	Ajuste Fino para la Adaptación de Datos . . . . .	76
4.5	Función para capturar imagen . . . . .	89
4.6	Función para realizar la predicción . . . . .	90
4.7	Función para reproducir audios . . . . .	90
5.1	Código para el funcionamiento del dispositivo IoT . . . . .	124



## INTRODUCCIÓN

Según la Organización Mundial de la Salud (OMS, 2010), existen aproximadamente 285 millones de personas con discapacidad visual en todo el mundo, de los cuales 39 millones son completamente ciegos. Estas cifras alarmantes son respaldadas por la Unión Mundial de Ciegos (UMC, 2023), que señala que cerca de 253 millones de personas en todo el mundo presentan deficiencias visuales. En el Perú existen cerca de 160,000 personas invidentes y casi 600,000 que sufren alguna discapacidad visual que compromete su calidad de vida, desarrollo integral y bien familiar según el Ministerio de Salud (MINSA, 2014). En la región de Arequipa se registra un 3.4% (% con respecto a la población total) de personas con limitaciones para ver aún utilizando lentes, mientras que el porcentaje de personas con limitación visual ligera, moderada, grave y completa ascienden a 12.8%, 67.4%, 16.4% y 2.2% respectivamente (% con respecto al total de personas con esta limitación). Esta realidad evidencia los desafíos significativos que enfrentan las personas con discapacidad visual en su vida cotidiana.

De acuerdo con Lizárraga (2018), generar herramientas que promuevan la inclusión e independencia a personas que presentan discapacidad visual es una tarea necesaria y en donde la tecnología puede jugar un papel fundamental. En este contexto, la investigación tiene como objetivo construir un dispositivo de asistencia ergonómico y de bajo costo para reconocimiento de billetes basado en visión artificial e IoT para personas con discapacidad visual en la ciudad de Arequipa. Para este fin, se presenta el diseño, la implementación y la evaluación del dispositivo de asistencia propuesto, que tiene el potencial de mejorar la calidad de vida de las personas con discapacidad visual en Arequipa y servir como modelo para soluciones similares en todo el mundo.

El resto del documento se organiza de la siguiente manera.

En el Capítulo 1, se aborda el planteamiento del problema, el objetivo general y específico, las preguntas de investigación, así como la delimitación de la investigación. Además, se justifica la relevancia del estudio y se describe la solución propuesta de manera conceptual.

El Capítulo 2 se enfoca en una revisión exhaustiva del estado del arte, explorando herramientas, técnicas y algoritmos de visión artificial, así como dispositivos de asistencia similares. Además, se proporciona una base teórica que profundiza en las terminologías relevantes.

En el Capítulo 3, se desarrolla el marco metodológico, abordando los alcances y limitaciones de la investigación, así como su tipo y nivel. Se detalla la población y muestra, junto con los métodos, técnicas e instrumentos utilizados.

El Capítulo 4 se centra en el desarrollo de la propuesta y los resultados obtenidos.

Finalmente, en el Capítulo 5 se presenta el análisis y validación de los resultados, seguido de las conclusiones, recomendaciones y trabajos futuros.



# CAPITULO I

# Desarrollo de un dispositivo de asistencia ergonómico y de bajo costo para reconocimiento de billetes basado en visión artificial e IoT para personas con discapacidad visual en la ciudad de Arequipa

## ORIGINALITY REPORT

3%

SIMILARITY INDEX

5%

INTERNET SOURCES

2%

PUBLICATIONS

2%

STUDENT PAPERS

## PRIMARY SOURCES

1

[www.learndatasci.com](http://www.learndatasci.com)

Internet Source

1%

2

[sedici.unlp.edu.ar](http://sedici.unlp.edu.ar)

Internet Source

1%

3

[hdl.handle.net](http://hdl.handle.net)

Internet Source

1%

4

[cybertesis.unmsm.edu.pe](http://cybertesis.unmsm.edu.pe)

Internet Source

1%

5

[imt.mx](http://imt.mx)

Internet Source

1%

Exclude quotes Off

Exclude matches < 1%

Exclude bibliography Off

## PLANTEAMIENTO DE LA INVESTIGACIÓN

### 1.1. Planteamiento del problema

Según la OMS (2010), existen aproximadamente 285 millones de personas con discapacidad visual en todo el mundo, de los cuales 39 millones son completamente ciegos. Como señala la UMC (2023), cerca de 253 millones de personas en todo el mundo presentan deficiencias visuales.

De acuerdo con las cifras del MINSA (2014), en el Perú existen cerca de 160,000 personas invidentes y casi 600,000 que sufren alguna discapacidad visual que compromete su calidad de vida, desarrollo integral y bien familiar.

Asimismo, hasta el año 2013, el Instituto Nacional de Estadística e Informática (INEI, 2013) reportó que 801 mil personas presentan limitación de forma permanente para ver incluso haciendo uso de anteojos. De este total, el 52,6 % está en el área urbana y el 44,8 % en el área rural.

En este sentido, según la Primera Encuesta Nacional Especializada sobre Discapacidad llevada a cabo por el INEI (2014) en cooperación con el Ministerio de la Mujer y Poblaciones Vulnerables, y el Consejo Nacional para la Integración de la Persona con Discapacidad (CONADIS), el porcentaje de personas con limitaciones para ver aún utilizando lentes, en la ciudad de Arequipa es del 3,4 % (% con respecto a la población total), mientras que el porcentaje de personas con limitación visual ligera, moderada, grave y completa ascienden a 12,8 %, 67,4 %, 16,4 % y 2,2 % respectivamente (% con respecto al total de personas con esta limitación).

Estas cifras indican que las limitaciones de estas personas para realizar actividades diarias no pueden ser ignoradas. Por ello, generar herramientas que promuevan la inclusión e independencia a personas que presentan discapacidad visual es una tarea necesaria y en donde la tecnología puede jugar un papel fundamental. Mediante el uso de la tecnología es posible diseñar, desarrollar, implementar y gestionar soluciones (realizar herramientas con IoT y generando soluciones innovadoras) a diversos problemas sociales (Thomas, 2008, como se citó en Hernández, 2022). Incrementar, mantener o mejorar las capacidades funcionales de personas con discapacidad visual debe ser uno de los desafíos que se debe de afrontar. (Thomas, 2008)

Según la OMS (2019), todas las personas con ceguera o deficiencia visual grave que no pueden ser tratadas pueden llevar una vida independiente si tienen acceso a servicios de rehabilitación. Por ello, algunas de las opciones disponibles actualmente son: a) lupas ópticas, b) lectura en Braille, c) buscadores de los teléfonos inteligentes y d) entrenamiento de orientación y movilidad con bastones.

Como afirma Hernández (2022), en los últimos años la tecnología ha logrado tomar un rol importante en cuanto a la mejora de la calidad de vida de las personas con alguna discapacidad mediante las llamadas tecnologías de asistencia. Las tecnologías de asistencia permiten a las personas ser más productivas, más independientes y lograr una mejor calidad de vida. Beheshti et al. (2023) encontraron que, a pesar de los beneficios derivados de las tecnologías de asistencia (TA, Assistive technologies por sus siglas en inglés AT), algunas partes del mundo tienen un acceso mínimo o nulo a las TA, en muchos países de ingresos bajos y medios solo entre el 5 y 15 % de las personas que necesitan TA tienen acceso a ellas. En este marco, para Soderback (2015) un dispositivo de asistencia para la discapacidad visual se define como “cualquier artículo, pieza de equipo o sistema de producto, ya sea adquirido

comercialmente, modificado o personalizado, que se utiliza para aumentar, mantener o mejorar las capacidades visuales funcionales de una persona con discapacidad” .

En la actualidad existen herramientas de asistencia para personas con discapacidad visual, que se enfocan en resolver problemas de desplazamiento, movilidad, orientación y reconocimiento de objetos. Muchas de estas herramientas han quedado como propuestas, proyectos o prototipos, pero ninguno ha sido comercializado. En Fernández (2021) se propone el desarrollo de un bastón sensorial geolocalizador inteligente para apoyar en el desplazamiento de personas invidentes. Del mismo modo, se ha encontrado en Lizárraga (2018) una propuesta de diseño de bastón electrónico para mejorar la calidad de vida del desplazamiento diario de personas invidentes. Por otro lado, en Vargas (2020) proponen el desarrollo de un prototipo de gafas vibratorias basado en Arduino que contribuye a la detección de obstáculos. En cambio, en Afif et al. (2021) proponen el desarrollo de un sistema para el reconocimiento de objetos baso en técnicas para el aprendizaje por transferencia. Por último, en López et al. (2006), desarrollan un prototipo para la identificación de colores de objetos mediante visión artificial y lógica difusa. De acuerdo con Gutiérrez (2019), efectivamente hay una escasez de tecnología y herramientas que sean ergonómicas y de bajo costo, además que se encuentren orientadas al apoyo y asistencia a personas que presentan discapacidad visual, en actividades como el reconocimiento de objetos. Conforme a Vasquez Salazar y Cardona Mesa (2019), esta situación radica a partir de los costos elevados de las herramientas existentes, la disponibilidad limitada de herramientas ergonómicas y la falta de características de ergonómicas en las herramientas ya existentes. Por consecuencia, Basterrechea y et al. (2011) afirma que estas causantes repercuten en la integración de las personas con discapacidad visual, el acceso de herramientas apropiadas y la dificultad en la realización de actividades diarias.

La mayoría de las soluciones existentes se orientan hacia la facilitación del desplazamiento seguro de personas con discapacidad visual y el reconocimiento de objetos en su entorno. En este contexto, la presente investigación plantea la creación de un dispositivo de asistencia destinado a personas con discapacidad visual, empleando algoritmos de visión artificial y tecnologías de IoT. Para abordar específicamente esta solución, el enfoque principal se centra en el reconocimiento de billetes, identificando así una brecha significativa en este ámbito.

## 1.2. Objetivos de la Investigación

### 1.2.1. Objetivo General

Construir un dispositivo de asistencia ergonómico y de bajo costo para reconocimiento de billetes basado en visión artificial e IoT para personas con discapacidad visual en la ciudad de Arequipa.

### 1.2.2. Objetivos Específicos

1. Analizar la situación actual de las personas con discapacidad visual, herramientas de asistencia, dispositivos de IoT, técnicas y algoritmos de Visión Artificial.
2. Crear el conjunto de datos de imágenes de billetes peruanos.
3. Implementar el módulo de visión artificial para la captura y procesamiento de imágenes en tiempo real.
4. Integrar módulo de IoT con técnicas de Visión Artificial.
5. Validar la integración del sistema.
6. Evaluar el sistema con personas invidentes de la Asociación Unión de Ciegos de Arequipa.

## 1.3. Preguntas de Investigación

1. ¿Será posible construir un dispositivo de asistencia ergonómico y de bajo costo para reconocimiento de billetes basado en visión artificial e IoT para personas con discapacidad visual en la ciudad de Arequipa?
2. ¿De qué manera el análisis de la situación actual de las personas con discapacidad visual, herramientas de asistencia, dispositivos de IoT, técnicas y algoritmos de Visión Artificial, ayuda a comprender el estado actual de la problemática identificada?
3. ¿Cómo la creación de un conjunto de datos de imágenes de billetes contribuye en el reconocimiento de billetes?
4. ¿La implementación del módulo de visión artificial contribuye al correcto funcionamiento del dispositivo de asistencia?
5. ¿La integración del módulo de IoT con técnicas de visión artificial contribuye al correcto reconocimiento de billetes?
6. ¿La integración del sistema con los componentes de IoT y visión artificial contribuye al correcto funcionamiento del dispositivo de asistencia?

## 1.4. Hipótesis, Variables e Indicadores

### 1.4.1. Hipótesis

La construcción de un dispositivo de asistencia ergonómico y de bajo costo para personas con discapacidad visual aumentará significativamente su capacidad para reconocer billetes.

### 1.4.2. Variables de investigación

- **Variable Dependiente:** Capacidad de las personas con discapacidad visual para reconocer billetes.
- **Variable Independiente:** Construcción del dispositivo de asistencia ergonómico y de bajo costo.

### 1.4.3. Indicadores

Tabla 1.1

*Indicadores*

Variable	Dimensión	Indicadores	Unidad de medida
<b>Variable Dependiente:</b> Capacidad de las personas con discapacidad visual para reconocer billetes.	Desempeño del Sistema	Precisión del reconocimiento de billetes	Porcentaje (%)
	(Reconocimiento de billetes)	Tasa de falsos positivos/negativos	Porcentaje
		Tiempo de procesamiento	Milisegundos (ms)
	Experiencia del Usuario (Facilidad de uso y diseño ergonómico)	Cuestionario sobre la experiencia de los Usuarios	Likert (1-7)
		Escala de usabilidad del Sistema	Likert (1-5)
<b>Variable Independiente:</b> Construcción del dispositivo de asistencia ergonómico y de bajo costo	Evaluación de Modelos de Aprendizaje Profundo	Precisión del modelo para el reconocimiento de billetes	Porcentaje (%)
		Tiempo de inferencia del modelo	Milisegundos (ms)
		Tamaño del modelo (parámetros y memoria requerida)	Megabytes (MB)
		Relación entre precisión y tiempo de predicción	Gráfico y análisis
	Eficiencia Energética	Duración de la batería	Horas
Ergonomía		Diseño ergonómico	NA
		Dimensiones del dispositivo	Milímetros
Costo		Costo de materiales	Soles Peruanos
		Costo de fabricación	Soles Peruanos

## 1.5. Línea y Sublínea de Investigación

1. Línea de Investigación: Inteligencia Artificial
2. Sublínea de Investigación: Visión Artificial

## 1.6. Palabras Clave

Visión Artificial; IoT; Aprendizaje Profundo; Discapacidad Visual; Tecnologías de asistencia, Tecnologías vestibles; Dispositivo Ergonómico y de Bajo Costo.

## 1.7. Solución Propuesta

### 1.7.1. Justificación e Importancia

Según la Organización de las Naciones Unidas (ONU, 2014), gran parte de la Convención sobre los derechos de las personas con discapacidad se basa en el acceso a bienes y servicios, como las tecnologías de apoyo, la atención de la salud y la educación. Estos servicios han de ser accesibles a las personas con discapacidad para que estas disfruten de sus derechos de igualdad de condiciones con las demás. Para ello puede ser necesario contar con servicios especializados en discapacidad, en tanto que otras veces tal vez sea menester que los servicios generales (por ejemplo, la educación) sean accesibles a esas personas.

Hoy en día en la ciudad de Arequipa no se encuentran soluciones tecnológicas para apoyar a personas invidentes en el reconocimiento de billetes, lo cual les resulta necesario al desempeñar sus actividades cotidianas, tal y como nos explican miembros de la Asociación Unión de Ciegos de Arequipa, que es una de las pocas asociaciones en la ciudad que junta a personas mayores de edad (18-55 años) con discapacidad visual para realizar actividades de integración y en algunos casos realizar actividades de rehabilitación. Para poder recabar información acerca de las necesidades de las personas con discapacidad visual, se tuvo una reunión en la cual nos dieron a conocer de cerca sus necesidades y acerca de las tecnologías que utilizan actualmente, las cuales en muchos casos no cubren por completo sus necesidades.

En este sentido, en colaboración con la Asociación Unión de Ciegos de Arequipa, la motivación de esta investigación radica en poder ayudar a las personas con discapacidad visual, brindándoles una mejor calidad de vida mediante el desarrollo de un dispositivo tecnológico ergonómico y de bajo costo que pueda suplir sus necesidades y les permita realizar sus actividades cotidianas.

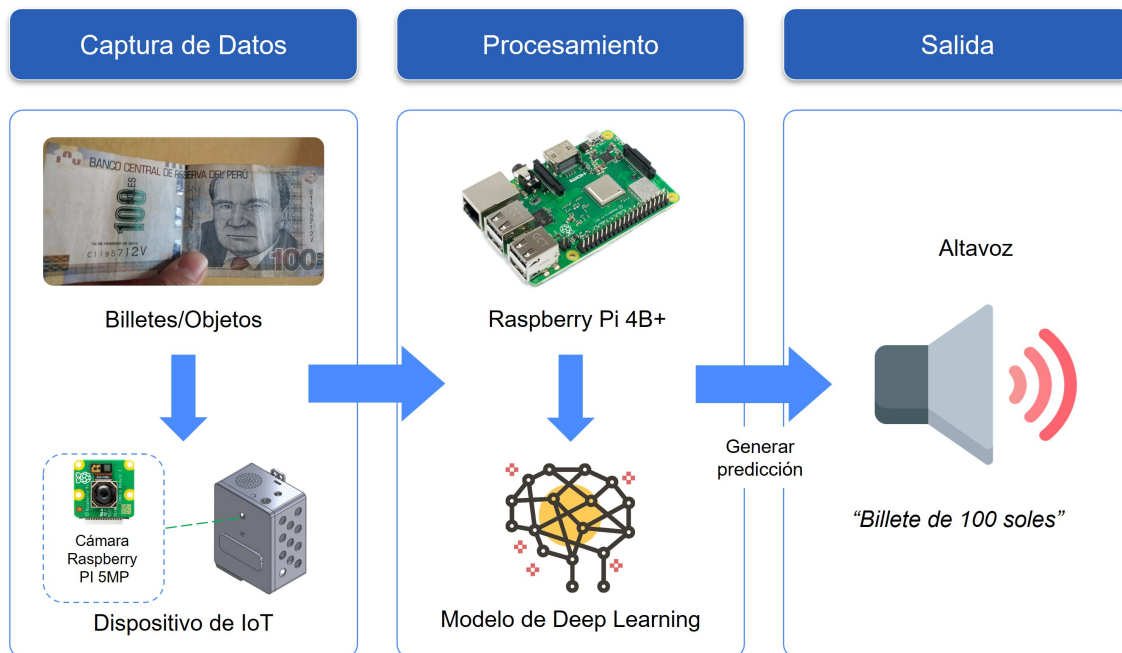
## 1.7.2. Descripción de la Solución

A continuación, se describe la solución propuesta:

### 1.7.2.1. Esquema general de la propuesta

En la figura 1.1, se presenta en diagrama de bloques del esquema general de propuesta, es decir, la secuencia de funcionamiento del dispositivo de reconocimiento:

**Figura 1.1:** Esquema general de la propuesta.



El proceso de funcionamiento del dispositivo se divide en tres etapas principales: Captura, Procesamiento y Salida de datos. A continuación, se presenta una descripción detallada de cada etapa:

1. **Captura de Datos:** En esta etapa, se emplea una cámara de alta resolución (Cámara Raspberry Pi de 5MP) para capturar imágenes en tiempo real de los billetes. Esta cámara actúa como el "ojo" del dispositivo, permitiendo la adquisición precisa de información visual del entorno circundante. La captura de imágenes se realiza de manera continua, garantizando una entrada constante de datos para su posterior análisis.
2. **Procesamiento de Datos:** La etapa de procesamiento de datos desempeña un papel crucial en la interpretación y comprensión de la información capturada. En primer lugar, las imágenes se someten a un proceso de preprocesamiento, que abarca el ajuste del tamaño y la eliminación de ruido. Estas optimizaciones preparan las imágenes para un análisis más preciso. A continuación, se utiliza un modelo de aprendizaje profundo previamente entrenado, que se ejecuta en una placa electrónica de computadora Raspberry Pi 4B+. Este modelo es capaz de identificar y distinguir entre diferentes billetes con una alta tasa de precisión. La elección de un modelo de aprendizaje profundo permite una interpretación avanzada de las imágenes, lo que garantiza resultados confiables y coherentes. (Shahira y Lijiya, 2021)

3. **Salida de Datos:** Una vez que las imágenes han sido procesadas y los billetes han sido identificados, se procede a la generación de la salida. Para lograrlo, se emplea un algoritmo de conversión de texto a voz (Text-to-Speech) que transforma los resultados del algoritmo en salidas de audio claras y comprensibles mediante un módulo Adafruit Max98357. Esto permite que las personas con discapacidad visual puedan obtener la información de manera auditiva, lo que contribuye a una experiencia más inclusiva y accesible.

Este proceso integral, que abarca desde la captura hasta la conversión de resultados en audio, asegura que las personas con discapacidad visual en la ciudad de Arequipa puedan acceder y comprender la información sobre billetes de manera efectiva y eficiente. La combinación de tecnologías de visión artificial e IoT ofrece una solución poderosa y de bajo costo para mejorar la independencia y calidad de vida de este grupo de personas.

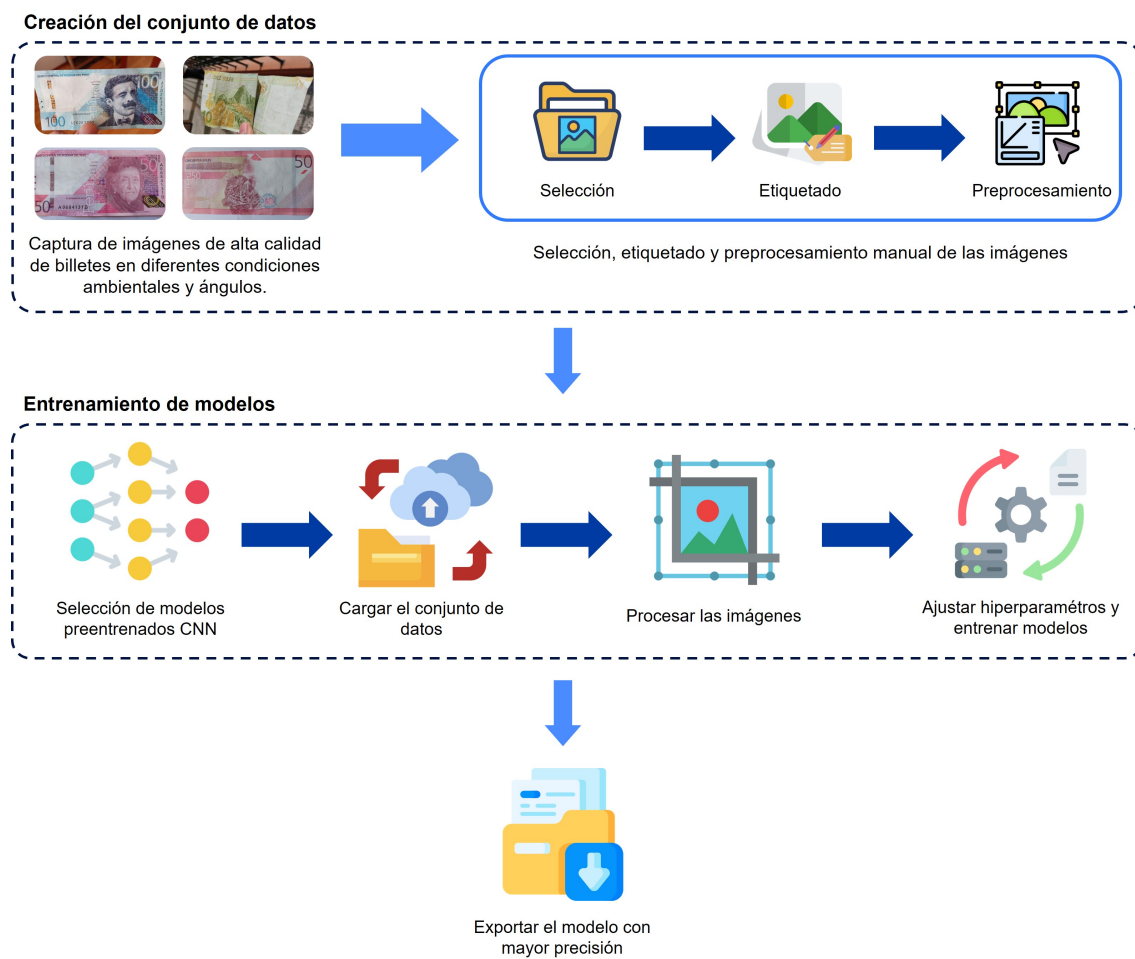
#### 1.7.2.2. Esquema del funcionamiento (Software - Visión artificial)

En la figura 1.2, se muestra el esquema del Software de la propuesta, que se compone principalmente de dos etapas fundamentales:

1. **Creación del Conjunto de Datos:** La primera etapa se enfoca en la obtención de imágenes de alta calidad de billetes, considerando diversas condiciones y ángulos de captura. Posteriormente, se procede a la selección, etiquetado y preprocesamiento manual de estas imágenes. En la fase de selección, se realiza una depuración de las imágenes, eliminando aquellas con problemas de captura o ruido, y se distribuyen en carpetas específicas. El etiquetado se orienta a renombrar las imágenes según su ubicación y el conjunto de datos a la cual pertenecen. Finalmente, el preprocesamiento se concentra en el redimensionamiento de las imágenes para que sean compatibles con los modelos de aprendizaje profundo.
2. **Entrenamiento de los Modelos:** La segunda etapa se enfoca en el entrenamiento de los modelos. En esta etapa, se inicia seleccionando los modelos preentrenados más adecuados para la tarea. Luego, se integra el conjunto de datos generado en la etapa anterior. La siguiente fase implica el procesamiento de las imágenes de acuerdo con los parámetros definidos por los algoritmos, seguido por el ajuste de los hiperparámetros de los algoritmos para seguidamente llevar a cabo el proceso de entrenamiento.

Finalmente, se exporta los pesos del modelo que obtenga la mayor precisión durante la fase de validación y pruebas.

**Figura 1.2:** Esquema del funcionamiento (Software - Visión artificial).

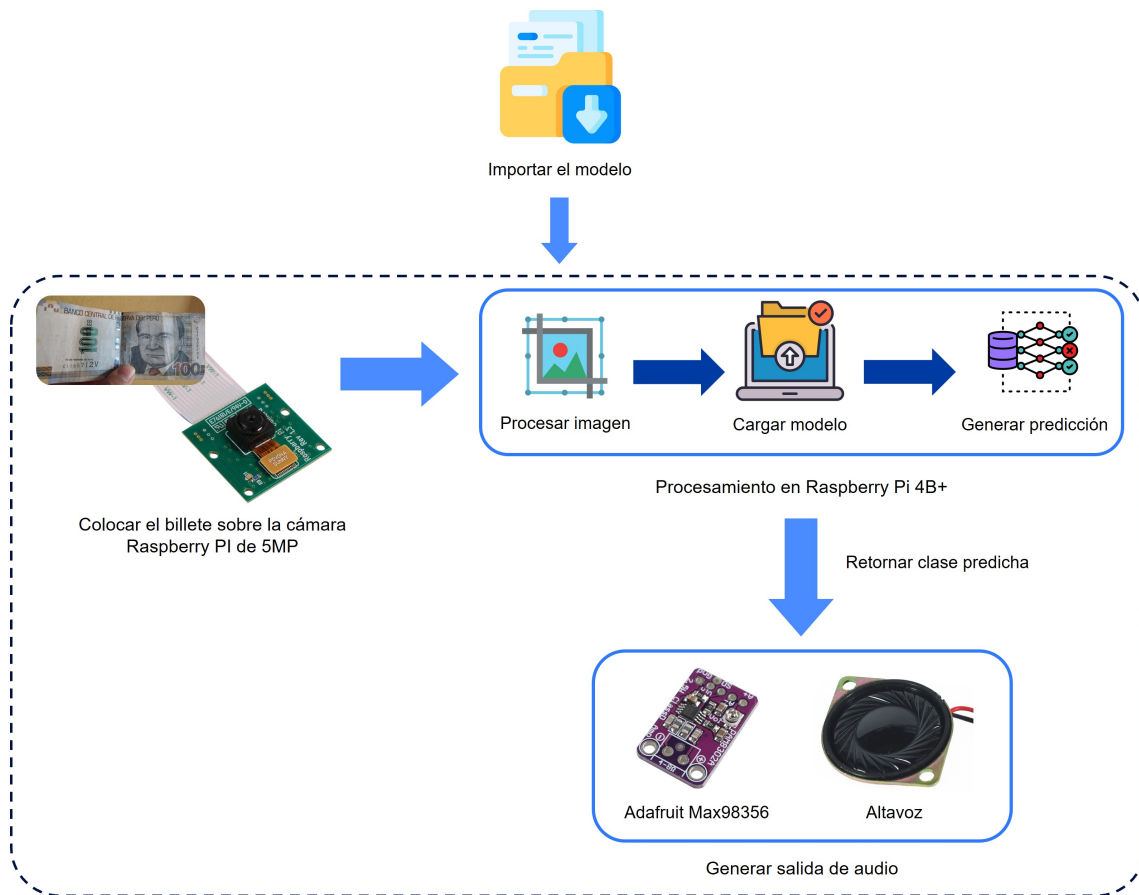


### 1.7.2.3. Esquema del funcionamiento (Hardware - IoT)

En la figura 1.3, se presenta el esquema del Hardware de la propuesta.

En primer lugar, importamos el modelo resultante de la etapa anterior. El dispositivo captura imágenes a través de una cámara de alta resolución, en este caso, la cámara Raspberry PI de 5MP. Estas imágenes se procesan en una placa electrónica de computadora Raspberry Pi 4B+. El procesamiento comienza con la redimensión y el preprocesamiento de las imágenes, siguiendo los parámetros establecidos por los algoritmos. Posteriormente, se carga el modelo para generar una predicción. El resultado del procesamiento se traduce en una clase, que luego se utiliza como entrada para producir una salida de audio a través de un módulo Adafruit Max98357 y un altavoz.

*Figura 1.3: Esquema del funcionamiento (Hardware - IoT).*



### 1.7.3. Aporte

El aporte de la investigación se distingue en varios aspectos clave:

1. **Utilización de Modelos de Aprendizaje Profundo Preentrenados:** La investigación demuestra cómo los modelos de Aprendizaje Profundo preentrenados pueden ser una herramienta valiosa para la detección precisa de billetes. Esta aplicación innovadora demuestra cómo la tecnología puede ser adaptada para abordar desafíos específicos, mejorando significativamente la independencia y la calidad de vida de las personas con discapacidad visual.
2. **Caracterización de la Complejidad de Procesamiento:** La investigación cuantifica la complejidad del procesamiento requerido por los modelos de aprendizaje profundo preentrenados. Ello proporciona una comprensión más profunda de los requisitos computacionales y permite una planificación más efectiva de los recursos, lo que es crucial para la implementación exitosa del sistema de asistencia.
3. **Diseño de un Sistema de Asistencia Ergonómico y de Bajo Costo:** La construcción de un sistema de asistencia ergonómico y económico, basado en IoT y visión artificial, constituye una contribución significativa. Esta solución tecnológica asequible y accesible puede tener un impacto directo en

la vida cotidiana de las personas con discapacidad visual, abriendo nuevas posibilidades para su inclusión y autonomía.

4. **Evaluación Empírica de la Precisión:** A través de experimentos reales, se evalúa empíricamente la precisión de sistema desarrollado. Esta evaluación proporciona resultados concretos sobre el rendimiento del sistema en situaciones del mundo real, lo que respalda la validez y la utilidad práctica del enfoque propuesto.

En última instancia, el desarrollo de este proyecto de tesis no solo aborda un problema tangible que enfrentan las personas con discapacidad visual en la ciudad de Arequipa, sino que también abre las puertas a un futuro más inclusivo y autónomo para este grupo de personas. La tecnología implementada tiene el potencial de mejorar su calidad de vida y empoderarlos para enfrentar sus actividades diarias con mayor independencia y confianza.





## CAPITULO II

## FUNDAMENTOS TEÓRICOS

### 2.1. Antecedentes de investigación

A lo largo de la historia se han desarrollado diversas herramientas para ayudar a las personas invidentes a desenvolverse en su entorno, como los bastones blancos, la escritura braille y el perro lazarillo. De acuerdo con la Organización Nacional de Ciegos Españoles (ONCE, 2011) e Ipland y Parra (2009), estas herramientas y recursos tradicionales han llegado a ser símbolo de las personas invidentes y los han acompañado por muchos años. Sin embargo, Paz y Sandoval (2013), afirman que con la globalización muchos aspectos de la vida han cambiado y con ello han surgido brechas para la integración de las personas invidentes en la sociedad, volviendo dichas herramientas obsoletas. En la actualidad, el desarrollo de herramientas de asistencia que pueda suplir las necesidades de los invidentes se convierte en un desafío constante, puesto que, no se llega a cubrir necesidades como la comodidad y accesibilidad. En concordancia con Torres y Rodríguez (2012), muchas de las herramientas de asistencia existentes son costosas y complejas de usar, por lo que mayormente se quedan en desuso.

En la investigación realizada por Ayala (2011), se diseñó un dispositivo electrónico capaz de detectar obstáculos, mediante señales ultrasónicas y haciendo uso de materiales de bajo costo y gran rendimiento, con el objetivo de poder ayudar en la orientación y movilidad de las personas que presentan discapacidad visual. El valor agregado de la propuesta radica que dicho dispositivo será adherible a un bastón blanco, por lo que no reemplazará a esta herramienta tradicional sino más bien se convierte en un accesorio complementario. Gracias al uso del sensor ultrasónico se ha observado que el dispositivo propuesto es capaz de detectar obstáculos con una precisión aceptable, además de ser de bajo costo y accesible. Asimismo, en Avila (2017) proponen la construcción de un bastón electrónico mediante Arduino, con su respectiva aplicación móvil, con el fin de suplir la necesidad de movilización de las personas no videntes. Los resultados de investigación indican que la integración del Arduino y los sensores para el desarrollo de la propuesta han impactado significativamente en la construcción de un dispositivo de bajo costo y práctico de usar, además ha facilitado su integración al bastón tradicional como complemento. El valor agregado de esta propuesta radica en la detección de obstáculos y emisión de vibraciones y sonidos de alerta mediante microcontroladores que son adaptados al bastón tradicional y configurados de acuerdo con el medio físico.

Por otro lado, en Dávila (2016) se propone el desarrollo de un bastón inteligente a través de un diseño amigable, personalizable y adaptable para personas con discapacidad visual. El desarrollo de la propuesta implicó un análisis morfológico del usuario no vidente, con el fin de poder extraer las necesidades principales que este posee. Dentro de las herramientas utilizadas para la construcción del bastón electrónico se encuentra la cámara Microsoft Kinect V2 Sensor, la cual ha permitido tener una alta precisión y mayor interacción entre el usuario no vidente y el espacio 360° que lo rodea. La integración de la cámara Microsoft Kinect V2 Sensor y el análisis morfológico previo, han sido factores importantes para brindar al usuario con discapacidad visual una experiencia agradable y personalizada al momento de hacer uso de esta herramienta de asistencia.

En Ortiz (2012) se propone un sistema de identificación de objetos para personas

invidentes haciendo uso de la tecnología RFID, el cual permite el almacenamiento y recuperación de los datos. Dicho sistema consta de un dispositivo móvil desarrollado mediante varios componentes electrónicos y de un software para PC que fue implementado con el lenguaje de programación Visual C# 2010 Express. En este contexto, los autores demuestran que el uso de RFID ha permitido cubrir con la necesidad de identificación de objetos, proporcionándoles a las personas invidentes un dispositivo de IOI capaz de reconocer ciertas características de objetos en su entorno mediante etiquetas de RFID. Por ello, este tipo de dispositivos tiene previsto utilizarse en escenarios de compra y venta de productos en supermercados, tiendas y almacenes, gracias a las funcionalidades que presenta como duración de suministros de energía, distancias y cobertura de lectura de etiquetas de RFID y la capacidad de almacenamiento.

Atendiendo a estas consideraciones, se puede deducir que la construcción de un dispositivo como el bastón electrónico mediante uso de herramientas de Arduino y sensores de ultrasonido y motores de vibración, han permitido mejor la adaptabilidad y configuración en superficies o medios físicos disponibles, convirtiendo estas propuestas como accesorios complementarios del bastón tradicional más no como un reemplazo de dicha herramienta. Además, para el desarrollo de sistemas de identificación de objetos u obstáculos ha sido necesario la integración de tecnologías emergentes que en su momento fueron las etiquetas de RFID y la cámara de Microsoft Kinect V2 Sensor. Estas tecnologías han permitido tener mayor funcionalidad respecto a la interacción con el usuario, facilidad en la realización de tareas y diseño amigable; de esta forma contribuyendo a cubrir la necesidad de un sector vulnerable, las personas invidentes.

Peralta y Urmendiz (2014) proponen el diseño de un prototipo funcional sobre un sistema guía de asistencia para personas invidentes conectado de manera inalámbrica con el dispositivo móvil del usuario, el cual facilite el desplazamiento de manera segura y confiable. El valor agregado de esta investigación consiste en la adaptación de un bastón tradicional, implementando un sistema de detección de obstáculos en tres niveles, inferior, frontal y superior, con el fin de facilitar la ubicación espacial de las personas invidentes; también manejando un componente que permite a los usuarios comunicarse con sus parientes por redes sociales. Los beneficios que proporciona esta propuesta consisten en el bajo costo de adquisición y el uso apropiado de la tecnología a favor de las personas invidentes. La propuesta es desarrollada sobre un escenario abierto por lo que se hará necesario comprobar o realizar pruebas en otros escenarios, ya que podría no tener el mismo resultado positivo o implicaría estudiar otro tipo de necesidades.

Por otra parte, en Hernandez et al. (2011) se plantean el desarrollo de un dispositivo tecnológico para la optimización del tiempo de aprendizaje del lenguaje braille en personas invidentes. Está investigación tiene un enfoque distinto, se orienta más en el ámbito de la educación, específicamente en el aprendizaje inicial de la escritura braille. Los resultados de la investigación radican en el fácil acceso y comprensión, mediante el sistema de síntesis de voz que genera palabras de manera artificial, con el objetivo de verificar si la escritura braille es correcta o no. El dispositivo electrónico propuesto, también permite una interacción positiva con el usuario invidente, y reduce de manera significativa el tiempo de aprendizaje gracias al sintetizador de voz que este proporciona que lo hace más fácil de conocer, localizar e identificar. Sin embargo, se debería tener en cuenta que dicha propuesta solo se aplica en el

contexto de aprendizaje de la escritura braille más no en otros tipos de aprendizaje, por lo que puede ser limitada y pocos beneficioso en otros aspectos.

## 2.2. Estado del Arte

En Shah et al. (2023) se hace uso de un conjunto de datos COCO, que contiene aproximadamente 123.287 imágenes etiquetadas a mano y clasificadas en 80 categorías diferentes. El objetivo del trabajo radica en detectar objetos cotidianos comunes y producir respuesta de voz para informar al usuario sobre la ubicación del objeto, mediante el uso del algoritmo de YOLOv4 y el módulo gTTs, respectivamente. La metodología utilizada consistió en cargar los pesos de los datos pre-entrenados de YOLO-v4-tiny, entrenarlos en el conjunto de datos COCO utilizando el módulo OpenCV Deep Neuronal Network (DNN). Luego, hacer uso de este módulo para la captura de los datos de la cámara y fotografías. Por último, usar el elemento reconocido para determinar su posición actual, dar como salida el texto traducido en el idioma local y emitir un segmento de audio mediante el módulo gTTs. Los resultados obtenidos demuestran que el algoritmo YOLOv4-tiny logra una mejor precisión (MAP) de 40.2 y un valor aceptable de fotografías procesados por segundo (FPS) de 330, a comparación de los algoritmos como YOLOv3-tiny y YOLOv4-tiny MAP. De esta forma, concluyendo que efectivamente el sistema inteligente propuesto es capaz de detectar objetos cotidianos comunes y produce respuesta de voz de la ubicación del objeto, con una precisión aceptable.

Asimismo, en Rusli et al. (2022) se hace uso del conjunto de datos pre-entrenados MS COCO 2017. El objetivo de la investigación consistió en desarrollar un sistema de navegación independiente para personas con discapacidad visual, de bajo costo, portátil y fácil de usar. La metodología se basó en hacer uso del modelo de aprendizaje YOLOv4-tiny debido a la simplicidad, confiabilidad, precisión, velocidad y conveniencia para balancear la precisión y velocidad de respuesta en tiempo real. Para luego, convertido y cuantificado en el modelo TensorFlow Lite, el cual funciona en Deep Simple Online y Real-time Tracking (DeepSORT) junto al reconocimiento de señales. Los resultados demuestran que el sistema propuesto posee un desempeño técnicamente aceptable y es compacto para usarlo, puesto que es independiente de los servicios de internet y eficiente en el uso de los recursos. La conclusión que llegaron fue que, para mejorar la experiencia del usuario, es necesario integrar algunas mejoras posibles en el diseño físico y método de salida de información. Gunupudi et al. (2021) hacen uso del algoritmo YOLO. El objetivo de la investigación radica en desarrollar un sistema basado en YOLO para el reconocimiento de escena y narración de vídeo mediante el uso de técnicas de aprendizaje profundo, para reducir los accidentes y facilitar la vida de las personas ciegas. La metodología del sistema consiste en reconocer y entender la escena presentada, para luego de forma dinámica dar una salida verbal. Los hallazgos demuestran el empleo de la técnica de bloques de construcción densos es ampliamente accesible para aproximarse a la estructura dispersa ideal predicha, de esta forma siendo esta técnica factible para mejorar las redes neuronales en la visión por computadora. Finalmente, concluyendo que es posible detectar objetos con mayor precisión e identificar los elementos de forma individual con la ubicación geográfica específica de los objetos en la imagen.

De la misma manera, en Mukhiddinov et al. (2022) se desarrolla un sistema orientado a las personas de las Islas Vírgenes Británicas y haciendo uso del modelo YOLOv4. El objetivo de la investigación radica en el desarrollo de un sistema de detección y notificación de incendios en tiempo real, como solución ante las dificultades que presentan las personas ciegas y deficientes visuales, de las Islas Vírgenes Británicas, ante escenarios de incendios en interiores. La metodología utilizada consistió en usar modelos de CNN profundos y un detector de objetos YOLOv4, que fue mejorado con un módulo de atención de bloque convolucional. Dentro de los resultados se demuestra que YOLOv4 mejorado, es más robusto y funciona ligeramente mejor que la versión YOLOv4, sobre el conjunto de datos de detección de incendios, con un 48.6 y 47.2 de AP, respectivamente. De esta forma, concluyendo la información brindada por este sistema ha sido útil para las personas de deficientes visuales a la hora de diferenciar los incendios cotidianos y navegar durante la evaluación.

De acuerdo con las investigaciones presentadas Shah et al. (2023), Rusli et al. (2022) y Mukhiddinov et al. (2022), es posible afirmar que el algoritmo YOLOv4 en combinación con conjuntos de datos pre-entrenados como MS COCO, permiten el desarrollo eficaz y robusta de sistemas de visión artificial para diversas aplicaciones como la detención de objetos en tiempo real. En, Shah et al. (2023), el algoritmo YOLOv4-tiny aplicado al conjunto de MS COCO, logra una precisión significativa con 17.6 % y una adecuada velocidad de procesamiento (FPS) de 330 FPS. Así mismo, en Rusli et al. (2022), haciendo uso del conjunto de datos de MS COCO, el algoritmo YOLOv4 logra .97 en precisión en la tarea de reconocimiento de señales de alto, .84 en reconocimiento de reloj, .79 en reconocimiento de autos, .76 en reconocimiento de personas y .69 en reconocimientos de ratones. Además, consigue un 2-5 en velocidad de procesamiento (FPS). Por otro lado, en Mukhiddinov et al. (2022) tras mejorar el algoritmo YOLOv4, se consigue una precisión de 72.6 %, en tareas de detección de incendios, a diferencia del algoritmo YOLOv3-tiny que posee una precisión de 43.9 % y YOLOv4-tiny con un 51.5 %. En general, los resultados demuestran que la integración de YOLOv4 con otras tecnologías como modelos de redes neuronales propuesto por Mukhiddinov et al. (2022), técnicas de aprendizaje profundo propuesto por Rusli et al. (2022), y el módulo gTTs propuesto por Shah et al. (2023), hace que sea más efectivo la detección y reconocimiento de objetos con mayor grado de asertividad. Sin embargo, también es posible destacar algunas de las limitaciones y posibles mejoras para el desarrollo de sistemas de detección de objetos como en el modelo de diseño físico y la salida de información para mejorar la experiencia del usuario y ofrecer una ergonomía al momento que haga y uso de este tipo de herramientas de asistencia (Rusli et al., 2022). En conclusión, el uso del algoritmo de aprendizaje YOLOv4 presenta grandes posibilidades para el desarrollo de herramientas de asistencia que sea inteligente, orientado a ayudar a las personas que presenta discapacidad o deficiencia visual en distintos ámbitos.

En Morparia et al. (2022), se propone una descripción general de varios sistemas desarrollados para ayudar a personas con discapacidad visual en la tarea de reconocimiento facial. El objetivo de esta investigación radica en proporcionar una descripción general del estado del arte en el contexto de reconocimiento de personas. La metodología empleada para la investigación consistió en destacar las fortalezas y debilidades de diferentes soluciones, además de analizar algunos problemas que deben de afrontarse y resolverse para acelerar el despliegue práctico y la aceptación generalizada de sistemas orientada a personas con discapacidad visual. Los

resultados obtenidos, indican que en los últimos años se han llegado a desarrollar varios sistemas prototipo orientado en ayudar a las personas deficiente visuales en la tarea de reconocimiento facial. De esta forma, concluyendo que el desarrollo de dichos sistemas para el reconocimiento facial es viable. Sin embargo, aún sigue siendo necesario abordar varios problemas como desafíos como la usabilidad, practicidad, portabilidad y operabilidad en tiempo real.

Por otro lado, Yuan et al. (2022) desarrollan un sistema orientado a ayudar a personas ciegas y con discapacidad visual. El objetivo del trabajo radica en proporcionar una evaluación completa y detallada de la computación perimetral para la detección de objetos con conectividad mmWave y sub-6 GHz en una aplicación con requisitos de alto ancho de banda y baja latencia. La metodología consistió en una evaluación de la visión combinada con una simulación de red inalámbrica de pila completa que permite poder determinar la distribución de rendimientos y retrasos con ruta de navegación reales y trazados a partir de nuevos modelos 3D de alta resolución en un escenario urbano. Los resultados de la investigación sugieren que la combinación de la computación perimetral móvil de baja latencia y los sistemas de bandas tradicionales de menos de 6 GHz, mejora la precisión y rango de detección. Sin embargo, aún es un desafío poder cumplir con los requisitos de extremo a extremo. De esta forma, llegando a la conclusión que las simulaciones inalámbricas brindan un alto nivel de realismo y pueden identificar las limitaciones claves en la computación perimetral de alta velocidad de datos. Además, que la computación perimetral móvil, junto con las capacidades de alta velocidad de datos mmWave, es una gran promesa para acceder a potentes análisis de vídeo por medio de dispositivos portátiles.

Valipoor y de Antonio (2022) proponen una revisión sistemática de mapas orientada en el reconocimiento de objetos y detección de obstáculos, como solución a la asistencia a personas ciegas. El objetivo de la presente investigación es proporcionar una guía para que los investigadores en el campo de la visión artificial para que comprendan los avances durante los últimos cuatro años y medio. La metodología utilizada consistió en proporcionar una descripción sobre los desafíos actuales y una comparación entre diferentes soluciones para indicar los pros y contras. Dentro de los resultados se menciona que existe una cantidad considerable de personas que se sienten incomodas al estar expuestas a dispositivos de asistencia que poseen una cámara, debido que amenaza la privacidad de los usuarios y las personas del entorno, de esta manera sugiriendo poner mayor hincapié a las razones por las que los usuarios abandonan las tecnologías de asistencia. De esta manera, llegando a la conclusión que existen soluciones que demuestran ser capaces de detectar personas, objetos u obstáculos, pero que los beneficios esperados por lo usuarios resultan ser vagos, puesto que la solución aplicada en un escenario simulado puede no tener la misma eficacia en un contexto diferente, además los bajos costos y la adaptación de una solución que sea específica para aplicación en contextos diferentes se vuelven factores limitantes.

En cuanto a Mouna Afif y de Antonio (2021), proponen el desarrollo de un sistema de identificación de puntos de referencia en interiores, eficiente y robusto basado en una red neuronal convolucional ligera (DCNN). La metodología consistió en hacer el uso de técnicas de aprendizaje por transferencia como MobileNet v1, v2 e inception v3 DCNN, para luego aplicarlos en el conjunto de datos MCIndoor. Los resultados demuestran que dicho sistema es genérico y lo suficiente eficaz como para

manejar la gran variación intraclase proporcionada en los conjuntos de entrenamiento y prueba. De esta forma llegando a la conclusión, el reconocimiento y clasificación de objetos en interiores presentan una tarea crucial en la inteligencia artificial para ayudar a las personas ciegas e invidentes a desarrollar sus actividades cotidianas.

Por otra parte, Vorapatratorn et al. (2021) plantean el desarrollo de un novedoso sistema de detección de obstáculos y navegación para invidentes utilizando cámaras estereoscópicas con técnicas de aprendizaje automático. La metodología de la investigación consistió en comparar el rendimiento de varios algoritmos de aprendizaje automático, por ejemplo, ANN, SVM, Naïve Bayes, Decision Tree, k-NN y Deep Learning, entre la precisión de la clasificación y la velocidad de la predicción. Los resultados demuestran que el uso ANN con el H-DAI y V-DAI alcanza un 96,45 % de precisión en la clasificación de obstáculos y 23,76 imágenes por segundo para el tiempo de procesamiento, que es 6,75 veces más rápido que la técnica de estimación del plano de tierra tradicional. De esta forma, concluyendo que el método de segmentación rápida de obstáculos es más rápido que la técnica tradicional de estimación de plano de tierra en 3D. Además, que la técnica de clasificación de obstáculos es rápida y precisa para la detección de obstáculos.

Asimismo, en la investigación realizada por Rajab Awad et al. (2022) se plantea el desarrollo de un nuevo sistema de clasificación de billetes en tiempo real basado en el aprendizaje profundo. La metodología utilizada consistió en desarrollar un modelo capaz de identificar la denominación de los papeles de dinar iraquí, para posteriormente convertirlos en comandos vocales específicos. Dentro de los hallazgos, se demuestra que el modelo CNN logra una precisión de predicción del 98.5 %, haciendo uso de 1975 imágenes para probar datos. Además, el modelo CNN ofrece una precisión de 97 % en cuanto reconocimiento para los datos de prueba. Por último, el modelo propuesto logra un 98.6 % de precisión, al combinar ambos lados, es decir los datos árabes con el de inglés del billete. De esta manera, concluyendo que el modelo desarrollado puede emplearse en algunas aplicaciones de reconocimientos de imágenes de billetes en tiempo real, lo cual ayuda de manera significativa a guiar a las personas ciegas brindándoles la categoría de la moneda mediante comandos de voz, en actividades de compras o cambios de dinero.

Por ultimo, Jarraya et al. (2020) desarrollaron un método de Clasificación de obstáculos profundo multicapa basado en perceptrones a partir de información visual parcial. El objetivo de la investigación es proponer un método consistente, fiable y robusto basado en smartphones para clasificar obstáculos en entornos desconocidos a partir de información visual parcial basada en visión por computador y basadas en técnicas de visión artificial y aprendizaje automático. La metodología utilizados consistió primeramente en la extracción de características de imágenes como el color RGB, histogramas de gradiente orientados (HOG), HSV (tono, saturación y valor), su fotograma y características de patrones binarios locales uniformes (LBP). Para luego, realizar la clasificación de obstáculos, el cual consta de dos pasos: (i) la preparación del conjunto de datos de aprendizaje y prueba; (ii) la generación de clasificadores. Los resultados experimentales indican que el método propuesto permite la categorización semántica para clasificar obstáculos y aumentar el conocimiento del entorno explorado. Así, llegando a registrar una precisión del 90.2 % en comparación de otros métodos para la clasificación de obstáculos. De esta forma, concluyendo que el método propuesto maneja altos niveles de ruido y mala resolución en fotogramas capturados con la cámara de un teléfono.

Las investigaciones propuestas por Morparia et al. (2022) y Valipoor y de Antonio (2022) destacan la importancia de abordar desafíos técnicos como la usabilidad, practicidad, portabilidad y operabilidad en tiempo real en el desarrollo de tecnologías de asistencia para personas con discapacidad visual. Mientras que Yuan et al. (2022) sugiere que la combinación de la computación perimetral móvil de baja latencia y los sistemas de banda tradicional mejora la precisión y el rango de detección en aplicaciones de asistencia para personas ciegas. Estos desafíos y avances técnicos son relevantes para lograr un despliegue práctico y una mayor aceptación de los sistemas de reconocimiento y detección. Por otro lado, Mouna Afif y de Antonio (2021) y Rajab Awad et al. (2022) demuestran la viabilidad del uso de técnicas de visión artificial, como las redes neuronales convolucionales, en la identificación de objetos en entornos interiores y en la clasificación de billetes en tiempo real, respectivamente. Estos enfoques técnicos ofrecen soluciones prometedoras para ayudar a las personas con discapacidad visual en actividades diarias y transacciones financieras. En cuanto a la detección de obstáculos y la navegación para personas invidentes, tanto Vorapatratorn et al. (2021) como Jarraya et al. (2020) resaltan los beneficios del aprendizaje profundo y las técnicas de visión por computadora. Estos enfoques técnicos permiten una mayor precisión en la detección de obstáculos y una velocidad de procesamiento mejorada en comparación con las técnicas tradicionales. En resumen, los avances técnicos en el reconocimiento y detección de objetos, la clasificación de billetes y la detección de obstáculos ofrecen soluciones prometedoras para ayudar a las personas con discapacidad visual. Sin embargo, es crucial abordar desafíos adicionales como la usabilidad, la portabilidad y la adaptabilidad de estas tecnologías para garantizar su eficacia y aceptación generalizada en diferentes contextos y escenarios.

En Zeinullin y Hersh (2022) se desarrolla un sistema especial para la accesibilidad de materiales educativos para personas que presentan discapacidad visual. El objetivo de trabajo es probar la practicidad de las personas con discapacidad visual que utilizan la aplicación móvil desarrollada en entornos educativos. El sistema propuesto consta de 3 componentes: los gráficos táctiles pre-etiquetados, una herramienta web de etiquetado interactivo y la aplicación telefónica. La metodología utilizada consistió primero se en etiquetar la versión digital de los gráficos utilizando la herramienta web desarrollada. Luego, la aplicación del teléfono que se encuentra desarrollada para plataforma Android, acompañara a dichos gráficos con el audio descripciones. Los resultados obtenidos mostraron que el uso de la aplicación permitió un intercambio de información más rápido con el GT en comparación con el texto Braille. Además, que, el porcentaje de respuestas correctas fue mayor con el aplicativo mas que un lector de pantalla. De esta manera, concluyendo que el aplicativo móvil desarrollado permite a las persona con discapacidad visual poder acceder a gráficos táctiles proporcionando audio descripciones en tiempo real y de una forma más práctica, requiriendo solo un teléfono inteligente.

Por otro lado, Das et al. (2021) proponen un marco para la localización de imágenes basado en aprendizaje profundo. El objetivo del trabajo fue desarrollar un sistema auxiliar de orientación de bajo costo, fácil de usar y confiable para personas con discapacidad visual en espacios interiores y exteriores. La metodología utilizada consistió en hacer uso del marco de localización de imágenes basado en aprendizaje profundo llamado PathLookup, el cual ayudara a proporcionar información precisa sobre el avance del camino para la orientación exterior para ayudar a las personas

con discapacidad visual con la navegación en espacios exteriores. Los resultados de la evaluación demuestran que PathLookup es rápido y altamente preciso, con un 99.5 % sobre el conjunto de datos de validación. En conclusión, se destaca como una herramienta valiosa para la integración futura en los sistemas de orientación al aire libre. Además, se sugiere el aprendizaje profundo y la visión por computadora puede tener un gran potencial para la mejora de la calidad de vida de las personas con discapacidad visual.

Por otra parte, Merchan et al. (2021) realizan una revisión de la literatura sobre los sistemas de ayuda a la navegación en el contexto de las Islas Vírgenes Británicas (BVI por sus siglas en inglés). La metodología utilizada consistió en presentar la experiencia de primera mano y los avances sobre el uso de la cámara comercial RGB-D de Microsoft Kinect para el desarrollo de un sistema de ayuda para la navegación interior para personas que presentan discapacidad visual. Luego, se discuten las limitaciones que pueden encontrarse al construir tal sistema y se presenta una visión general de las nuevas vías de investigación en navegación interior para de las Islas Vírgenes Británicas. Los resultados demuestran que efectivamente existen nuevas vías de investigación que pueden mejorar la eficacia y la precisión de estos sistemas orientado en la ayuda a la navegación de personas con discapacidad visual, como la integración de algoritmos de visión por computadora, aprendizaje profundo para la clasificación de objetos y desarrollos recientes con visión profunda estéreo. En conclusión, los sistemas de asistencia para la navegación interior basados en visión, como la cámara comercial RGB-D de Microsoft Kinect, ofrecen posibilidades para mejorar la calidad de vida de las personas BVI.

Por otra parte, El-taher et al. (2021) presentan una revisión exhaustiva de la literatura directamente o relevante a la navegación asistida para personas con discapacidad visual. La metodología utilizada consistió en desglosar el área de navegación en una serie de fases y tareas de navegación. Luego, hacen uso la estructura para el desarrollo de la revisión sistemática de la investigación, para ello se analiza os artículos, métodos, conjunto de datos y limitaciones actuales por tarea. Adicionalmente, proporcionan una descripción general de las aplicaciones de navegación comerciales y no comerciales para personas con discapacidad visual. Los resultados de la investigación indican que los sistemas electromagnéticos basados en radar superan a los sistemas basados en sensores, los cuales se utilizan principalmente en tareas de evasión de obstáculos. Además, los sistemas basados en cámaras se ven afectados por las condiciones climáticas y de iluminación, pero brindan más detalles sobre los obstáculos, como la forma y el color. Por último, los sistemas basados en teléfonos inteligentes poseen una ventaja, ya que es un dispositivo que contiene diferentes componentes útiles que son necesarios para las tareas de navegación. De esta forma, se concluye que aún se necesita más trabajo en este campo para presentar un dispositivo de navegación confiable y completo para personas que presentan discapacidad visual, enfatizando la necesidad de transferir el aprendizaje entre otros dominios a este mismo, como otros dominios refiriéndose al de los automóviles automatizados, asistencia al conductor y navegación robótica.

Las investigaciones presentadas abordan diferentes aspectos de la accesibilidad y la navegación asistida para personas con discapacidad visual. En Zeinullin y Hersh (2022), se centra en el desarrollo de una aplicación móvil para mejorar la accesibilidad de materiales educativos, mostrando resultados positivos en términos de intercambio de información y respuestas correctas. De esta forma, se sugiere el uso

de dispositivos al alcance del usuario como un teléfono inteligente para poder implementar la solución de manera más fácil, así mismo para que pueda ser usada por los usuarios con discapacidad visual sin mucha complejidad, es decir su accesibilidad sea más práctica. Por su parte, en Merchan et al. (2021) se enfoca en la navegación interior utilizando la cámara comercial RGB-D de Microsoft Kinect, destacando las posibilidades de mejora en la eficacia y precisión de los sistemas. En esta investigación, se sugiere la integración de algoritmos de visión por computadora, aprendizaje profundo para la clasificación de objetos y desarrollo recientes con visión profunda estéreo, ya que permitieron mejorar la eficacia y la precisión de estos sistemas orientado en la ayuda a la navegación de personas con discapacidad visual. Por último, El-taher et al. (2021) realiza una revisión exhaustiva de la navegación asistida, concluyendo que se requiere más trabajo para desarrollar un dispositivo de navegación confiable y completo, y enfatizando la importancia de transferir el aprendizaje de otros dominios, como los automóviles automatizados y la navegación robótica. Sin embargo, afirma que aún es necesario mayores esfuerzos para desarrollar un dispositivo de navegación confiable y completo para personas con discapacidad visual. De esta manera, la investigación sugiere la transferencia de aprendizaje desde otros dominios, como el de los automóviles automatizados y la navegación robótica. En conjunto, estos artículos proporcionan contribuciones significativas en el campo de la accesibilidad y la navegación asistida, ofreciendo perspectivas y oportunidades para mejorar la calidad de vida de las personas con discapacidad visual.

En esa misma línea, Croce et al. (2019) presentan un sistema que permite a las personas con discapacidad visual navegar de forma autónoma en un entorno interior y exterior desconocido. La metodología empleado consiste primero en utilizar sensores inerciales y la cámara integrada en un teléfono inteligente como sensores para la navegación. Luego los autores plantean puntos de referencias especiales para ayudar a los usuarios en la localización de rutas predefinidas. Después, se realizaron pruebas experimentales en ambientes interiores controlados y en instalaciones exteriores reales para demostrar la eficacia del enfoque propuesto, además se desarrolló una comparación con métodos de aprendizaje profundo. Los resultados de la investigación muestran que el sistema es efectivo en ambientes interiores como exteriores, ya que puede proporcionar estimaciones precisas para la orientación de los usuarios. Como conclusión se llega que el sistema de navegación para personas con discapacidad visual que es efectivo y puede ser generalizado a otros usuarios. Además, el enfoque propuesto donde se hace uso de sensores inerciales y la cámara de un teléfono inteligente, es mucho más efectivo que los métodos de aprendizaje profundo.

Por último, Portilla (2022) proponen la creación de un perro robot dotado con inteligencia artificial que pueda guiar a las personas dependientes o con discapacidad. La metodología de la investigación consistió en la integración de GPS y herramientas de Google maps para la navegación en exteriores, junto a la implementación de algoritmo de visión artificial para la identificación de objetos y señales de tráfico. Los resultados de la investigación indican que la creación de un robot con forma de perro mediante uso de inteligencia artificial como algoritmos de visión artificial lo dotan de capacidad para distinguir de forma inequívoca entre un objeto y una persona, de comunicar información en tiempo real mediante voz; navegar en exteriores y distinguir objetos y señales de tráfico. De esta forma, llegando a la conclusión que el perro robot Tefi puede guiar a las personas hasta diferentes lugares, como tiendas,

restaurantes y hospitales, y puede comunicar información en tiempo real, como la situación del tráfico, a través de su conexión a Google. Los investigadores destacan que el robot también podría ser útil para asistir a personas mayores con demencia o enfermedades como el alzhéimer.

Las investigaciones realizadas por Das et al. (2021), Croce et al. (2019) y Portilla (2022) presentan diferentes enfoques para mejorar la navegación y orientación de las personas con discapacidad visual en distintos entornos, tanto interiores como exteriores, mediante el uso de tecnologías como la visión por computadora, el aprendizaje profundo y la inteligencia artificial. En el primer estudio Das et al. (2021), se destaca la necesidad de desarrollar herramientas auxiliares de orientación confiables y de bajo costo para complementar e integrar los sistemas basados en satélites existentes. Además, se afirma que la integración del aprendizaje profundo y la visión artificial permitió el desarrollo de un sistema capaz de identificar y localizar imágenes en tiempo real con una precisión del 99.5 %, según las pruebas realizadas sobre un conjunto de validación. Por otro lado, en el estudio Croce et al. (2019), se mencionan algunos desafíos técnicos durante la implementación de un sistema de navegación efectivo para personas con discapacidad visual. Se destaca que el sistema propuesto ARIANNA, que fue desarrollado mediante algoritmos de visión por computadora, logra un margen de error de 0.54 cm, lo que demuestra una mayor precisión en comparación con otras propuestas basadas en redes neuronales convolucionales o sistemas que integran GPS, como es el caso del sistema basado en PDR (Navegación inercial). Sin embargo, se muestra que el sistema ARIANNA puede verse afectado por obstáculos y cambios en la geometría del entorno, lo que podría reducir su precisión. Por último, en el estudio Portilla (2022), se ha desarrollado un robot con forma de perro dotado de inteligencia artificial para guiar a personas con discapacidad visual. Así mismo, se menciona que el uso de algoritmos de visión artificial y la conexión a servicios en línea, como Google Maps, permitirían al robot poder distinguir objetos y personas con alta precisión, además de proporcionar información en tiempo real. Sin embargo, se plantean cuestiones éticas y de privacidad relacionadas con la recopilación y uso de datos sensibles, puesto que la propuesta requiere conexión a internet y el uso de una cámara para funcionar adecuadamente. En general, estos tres estudios presentan enfoques interesantes y prometedores para mejorar la navegación y orientación de las personas con discapacidad visual. No obstante, se sugiere proporcionar mayor transparencia y detalles en cuanto a los escenarios de prueba, las limitaciones y las comparaciones con otros enfoques existentes, a fin de realizar una evaluación más completa de su eficacia y viabilidad en diferentes entornos y situaciones reales, teniendo en cuenta aspectos éticos y de privacidad en el uso de datos.

A continuación, exploraremos un conjunto de investigaciones centradas en el desarrollo de sistemas de reconocimiento de billetes diseñados para ayudar a personas con discapacidad visual en la identificación y gestión de diferentes denominaciones de moneda. Estos estudios hacen uso de diversas tecnologías avanzadas, incluyendo el aprendizaje automático, la visión por computadora y el procesamiento de imágenes, para abordar los desafíos específicos asociados con esta tarea crucial en la vida cotidiana.

En la investigación realizada por Sufri et al. (2019), se enfoca en el desarrollo de un sistema de reconocimiento de billetes basado en visión con el propósito de ayudar a las personas con discapacidad visual a identificar distintos tipos de billetes de Malasia (RM 1, RM 5, RM 10, RM 20, RM 50 y RM 100). El estudio se propone analizar en detalle el efecto de la región y la orientación en el rendimiento de enfoques de Aprendizaje Automático (Machine Learning) y Aprendizaje Profundo (Deep Learning). Esto se logra mediante el uso de imágenes de billetes capturadas en diferentes regiones y orientaciones con una cámara de teléfono móvil en un entorno controlado. El estudio se divide en dos experimentos: uno con imágenes de billetes de diferentes regiones y otro con imágenes de billetes en diversas orientaciones. Se emplean algoritmos de clasificación de Aprendizaje Automático, como k-Nearest Neighbors (kNN), Decision Tree Classifier (DTC), Support Vector Machine (SVM) y Bayesian Classifier (BC), utilizando características extraídas de los valores RGB de las imágenes. Además, se utiliza la red neuronal convolucional AlexNet para las imágenes con diferentes orientaciones. Los resultados revelan que SVM y BC alcanzan una precisión del 100 %, mientras que kNN y DTC logran un 99.7 % de precisión. También se concluye que la orientación afecta al rendimiento del modelo AlexNet, que presenta un mejor desempeño cuando se entrena con datos de orientación similar. Este estudio subraya la importancia de tener en cuenta la orientación al diseñar sistemas de reconocimiento de billetes basados en visión.

En contraste, el estudio de Castelar et al. (2022) se enfoca en el desarrollo de algoritmos avanzados para la clasificación precisa de imágenes de billetes centrados. Su objetivo principal es determinar tanto la denominación como los lados visibles de los billetes, centrándose en el desafío de clasificar un nuevo billete de L200 emitido con motivo del bicentenario de Honduras. El estudio presenta dos métodos de clasificación de alto rendimiento. El primero aprovecha descriptores locales avanzados como ORB o SIFT para establecer correspondencias entre puntos clave en la imagen de entrada y plantillas previamente definidas. Esto genera vectores de características altamente informativos que alimentan algoritmos de clasificación como Máquinas de Soporte Vectorial (SVM) o Bosques Aleatorios. El segundo método representa un avance significativo al introducir una Red Neuronal Convolucional (CNN) denominada LempiraNet. Esta CNN utiliza transferencia de aprendizaje para abordar las limitaciones de datos. Ambos métodos pueden aprovechar el preprocesamiento de imágenes para localizar de manera eficiente el billete en la imagen de entrada, lo que facilita su clasificación precisa. Los resultados son notables, con ambos métodos alcanzando un impresionante 98 % o más en puntuación F1 cuando se utiliza el preprocesamiento de imágenes para la ubicación del billete. Se destaca que SIFT supera a ORB en términos de rendimiento. Además, en cuanto al tiempo de ejecución, LempiraNet muestra una eficiencia sobresaliente, siendo al menos 20 veces más rápido que el otro método. En resumen, este estudio presenta una solución sólida y eficaz para la clasificación de billetes, con un enfoque particular en la adaptación a la incorporación de un nuevo billete en circulación, demostrando su potencial para aplicaciones prácticas y el fortalecimiento de sistemas de reconocimiento de billetes en el futuro.

Por otro lado, en Huang y Gai (2020), se presenta un enfoque innovador para la clasificación de billetes basado en la transformada de ondículas cuaterniónicas (QWT) y una red neuronal convolucional profunda. La metodología comienza aplicando la QWT para describir tanto la fase como la magnitud de diversas imágenes de billetes, aprovechando su sensibilidad direccional inherente y su estructura multiescala. Posteriormente, se desarrolla una red neuronal convolucional profunda, la cual se entrena utilizando imágenes de billetes junto con los coeficientes de ondículas cuaterniónicas de fase y magnitud. Los pesos neuronales se asignan en función de las probabilidades de salida de la red y se actualizan mediante el algoritmo de retropropagación. Finalmente, las redes entrenadas, junto con la decisión de una suma ponderada y las redes de ondículas cuaterniónicas de fase y magnitud, se utilizan para la clasificación de imágenes de billetes. Los resultados experimentales demuestran un rendimiento superior en comparación con otros algoritmos de clasificación de billetes de última generación, y además, se cumple con los requisitos de tiempo real para sistemas de clasificación de billetes, lo que sugiere la viabilidad de esta metodología para aplicaciones prácticas y el fortalecimiento de sistemas de reconocimiento de billetes en el ámbito financiero.

En la investigación realizada por Awad et al. (2022), se enfoca en la creación de un sistema de clasificación de billetes iraquíes basado en Deep Learning y tecnología de visión por computadora, con el propósito de asistir a personas con discapacidad visual en la identificación de los billetes. El objetivo principal es desarrollar un modelo de clasificación multiclase que pueda distinguir entre las distintas denominaciones de billetes iraquíes y proporcionar comandos de voz equivalentes a la imagen categorizada de los billetes para informar a las personas con discapacidad visual sobre su valor. Para llevar a cabo esta tarea, se utilizó una base de datos que incluye 3,961 muestras de imágenes de las siete categorías de billetes de papel iraquíes. Se entrenó una red neuronal convolucional (CNN) de diecinueve capas utilizando esta base de datos para lograr la clasificación precisa de los billetes, y el sistema desarrollado alcanzó una precisión del 98.6%. Estos resultados demuestran la viabilidad de la propuesta y su capacidad para mejorar la independencia financiera de las personas con discapacidad visual. En resumen, este estudio demuestra cómo la aplicación de tecnologías de visión por computadora y algoritmos de Deep Learning puede tener un impacto significativo en la vida de las personas con discapacidad visual, al proporcionarles una herramienta efectiva para identificar y distinguir entre los billetes iraquíes.

En el estudio de Fan et al. (2022), se centra en el desarrollo de un sistema avanzado para el reconocimiento de billetes etíopes utilizando una red neuronal convolucional (CNN) y su posterior implementación en una plataforma integrada. El objetivo principal de esta investigación es diseñar y poner en práctica un sistema eficaz para la detección y clasificación de las denominaciones de los billetes etíopes, un componente esencial en diversas aplicaciones relacionadas con transacciones monetarias automáticas. Para alcanzar este propósito, se recopilaron imágenes de billetes etíopes y se llevaron a cabo evaluaciones exhaustivas de distintas arquitecturas de CNN y técnicas de optimización con el fin de obtener un alto rendimiento en la tarea de clasificación. La arquitectura que destacó por su desempeño fue MobileNetV2, implementada con optimización RMSProp, logrando una notable precisión del 96.4%. Además, se llevó a cabo la implementación de este modelo en una plataforma integrada mediante el uso de Raspberry Pi 3 B+ y se desarrolló una interfaz

de usuario basada en web para permitir la detección en tiempo real de billetes falsos. Estos resultados subrayan la viabilidad de emplear tecnologías de aprendizaje profundo y sistemas integrados para fortalecer la autenticación de billetes en el contexto etíope, planteando posibilidades de aplicación en otras regiones y en una variedad de escenarios financieros y comerciales. De manera concluyente, se puede inferir que la aplicación de CNN y técnicas de optimización en el reconocimiento de billetes representa una solución promisoría para abordar los desafíos vinculados a la autenticación de moneda en naciones con sistemas monetarios diversos. Además, el uso de sistemas integrados, como el Raspberry Pi, tiene el potencial de llevar esta tecnología a entornos prácticos, incluyendo cajeros automáticos y máquinas de cambio, lo que contribuiría significativamente a fortalecer la seguridad en las transacciones financieras.

En el estudio de Tamayo Zapata (2018), se enfoca en abordar el desafío del reconocimiento de billetes por parte de personas con discapacidad visual en Colombia. El objetivo principal consiste en desarrollar un sistema de clasificación de once denominaciones de billetes colombianos utilizando técnicas de procesamiento de imágenes y reconocimiento de patrones. Para lograr esto, se implementa un prototipo que incluye un sistema de iluminación frontal para eliminar sombras y realzar detalles de texturas en los billetes. La metodología se divide en dos etapas: detección y identificación de billetes. La detección se realiza mediante procesamiento de imágenes y binarización para eliminar el fondo. Luego, se extraen puntos de interés uniformes y descriptores en un espacio de color distinto. La identificación se basa en una base de datos de muestras de billetes utilizada para entrenar redes neuronales MLP, optimizando parámetros para lograr la mayor precisión posible. Los resultados demuestran una precisión inicial del 85 % mediante validación cruzada, mejorada al 95 % al aumentar el conjunto de datos y manipular las muestras. El sistema se implementa en una Raspberry Pi 3 para aplicaciones prácticas, y se realiza una prueba final exitosa con 240 imágenes capturadas en tiempo real, destacando la eficiencia del sistema. En conclusión, este estudio presenta un enfoque innovador para mejorar la autonomía y calidad de vida de personas con discapacidad visual en Colombia al desarrollar un sistema preciso de clasificación de billetes mediante tecnologías de procesamiento de imágenes y redes neuronales. Además, se sugiere que esta metodología podría aplicarse en otros contextos similares para abordar desafíos relacionados con la identificación de moneda en circulación.

Finalmente, en la investigación realizada por Khalil et al. (2022), se presenta una aplicación móvil respaldada por inteligencia artificial diseñada para el reconocimiento de billetes, dirigida a personas con discapacidad visual. La aplicación utiliza una red neuronal convolucional MobileNet V2 previamente entrenada, técnicas de aprendizaje por transferencia, transformación masiva y conversión de texto a voz para detectar y clasificar billetes de los Emiratos Árabes Unidos y proporcionar una señal auditiva. El objetivo principal es facilitar a las personas con discapacidad visual la identificación y diferenciación de billetes utilizando smartphones comunes. La metodología incluyó la recopilación de imágenes de alta definición de los billetes, el entrenamiento del modelo de inteligencia artificial y la implementación en una aplicación móvil. Los resultados muestran que el sistema puede ser una herramienta efectiva, con una precisión promedio del 70 % en pruebas y 88 % en validación cruzada quintuple. Esto sugiere que la tecnología móvil y la inteligencia artificial pueden desempeñar un papel importante en mejorar la independencia de las personas con

discapacidad visual en las transacciones financieras cotidianas.

En resumen, a partir de los estudios mencionados, se observa un creciente interés en sistemas de reconocimiento de billetes que utilizan diversas tecnologías para asistir a personas con discapacidad visual en sus transacciones financieras, promoviendo su independencia. Varios estudios resaltan la importancia de considerar factores como la región y la orientación en el diseño de sistemas de reconocimiento de billetes basados en visión. Además, se presentan enfoques innovadores que combinan tecnologías como la transformada de ondículas cuaterniónicas y redes neuronales convolucionales profundas. Estos sistemas demuestran un alto rendimiento y la capacidad de adaptarse a la incorporación de nuevos billetes en circulación. También se destaca la relevancia de la accesibilidad y la utilidad de la tecnología móvil en la vida diaria de las personas con discapacidad visual. En conjunto, estos estudios ofrecen soluciones versátiles y prometedoras que pueden mejorar la independencia financiera y la calidad de vida de las personas con discapacidad visual en todo el mundo.

A continuación, se presentan algunas investigaciones desde el campo nacional relacionadas con el problema de investigación escasez de herramientas ergonómicas de bajo costo para personas con discapacidad visual.

En la ciudad de Lima, se realizó una investigación por Vela (2019), el cual propone el diseño e implementación de un bastón ergonómico con GPS para el desplazamiento de personas ciegas que permitan mejorar la autonomía de las personas y su entorno en el centro de Unión Nacional de Ciegos del Perú. La metodología utilizada consistió en el desarrollo de una investigación aplicada, en la rama de ingeniería de sistemas de monitoreo, específicamente para el desarrollo de programas computacionales. Los resultados obtenidos indican que efectivamente el bastón ergonómico logra incrementar el desplazamiento de personas, siendo además rastreado por su familia. La conclusión que llegaron fue que para poder lograr una ergonomía en un instrumento como es el bastón con sensores ultrasonido, debemos considerar las características, sobre la forma, tamaño, peso del bastón a realizar.

Por otro lado, en Chiclayo, se realizó una investigación por Vilchez (2021) donde propone el desarrollo de un Bastón sensorial geolocalizador inteligente para apoyar en el desplazamiento de personas invidentes en la organización regional de ciegos del Perú. El objetivo de la investigación es superar uno de los distintos aspectos que limitan la normal actividad de las personas que padecen de discapacidad visual parcial o total. Para la investigación se usó la metodología RUP, la cual se basa en componentes e interfaces, haciendo uso del Lenguaje Unificado de Modelado (UML). Además, puesto que es una de las metodologías más populares para el análisis, implementación y documentación de los sistemas orientados a objetos. Además, de la metodología de sistemas embebidos, el cual permitió establecer la base científica para el desarrollo del bastón sensorial, especificando cada uno de las etapas y actividades de su ciclo de vida para el desarrollo del prototipo que se define en la tesis. Los resultados obtenidos indican que este bastón sensorial puede hacer la vida más dinámica y segura para las personas con discapacidad visual. Además, que la fusión de las dos metodologías RUP y sistemas embebidos favoreció en el diseño del prototipo de bastón sensorial geolocalizador inteligente. De esta forma, concluyendo que el bastón sensorial diseñado permitirá evitar obstáculos durante su desplazamiento, ubicar a la persona en tiempo real y comunicarse con sus familiares.

En cambio, en Arequipa, Lizárraga (2018) realizó una investigación orientada a mejorar la calidad del desplazamiento diario de las personas con discapacidad visual, mediante el uso de sensores ultrasónicos. El objetivo del trabajo radica en elaborar un Diseño de un Bastón Electrónico para personas invidentes con materiales que se puedan conseguir en el mercado que permita mejorar la movilidad de su desplazamiento diario. La metodología utilizada consistió en primero realizar un estudio descriptivo-explicativo en el departamento de Arequipa en el periodo de junio del 2018 a marzo del 2019. En segundo, se realiza el diseño experimental de los diferentes sensores que se va a utilizar y deben de ser probados y calibrados para la mejor movilidad de personas con alguna discapacidad visual. Los resultados de la investigación indican que el tener un Bastón Electrónico que tenga varias opciones de configuración para poder saber si tienen un objeto o algún obstáculo en su desplazamiento les parece muy bueno porque así podrán reducir los accidentes que suelen tener habitualmente en su desplazamiento diario. De esta forma, llegando a la conclusión que la propuesta del Bastón Electrónico para personas con discapacidad visual se tuvo en cuenta las necesidades diarias de este grupo vulnerable y se investigaron las diferentes clases de sensores disponibles en el mercado para determinar cuáles eran los más adecuados para el propósito del dispositivo. Además, se realizó una encuesta para conocer directamente las necesidades y problemáticas de las personas con discapacidad visual, lo que contribuyó a garantizar que el bastón fuera eficiente y pudiera dar solución a sus inconvenientes en su desplazamiento diario.

De acuerdo con las tres investigaciones se puede inferir que localmente los investigadores se enfocan en el diseño y la implementación de bastones electrónicos con diferentes características y metodologías para mejorar la movilidad de personas invidentes. Además, que el uso de bastones electrónicos con sensores ultrasónicos y GPS tiende ser una herramienta efectiva para aumentar la autonomía y seguridad de las personas invidentes en su desplazamiento cotidiano. Del mismo modo, se destaca la importancia de considerar las necesidades específicas de las personas con discapacidad visual y de realizar investigaciones que incluyan su participación, para la identificación de problemas. Finalmente, que la fusión de diferentes metodologías, como RUP y sistemas embebidos, puede contribuir en el diseño y desarrollo de prototipos efectivos y eficientes. En resumen, dichas investigaciones presentan posibilidades interesantes para la mejora de la calidad de vida de personas invidentes en el Perú y son útiles para futuras investigaciones y desarrollos tecnológicos en este campo.

## 2.3. Bases Teóricas de la Investigación

### 2.3.1. Familias de billetes de 10, 20, 50 y 100 nuevos soles

De acuerdo con el Banco Central de Reserva del Perú (BCRP, 2020), se pone en circulación nuevos billetes de S/ 10, S/ 20, S/ 50, S/ 100 y S/ 200 a partir del año 2019, que incorporan modernos diseños, nuevos motivos y renovados elementos de seguridad.

#### 2.3.1.1. Nueva familia de billetes

Los billetes de la nueva familia se caracterizan por contar con renovados elementos de seguridad de fácil reconocimiento por parte del público. Su temática está enfocada en personajes del siglo XX que han influido notablemente en el arte, la literatura, la historia y la ciencia. También se destaca la temática de la biodiversidad incluyendo motivos de nuestra flora y fauna. (BCRP, 2020)

La nueva familia de billetes cuenta con los siguientes motivos:

1. **S/ 10:** En el anverso del billete de S/ 10 se muestra la imagen de María Isabel Granda y Larco conocida mundialmente como Chabuca Granda. En el reverso se muestran a la vicuña, el Escudo de Armas y la Flor de Amancaes (véase Figura 2.1).

*Figura 2.1: Billete de S/ 10 Nueva Familia.*



*Fuente: Banco Central de Reserva del Perú.*

2. **S/ 20:** En el anverso del billete de S/ 20 se muestra la imagen del literato y etnólogo José María Arguedas Altamirano, figura fundamental de la literatura indigenista del Perú. En el reverso se muestran imágenes del cóndor y de la Flor de la Cantuta (véase Figura 2.2).
3. **S/ 50:** En el anverso del billete de S/ 50 figura el retrato de la historiadora e investigadora del mundo andino María Rostworowski Tovar. En el reverso se muestra el jaguar y la Puya Raimondi. También se aprecia el Escudo de Armas (véase Figura 2.3).
4. **S/ 100:** En el anverso del billete de S/ 100 figura el retrato del científico, pionero de la astronáutica y de la era espacial Pedro Paulet Mostajo. En el reverso se muestran imágenes del Colibrí Cola de Espátula y de la orquídea *Phragmipedium kovachii*, ambas especies endémicas del Perú. También se aprecia el Escudo de Armas (véase Figura 2.4).

**Figura 2.2:** Billete de S/ 20 Nueva Familia.



*Fuente:* Banco Central de Reserva del Perú.

**Figura 2.3:** Billete de S/ 50 Nueva Familia.



*Fuente:* Banco Central de Reserva del Perú.

**Figura 2.4:** Billete de S/ 100 Nueva Familia.



*Fuente:* Banco Central de Reserva del Perú.

5. S/ 200: No se encuentra en circulación.

### 2.3.1.2. Familia de billetes 2011

Esta familia de billetes fue lanzada el año 2011 y todavía continúa en circulación.

1. **S/ 10:** En el anverso del billete de S/. 10 se muestra a José Abelardo Quiñones Gonzales, héroe de la aviación militar del Perú y el vuelo acrobático que realizó el día de su graduación. En el reverso se aprecia la ciudadela de Machu Picchu en Cusco (véase Figura 2.5).

*Figura 2.5: Billete de S/ 10 Familia 2011.*



Anverso

Reverso

*Fuente: Banco Central de Reserva del Perú.*

2. **S/ 20:** En el anverso del billete de S/. 20 se muestra a Raúl Porras Barrechea, distinguido catedrático, diplomático y periodista y la fachada del Palacio de Torre Tagle, sede de la Cancillería del Perú. En el reverso figura una de las construcciones del Complejo Arqueológico de Chan Chan en La Libertad (2.6).

*Figura 2.6: Billete de S/ 20 Familia 2011.*



Anverso

Reverso

*Fuente: Banco Central de Reserva del Perú.*

3. **S/ 50:** En el anverso del billete de S/. 50 figuran las imágenes del escritor Abraham Valdelomar Pinto y de la fachada del Palais Concert, centro de reunión habitual de los intelectuales de la época. En el reverso se observa las ruinas del Templo Nuevo de Chavín de Huantar en Ancash (2.7).
4. **S/ 100:** En el anverso del billete de S/. 100 se observa la imagen del ilustre historiador Jorge Basadre Grohmann y una vista del Paseo Cívico de Tacna, su ciudad natal. En el reverso figura una de las construcciones del Gran Pajatén ubicada en San Martín (2.8).

*Figura 2.7: Billete de S/ 50 Familia 2011.*



*Fuente: Banco Central de Reserva del Perú.*

*Figura 2.8: Billete de S/ 100 Familia 2011.*



*Fuente: Banco Central de Reserva del Perú.*

5. **S/ 200:** En el anverso del billete de S/. 200 se presenta la imagen de Santa Rosa de Lima, Patrona de América, Indias y Filipinas, así como el pozo de su santuario. En el reverso figura la Ciudad Sagrada de Caral ubicada en Lima (2.9).

*Figura 2.9: Billete de S/ 200 Familia 2011.*



*Fuente: Banco Central de Reserva del Perú.*

Basado en la entrevista llevada a cabo con los miembros de la Asociación Unión de Ciegos de Arequipa, se identificó que las denominaciones de billetes más comúnmente utilizadas son S/ 10, S/ 20, S/ 50 y S/ 100. En contraste, los billetes de S/ 200 no son de uso frecuente. Siguiendo esta información, en el enfoque de esta investigación se concentró en las denominaciones mencionadas. Esto asegura que el sistema esté

optimizado para el reconocimiento de los billetes más relevantes y empleados por la comunidad de personas con discapacidad visual en la ciudad de Arequipa.

### 2.3.2. Discapacidad Visual

De acuerdo con Torres-Carazo et al. (2021), existe una amplia gama de términos, definiciones y clasificaciones relacionados con la discapacidad. En estos tiempos de globalización, es necesario que haya una idea clara de los conceptos implícitos y su clasificación para que se pueda utilizar un lenguaje global y común.

Por este motivo, la OMS aprobó la Clasificación Internacional del Funcionamiento, de la Discapacidad y de la Salud (CIF) o International Classification of Functioning (ICF) (World Health Assembly, (WHA, 2001) citado por Torres-Carazo et al., 2021); cuyo objetivo es aproximar, simplificar y unificar la terminología e identificar los problemas relacionados con la discapacidad. La CIF complementa a la Clasificación Internacional de Enfermedades y Problemas Relacionados con la Salud (CIE-11) o International Classification of Diseases-11 (ICD-11) según la OMS (2018) citado por Torres-Carazo et al. (2021), publicada anteriormente por la OMS.

En tal sentido, la CIF define las **deficiencias** como problemas en las funciones y estructuras corporales, como una desviación o pérdida significativa, las **limitaciones en la actividad** como dificultades que una persona puede tener para realizar actividades, y las **restricciones en la participación** como problemas que una persona puede experimentar para implicarse en situaciones vitales. El término **discapacidad** está relacionado con las deficiencias, las limitaciones en la actividad y las restricciones en la participación. Este último denota los aspectos negativos de la interacción entre la(s) condición(es) de salud de una persona y los factores contextuales de esa persona (Factores ambientales y personales). WHA (2001); Torres-Carazo et al. (2021)

En este contexto, la CIE-11 OMS (2018) categoriza el deterioro de la visión en dos grupos: (i) deterioro de la visión cercana, con una Agudeza visual cercana inferior a N6 o M.08 a 40 cm; y (ii) deterioro de la visión distante, dividido en 4 subgrupos, los cuales son: Leve- Agudeza visual inferior a 6/12 o igual o superior a 6/18, Moderado- Agudeza visual inferior a 6/18 o igual o superior a 6/60, Grave- Agudeza visual inferior a 6/60 o igual o superior a 3/60, y Ceguera- Agudeza visual inferior a 3/60.

La experiencia individual de la discapacidad visual varía dependiendo de muchos factores diferentes. En ello influye, por ejemplo, la disponibilidad de intervenciones de prevención y tratamiento, el acceso a la rehabilitación de la visión (incluidos los productos de apoyo como gafas o bastones blancos), y si la persona tiene problemas debido a la inaccesibilidad de los edificios, los medios de transporte y la información. (OMS, 2022)

Este estudio se centra en personas adultas con un rango de edad entre 20-50 años con deterioro de la visión Grave y Ceguera.

### 2.3.3. Tecnologías de asistencia

De acuerdo con Hersh y Johnson (2008) citado por Zeinullin y Hersh (2022), los sistemas de tecnologías de asistencia se definen como “equipos, dispositivos y sistemas que pueden utilizarse para superar las barreras sociales, de infraestructura y de otro tipo que experimentan las personas discapacitadas y que impiden su participación plena y en igualdad de condiciones en todos los aspectos de la sociedad”. Para Beheshti et al. (2023), las tecnologías de asistencia permiten a las personas ser más productivas, más independientes y alcanzar una mayor calidad de vida.

En este sentido, según Valipoor y de Antonio (2022) “el desarrollo de tecnologías de asistencia para personas con discapacidad visual/ciegas, ha ayudado a afrontar diversos retos en sus vidas al proporcionar servicios como la detección de obstáculos, la navegación en interiores/exteriores, la descripción de escenas, la lectura de textos, el reconocimiento facial, entre otros.”

### 2.3.4. Tecnologías vestibles

Según Pascolini y Mariotti (2012) citado por Godfrey et al. (2018), la tecnología vestible (Wearable Technology por sus siglas en inglés WT) engloba una plétora de dispositivos que se llevan directamente sobre la persona o adheridos a ella. Entre estos últimos se encuentran los teléfonos inteligentes, que han adquirido gran popularidad y funcionalidad.

Asimismo, Strategic Policy Forum on Digital Entrepreneurship (2016) citado por Godfrey et al. (2018), afirma que los dispositivos vestibles se pueden subdividir en dos categorías: (i) primarios, que operan de forma independiente y funcionan como conectores centrales para otros dispositivos y/o información (por ejemplo, un monitor de actividad física que se lleva en la muñeca o en un smartphone); y (ii) secundarios, que capturan acciones específicas o realizan una medición (por ejemplo, un monitor de frecuencia cardíaca que se lleva en el pecho) y se envían a un dispositivo vestible primario para su análisis.

De acuerdo con Piculo-Dos Santos et al. (2021), se han desarrollado dispositivos vestibles para mejorar la navegación de las personas ciegas o con deficiencias visuales. Con los avances tecnológicos, la aplicación de los dispositivos vestibles ha ido en aumento. Asimismo, “los dispositivos vestibles recogen información sobre el usuario o el entorno, la procesan (local o globalmente) y la devuelven al usuario en tiempo real mediante señales acústicas o táctiles” (Piculo-Dos Santos et al., 2021).

### 2.3.5. Tecnologías de asistencia e IoT

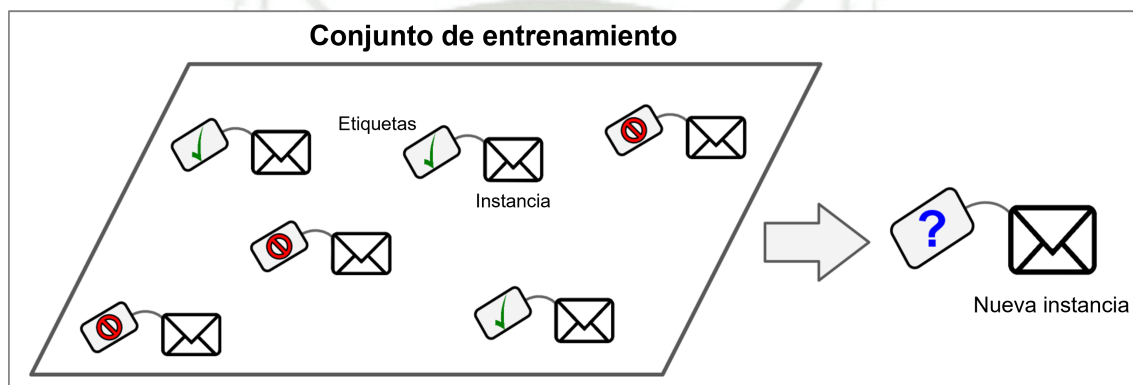
De acuerdo con Cocconcelli et al. (2023), el Internet de las Cosas (Internet of Things por sus siglas en inglés IoT) es ampliamente utilizado en aplicaciones orientadas a la salud, incluyendo soluciones muy específicas, que apoyan las necesidades de las personas con discapacidad, o técnicas para la monitorización continua y no intrusiva de parámetros fisiológicos relevantes. El uso de sensores vestibles está muy extendido, y se busca el enfoque no vestible para reducir aún más la carga del usuario. La adopción de técnicas de Inteligencia Artificial (IA) permite la evaluación indirecta de la salud mediante el reconocimiento de la actividad y el análisis del comportamiento. “IoT puede mejorar la calidad de vida de las personas mayores” (Mettouris et al., 2023).

El objetivo a largo plazo de nuestro equipo de investigación es mejorar la calidad de vida de las personas con discapacidad visual (deterioro de la visión Grave y Ceguera), proporcionándoles soluciones portátiles que fomenten una verdadera independencia funcional. (Beheshti et al., 2023)

### 2.3.6. Aprendizaje Supervisado

Según Géron (2019), en el aprendizaje supervisado, “los datos de entrenamiento que se introducen en el algoritmo incluyen las soluciones deseadas, denominadas *labels* o *etiquetas*” (Véase la Figura 2.10).

**Figura 2.10:** Un conjunto de entrenamiento etiquetado para aprendizaje supervisado (por ejemplo, clasificación de spam).



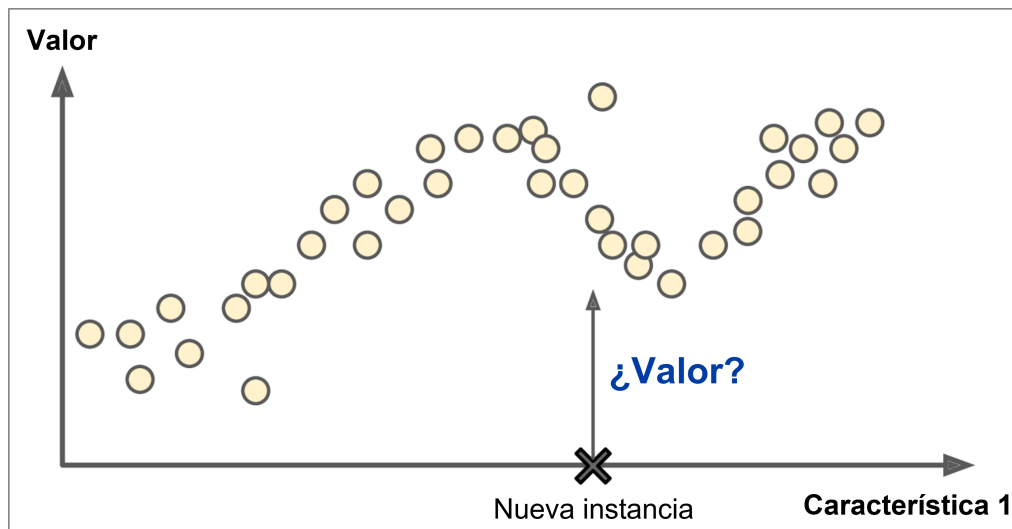
Fuente: Adaptado de Géron (2019).

Una tarea típica de aprendizaje supervisado es la *clasificación*. El filtro de spam es un buen ejemplo de ello: se entrena con muchos correos electrónicos de ejemplo junto con su clase (spam o ham), y debe aprender a clasificar los nuevos correos. Otra tarea típica es predecir un valor numérico objetivo, como el precio de un coche, a partir de un conjunto de *características* (kilometraje, marca, etc.) denominadas *predictores*. Este tipo de tarea se denomina *regresión* (Figura 2.11). Para entrenar el sistema, hay que darle muchos ejemplos de coches, incluyendo tanto sus predictores como sus etiquetas (es decir, sus precios).

Algunos algoritmos de regresión pueden utilizarse también para la clasificación, y viceversa. Por ejemplo, la *regresión logística* se utiliza habitualmente para la clasificación, ya que puede dar como resultado un valor que corresponde a la probabilidad de pertenecer a una clase determinada (por ejemplo, 20 % de probabilidades de ser spam). A continuación, se listan algunos de los algoritmos de aprendizaje supervisado más importantes (Géron, 2019):

- K-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees y Random Forests

*Figura 2.11: Ejemplo de regresión.*



*Fuente: Adaptado de Géron (2019).*

- Neural Networks

### 2.3.7. Redes Neuronales Convolucionales

Según Géron (2019) las redes neuronales convoluciones (Convolutional Neural Network, por sus siglas en inglés CNN), surgieron del estudio de la corteza visual del cerebro y se utilizan en el reconocimiento de imágenes desde la década de 1980. En los últimos años, gracias al aumento de la potencia computacional, la cantidad de datos de entrenamiento disponibles y los trucos existentes para entrenar las redes profundas, las CNN han logrado alcanzar un rendimiento sobrehumano en algunas áreas visuales complejas. Son la base de servicios de búsqueda de imágenes, coches autoconducidos, sistemas de clasificaciones automática de video y muchos más.

#### 2.3.7.1. Elementos básicos que componen una Red Neuronal

A continuación se puede ver, en la Figura 2.12, un esquema de una red neuronal:

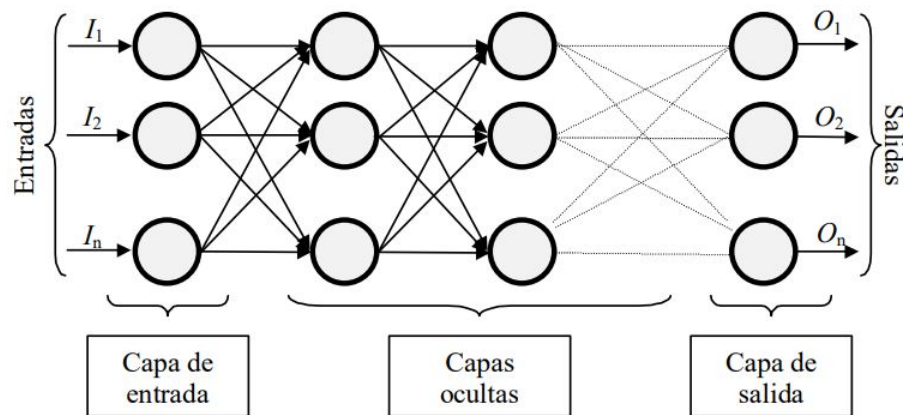
Está, esta constituida por neuronas interconectadas y arregladas en tres capas (est último puede variar). Los datos ingresan por medio de la "Capa de entrada", pasan a través de la "Capa oculta" salen por la "Capa de salida". Cabe mencionar que la capa oculta puede estar constituida por varias capas. (Matich, 2001)

De acuerdo con Matich (2001), antes de comenzar el estudio sobre las redes neuronales, se debe aprender algo sobre las neuronas y de cómo ellas son utilizadas por una red neuronal. En la figura 2.13 se compara una neurona biológica con una neurona artificial. En la misma se pueden observar similitudes entre ambas (tienen entradas, utilizan pesos y generan salidas).

Mientras una neurona es muy pequeña en sí misma, cuando se combinan cientos, miles o millones de ellas pueden resolver problemas complejos. Por ejemplo el cerebro humano se compone de billones de tales neuronas.

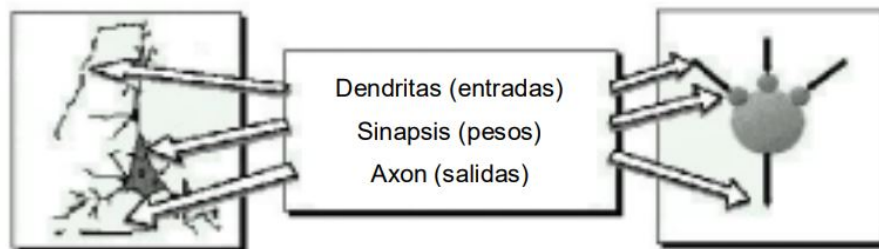
A continuación, se describe los elementos básicos que componen una Red Neuronal. (Garnica et al., 2018)

*Figura 2.12: Ejemplo de una red neuronal totalmente conectada.*



*Fuente: Extraído de Matich (2001).*

*Figura 2.13: Comparación entre neurona biológica y una neurona artificial.*



*Fuente: Extraído de Matich (2001).*

### 1. Función de entrada (Input Function)

La neurona trata muchos valores de entrada como si fueran uno solo, esto se conoce con el nombre de entrada global. Algunas de las funciones de entrada más comúnmente utilizadas son: (a) sumatoria de las entradas pesadas, (b) productos de las entradas pesadas y (c) máximo de las entradas pesadas.

### 2. Función de activación (Activation Function)

La función de activación calcula el estado de actividad de una neurona; transforma la entrada global en un valor (estado) de activación, cuyo rango normalmente va de 0 a 1 o de -1 a 1. Entonces, una neurona puede estar totalmente inactiva (0 o -1) o activa (1).

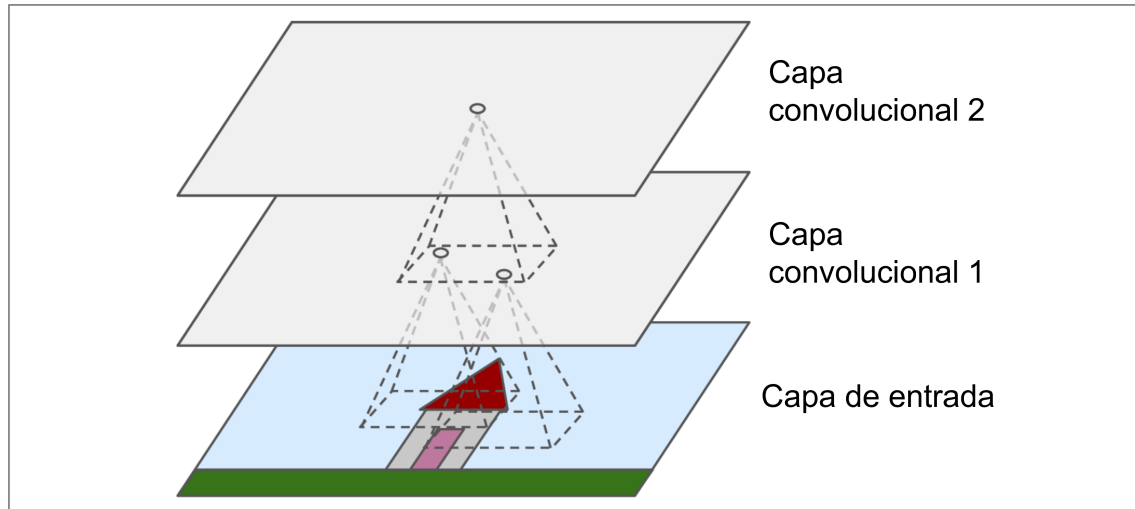
### 3. Función de salida (Output Function)

El último componente que necesita una red neuronal es la función de salida. El valor resultante de esta función es la salida de la neurona  $i$ ; por lo tanto, la función de salida determina qué valor se transfiere a las neuronas vinculadas. Si la función de activación está por debajo de un umbral determinado, ninguna salida se pasa a la neurona subsiguiente. Normalmente, no cualquier valor es permitido como una entrada para una neurona; por lo tanto, los valores de salida están comprendidos en el rango  $[0, 1]$  o  $[-1, 1]$ . (Matich, 2001)

### 2.3.7.2. Capas de convolución

De acuerdo con Géron (2019), el bloque de construcción más importante de una CNN es la capa convolucional: la neuronas de la primera capa convolucional no están conectadas a todos y cada uno de los píxeles de la imagen de entrada, sino sólo a los píxeles de sus campos receptivos (Figura 2.14).

**Figura 2.14:** Capas CNN con campos receptivos locales rectangulares.



Fuente: Adaptado de Géron (2019).

### 2.3.7.3. Capas de Pooling

Para Géron (2019), el objetivo de las capas de convolución es submuestrear (es decir, reducir) la imagen de entrada para reducir la carga computacional, el uso de memoria y el número de parámetros (limitando así el riesgo de sobreajuste u *overfitting*).

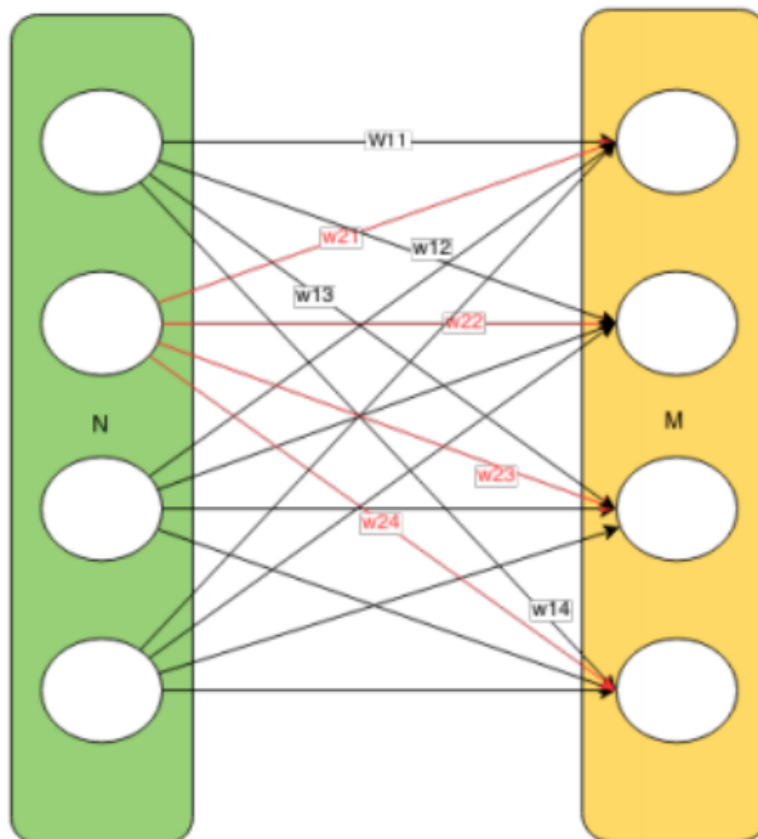
Al igual que en las capas convolucionales, cada neurona de una capa de agrupación está conectada a las salidas de un número limitado de neuronas de la capa anterior, situadas dentro de un pequeño campo receptivo rectangular. Debe definir su tamaño, la zancada y el tipo de relleno, igual que antes. Sin embargo, una neurona de agrupación no tiene pesos; todo lo que hace es agregar las entradas utilizando una función de agregación como la máxima o la media.

1. *Max-Pooling*: Géron (2019) afirma que la capa de agrupación máxima o Max-Pooling layer, es el tipo más común de capa de agrupación.
2. *Average-Pooling*: Asimismo, Géron (2019) afirma que las capas de agrupación de medias o Average-Pooling layer, solían ser muy populares, pero ahora se utiliza sobre todo capas Max-Pooling, ya que en general funcionan mejor.

#### 2.3.7.4. Capas fully-connected

Según Aguado (2020), tras haber pasado la imagen de entrada por las capas de convolución y reducción, normalmente se añaden 2 o más capas denominadas *Fully connected* o totalmente conectada. Deben este nombre a que todas las neuronas en una capa totalmente conectada se conectan a las neuronas de la capa anterior.

*Figura 2.15: Ejemplo de capa fully-connected.*



*Fuente: Extraído de Aguado (2020).*

#### 2.3.7.5. Capas de softmax y clasificación

Asimismo, Aguado (2020) afirma que si nos centramos en problemas de clasificación, tras la última capa Fully connected son necesarias una capa *softmax* y una capa de clasificación. La función de activación de la unidad de salida para esta capa es la función *softmax*, descrita previamente en el punto 2.3.6.1-2 de este apartado.

Por último, la capa de clasificación toma los valores de la función *softmax* y asigna cada entrada a una de las clases existentes mutuamente excluyentes usando la probabilidad de pertenencia a las clases.

### 2.3.8. Visión Artificial

De acuerdo con el libro de Szeliski (2010), la visión artificial se define como el campo de la informática que se ocupa de la adquisición, procesamiento, análisis y comprensión de imágenes y vídeos digitales. Conforme a Prince (2012), comprende la visión artificial como el proceso de automatizar la tarea de interpretación de imágenes y vídeos, utilizando técnicas y algoritmos de procesamiento de imágenes y aprendizaje automático. En cambio Forsyth y Ponce (2012), define el término como disciplina interdisciplinaria que se ocupa del análisis de imágenes y vídeos digitales, que tiene como objetivo automatizar la tarea de interpretar información visual, utilizando técnicas de procesamiento de imágenes y aprendizaje automático. En tal sentido, es preciso comprender a este término como aquel conjunto de actividades relacionadas entre sí, que tiene como principal objetivo la automatización de tareas de interpretación de imágenes y vídeos, que hace uso de algoritmos y técnicas de inteligencia artificial para el análisis de imágenes y vídeos digitales.

### 2.3.9. Algoritmos de Visión Artificial

De acuerdo con Szeliski (2010) y Gonzalez y Woods (2008), los algoritmos de visión artificial se basan en técnicas avanzadas de procesamiento de imágenes, aprendizaje automático y análisis estadístico para extraer características visuales, reconocer objetos, segmentar regiones y llevar a cabo diversas tareas relacionadas con la interpretación visual. Su aplicación abarca una amplia gama de áreas, incluyendo el reconocimiento facial, la detección de objetos y el seguimiento de movimiento. Según Gonzalez y Woods (2008), algunos de los algoritmos más usados en el campo se dividen en los siguientes tipos: algoritmos de detección de bordes, algoritmos de filtrado espacial, algoritmos de transformación de dominio frecuencial y algoritmos de reconocimiento de patrones. Dichos algoritmos permiten la extracción de características clave de imágenes, como bordes y texturas, además favorecen en la calidad de imágenes, identificación y clasificación de objetos; permiten poder analizar la distribución de frecuencias en el dominio de la imagen. Por otro lado, Szeliski (2010) menciona que los algoritmos de visión artificial pueden clasificarse en algoritmos de detección de características, algoritmos de reconocimiento de objetos, algoritmos de segmentación y algoritmos de seguimiento de movimiento. En primera instancia, los algoritmos de detección de características identifican y extraen características visuales clave, como esquinas y bordes, mientras que los algoritmos de reconocimiento de objetos permiten clasificar y reconocer objetos específicos en las imágenes. En cambio, los algoritmos de segmentación dividen una imagen en regiones o segmentos más pequeños, agrupando píxeles o regiones similares. Por último, los algoritmos de seguimiento de movimiento rastrean y siguen el movimiento de objetos o regiones en secuencias de imágenes o vídeos, en efecto siendo útil para analizar el comportamiento dinámico de los elementos visuales. En resumen, conforme a Szeliski (2010), Gonzalez y Woods (2008) y De Frías (2019), estas clasificaciones y técnicas de algoritmos de visión artificial desempeñan un papel primordial en una amplia variedad de aplicaciones, desde el reconocimiento facial y la detección de objetos hasta el seguimiento y análisis de movimiento en tiempo real, de esta forma aportando en el desarrollo de soluciones innovadoras y creativas.

### 2.3.10. Transferencia de aprendizaje y modelos de deep learning preentrenados

Se evaluaron modelos preentrenados mediante la técnica de transferencia de aprendizaje (transfer learning). La transferencia de aprendizaje, una técnica estrategia basada en el aprendizaje automático, se utiliza con frecuencia para abordar problemas de visión por computadora. (Meshram et al., 2023)

De acuerdo con Kulkarni et al. (2023) la transferencia de aprendizaje en la clasificación de billetes ofrece una valiosa herramienta para la creación de modelos precisos, incluso con conjuntos de datos de tamaño limitado. Al emplear modelos preentrenados, se incorpora información y representaciones generalizadas de diversas imágenes, lo que mejora la capacidad del modelo para identificar características distintivas y aplicarlas a billetes no observados. Esto agiliza el desarrollo y despliegue del modelo al reutilizar pesos previamente entrenados, evitando su desperdicio en el proceso de entrenamiento. Sin embargo, para superar desafíos potenciales, como el cambio de dominio entre modelos preentrenados y la tarea específica de clasificación de billetes, es necesario evaluar cuidadosamente aspectos como la arquitectura, los hiperparámetros y la selección de capas adecuadas, además de considerar enfoques de adaptación de dominio cuando sea pertinente.

En este contexto, se consideraron cinco modelos preentrenados para su evaluación, a saber, VGG16, VGG19, Xception, ResNet50 y MobileNetV2. Estos modelos fueron seleccionados consultando una lista de modelos preentrenados proporcionada por la biblioteca de Redes Neuronales. (Keras, 2023)

A continuación, se profundiza en la arquitectura y configuraciones de los modelos preentrenados seleccionados:

#### 1. VGG16

El modelo VGG, o VGGNet, que admite 16 capas también se conoce como VGG16 Ko Min-soo (2023), y es un modelo de red neuronal convolucional propuesto por Simonyan y Zisserman (2015) de la universidad de Oxford. Estos investigadores publicaron su modelo en el artículo de investigación titulado "Very Deep Convolutional Networks for Large-Scale Image Recognition".

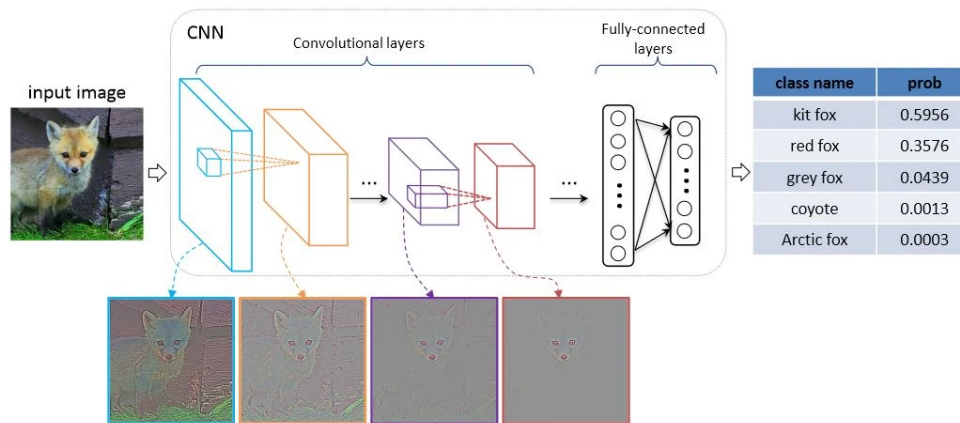
El modelo VGG16 logra casi un 92.7% de precisión de prueba entre los cinco primeros en ImageNet. ImageNet es un conjunto de datos que consta de más de 14 millones de imágenes pertenecientes a casi 1000 clases (Ko Min-soo, 2023). Además, fue uno de los modelos más populares presentados en *ILSVRC-2014*. Reemplaza los filtros de gran tamaño por varios filtros de tamaño  $3 \times 3$  uno tras otro, lo que supone una mejora significativa respecto a AlexNet. El modelo VGG16 se entrenó utilizando GPUs Nvidia Tital Black durante varias semanas.

Como se mencionó anteriormente, la VGGNet-16 consta de 16 capas y puede clasificar imágenes en 1000 categorías de objetos, que incluyen teclados, animales, lápices, ratones, etc. Además, el modelo tiene un tamaño de entrada de imagen de  $224 \times 224$ . (Ko Min-soo, 2023)

### Arquitectura VGG

Para comprender adecuadamente la arquitectura del modelo VGG16, es fundamental tener conocimiento de cómo se estructuran las VGGNets. Estas redes se fundamentan en los aspectos fundamentales de las redes neuronales convolucionales (CNN). El gráfico a continuación ilustra un ejemplo del funcionamiento de una CNN, que es relevante en este contexto.

**Figura 2.16:** Ejemplo del funcionamiento de una CNN.



*Fuente: Extraído de Ko Min-soo (2023).*

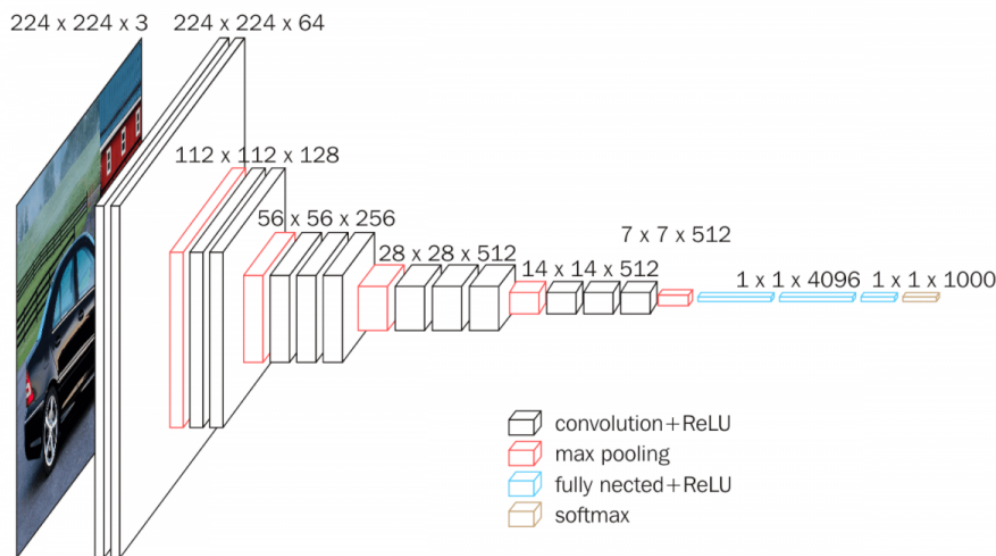
En la imagen 2.16, los datos de la imagen son la entrada de la CNN; la salida del modelo proporciona categorías de predicción para las imágenes de entrada.

En este sentido, la arquitectura VGG es la base de modelos revolucionarios de reconocimiento de objetos. Desarrollada como una red neuronal profunda, la VGGNet también supera los valores de referencia en muchas tareas y conjuntos de datos más allá de ImageNet. Además, todavía es una de las arquitecturas de reconocimiento de imágenes más populares. La imagen 2.17 muestra la arquitectura de una VGGNet.

La red VGG se construye con filtros convolucionales muy pequeños. El modelo VGG-16 consta de 13 capas convolucionales y tres capas completamente conectadas. A continuación, se resume la arquitectura VGG propuesta por (Simonyan y Zisserman, 2015):

- **Entrada:** VGG toma una entrada de imagen de tamaño  $224 \times 224$ .
- **Capas convolucionales:** Las capas convolucionales de VGG utilizan un campo receptivo mínimo, es decir,  $3 \times 3$ , el tamaño más pequeño posible que aún captura arriba/abajo e izquierda/derecha. Además, también hay filtros de convolución de  $1 \times 1$  que actúan como una transformación lineal de la entrada. Esto se sigue de una unidad ReLU, que es una gran innovación de AlexNet que reduce el tiempo de entrenamiento. ReLU significa función de activación de unidad lineal rectificadora (Ko Min-soo, 2023); es una función lineal por tramos que dará como salida la entrada si es positiva; de lo contrario, la salida es cero. La longitud de zancada (stride) de convolución está fijada en 1 píxel para mantener la resolución

*Figura 2.17: Arquitectura de una VGGNet.*



*Fuente: Extraído de Ko Min-soo (2023).*

espacial preservada después de la convolución (la zancada es la cantidad de desplazamientos de píxeles sobre la matriz de entrada).

- **Capas ocultas:** Todas las capas ocultas en la red VGG utilizan ReLU. VGG generalmente no utiliza la Normalización Local de Respuesta (LRN) ya que aumenta el consumo de memoria y el tiempo de entrenamiento. Además, no mejora la precisión general.
- **Capas completamente conectadas:** VGGNet tiene tres capas completamente conectadas. De las tres capas, las dos primeras tienen 4096 canales cada una, y la tercera tiene 1000 canales, uno por cada clase.

### **Arquitectura VGG16**

El número 16 en el nombre VGG se refiere al hecho de que es una red neuronal profunda de 16 capas (VGGnet). Esto significa que VGG16 es una red bastante extensa y tiene un total de alrededor de 138 millones de parámetros. Incluso según los estándares modernos, es una red enorme. Sin embargo, la simplicidad de la arquitectura de VGGNet16 es lo que hace que la red sea más atractiva. Solo con mirar su arquitectura, se puede decir que es bastante uniforme. (Ko Min-soo, 2023)

De acuerdo con la figura 2.18, a continuación se resume la arquitectura del modelo VGG16:

- La entrada al modelo es una imagen RGB de tamaño fijo de  $224 \times 224$  píxeles.
- El preprocesamiento implica restar la media de los valores RGB del conjunto de entrenamiento a cada píxel.
- 13 capas de convolución
  - Paso fijo de 1 píxel
  - Relleno (padding) de 1 píxel para filtros de  $3 \times 3$ .
- Capas de agrupación espacial
  - Esta capa no se incluye en la profundidad de la red por convención.
  - La agrupación espacial se realiza utilizando capas de max-pooling.
  - Tamaño de ventana de  $2 \times 2$  píxeles.
  - Paso fijo de 2 píxeles.
  - Se utilizan 5 capas de max-pooling en las redes convolucionales.
- Capas totalmente conectadas
  - 1ra: 4096 canales (ReLU)
  - 2da: 4096 canales (ReLU)
  - 3ra: 1000 canales (Softmax)

### *Configuración de la Arquitectura*

En la figura 2.19 se muestra la configuración de la red neuronal convolucional de la red VGG propuesta por Simonyan y Zisserman (2015), con las siguientes capas:

- VGG-11
- VGG-11 (LRN)
- VGG-13
- VGG-16 (Conv1)
- VGG-16
- VGG-19

Las configuraciones de la red neuronal convolucional se mencionan en la parte superior, una por columna.

Seguidamente, las redes se denominan por sus nombres (A - E). Todas las configuraciones siguen el diseño tradicional y se diferencian sólo en la profundidad: desde 11 capas de peso en la red A es decir 8 capas convolucionales y 3 capas completamente conectadas, a 19 capas de peso en la red E, es decir, 16 capas convolucionales y 3 capas completamente conectadas. El ancho de cada capa de convolución es el número de canales, que comienza desde 64 en la primera capa y luego aumenta por un factor de 2 después de cada capa max-pooling hasta que alcanza 512.

También se describe el número de parámetros para cada configuración. A pesar de tener una gran profundidad, el número de pesos en las redes no es mayor que el número de pesos en una red más superficial con anchos de capa de convolución y campo receptivo más grandes.

## 2. VGG19

El concepto del modelo VGG19 (también conocido como VGGNet-19) es el mismo que el de VGG16, con la diferencia de que admite 19 capas. Los números “16” y “19” representan la cantidad de capas de peso en el modelo (capas convolucionales). Esto significa que VGG19 tiene tres capas convolucionales adicionales en comparación con VGG16.

## 3. MobileNetV2

### *¿Qué es MobileNet?*

Como su nombre indica, el modelo MobileNet está diseñado para su uso en aplicaciones móviles y es el primer modelo de visión artificial móvil de TensorFlow. (Howard et al., 2017)

Asimismo, según Howard et al. (2017), MobileNet es una red neuronal convolucional simple pero eficiente y no muy intensiva en términos de cálculos, diseñada para aplicaciones de visión en dispositivos móviles. MobileNet se utiliza ampliamente en muchas aplicaciones del mundo real, que incluyen la detección de objetos, clasificaciones detalladas, atributos faciales y localización.

MobileNet utiliza convoluciones separables en profundidad. Esto reduce significativamente el número de parámetros en comparación con una red que utiliza convoluciones regulares con la misma profundidad. El resultado es una red neuronal profunda ligera. (Ko Min-soo, 2023)

Una convolución separable en profundidad se compone de dos operaciones:

- Convolución en profundidad.
- Convolución punto a punto.

### *Arquitectura de MobileNet*

Como se observa en la figura 2.20, la red completa de MobileNets consta de 30 capas. El diseño de la red es bastante sencillo:

- Capa convolucional con paso de 2
- Capa de profundidad
- Capa de punto que duplica el número de canales
- Capa de profundidad con paso de 2
- Capa de punto que duplica el número de canales
- Capa de profundidad
- Capa de punto
- Capa de profundidad con paso de 2
- Capa de punto que duplica el número de canales
- Y así sucesivamente...

Después de la primera capa (una convolución regular), las capas de profundidad y punto se alternan. A veces, la capa de profundidad tiene un paso de 2 para reducir el ancho y alto de los datos a medida que fluyen a través de la red. En ocasiones, la capa de punto duplica el número de canales en los datos. Todas las capas convolucionales están seguidas por una función de activación ReLU. (Howard et al., 2017)

Este proceso continúa hasta que la imagen original de  $224 \times 224$  píxeles se reduce a  $7 \times 7$  píxeles, pero ahora tiene 1024 canales. Después de esto, hay una capa de promediado que trabaja en toda la imagen, de modo que se termina con una imagen de  $1 \times 1 \times 1024$ , que es realmente solo un vector de 1024 elementos.

A pesar de que MobileNet ya está diseñado para ser bastante rápido, es posible utilizar una versión reducida de esta arquitectura. Hay tres hiperparámetros que se pueden configurar para determinar el tamaño de la red (Howard et al., 2017):

- **Multiplicador de ancho (el artículo llama a esto “alfa”):** Este hiperparámetro reduce el número de canales. Si el multiplicador de ancho es 1, la red comienza con 32 canales y termina con 1024.
- **Multiplicador de resolución (“rho” en el artículo):** Esto reduce las dimensiones de la imagen de entrada. El tamaño de entrada predeterminado es de  $224 \times 224$  píxeles.
- **Superficial o profundo:** La red completa tiene un grupo de 5 capas en el medio que se pueden omitir.

Estos ajustes se pueden utilizar para hacer que la red sea más pequeña, y por lo tanto más rápida, pero a costa de la precisión de las predicciones. (Howard et al., 2017)

### ***MobileNetV2***

MobileNetv2 sigue utilizando convolucionales separables en profundidad, pero su bloque de construcción principal ahora se estructura como en la figura 2.21. (Sandler et al., 2019)

MobileNet V2 introduce un bloque de construcción con tres capas convolucionales. A diferencia de su predecesor, V1, en V2 la capa de convolución  $1 \times 1$  reduce el número de canales, siendo ahora denominada capa de proyección. Por ejemplo, puede reducir un tensor de 144 canales a tan solo 24. Además, se agrega una capa de expansión con un factor de ampliación predeterminado de 6 que aumenta la cantidad de canales antes de la convolución en profundidad. (Sandler et al., 2019)

La arquitectura de MobileNet V2 incluye una conexión residual y mantiene la normalización por lotes y la función de activación ReLU6. A diferencia de V1, la capa de proyección no aplica una función de activación. El modelo consta de 17 bloques de construcción consecutivos, seguidos de una convolución  $1 \times 1$ , una capa de promediado global y una capa de clasificación. El primer bloque es ligeramente diferente, utilizando una convolución  $3 \times 3$  en lugar de la capa de expansión. Esta arquitectura se ha diseñado para reducir la cantidad de datos

que fluyen a través de la red y mantener el flujo de gradientes en la red para mejorar la eficiencia y el rendimiento.

#### 4. ResNet50

Antes de adentrarnos en la arquitectura de ResNet, es importante comprender algunos conceptos importantes relacionados con este tipo de arquitecturas.

##### *¿Qué es una Red Neuronal Profunda Residual (Deep Residual Network)?*

Una Red Neuronal Profunda Residual es casi similar a las redes que tienen capas de convolución, capas de pooling, activación y capas completamente conectadas apiladas una sobre la otra. La única construcción que convierte a una red simple en una red residual es la conexión de identidad entre las capas (Ko Min-soo, 2023). La figura 2.22 muestra el bloque residual utilizado en la red (He et al., 2015). Se puede apreciar la conexión de identidad como la flecha curva que se origina desde la entrada y se dirige hacia el final del bloque residual.

##### *¿Qué es una Función Residual?*

De acuerdo con Ko Min-soo (2023), la función residual (también conocida como mapeo residual) es la diferencia entre la entrada y la salida del bloque residual en cuestión. En otras palabras, el mapeo residual es el valor que se agregará a la entrada para Aproximar la función final ( $A_1, A_2, A_3, \dots, A_n$ ) del bloque. También se puede suponer que el mapeo residual es la cantidad de error que se puede agregar a la entrada para llegar al destino final, es decir, para aproxima la función final. Se puede visualizar el Mapeo Residual en la figura 2.23. Se puede apreciar que el Mapeo Residual actúa como un puente entre la entrada y la salida del bloque. Las capas de pesos y la función de activación no se muestran en el diagrama, pero se encuentran presentes en la red.

##### *Características Clave de ResNet*

ResNet utiliza Normalización por Lotes en su núcleo (He et al., 2015). La Normalización por Lotes ajusta la capa de entrada para aumentar el rendimiento de la red. Se mitiga el problema del cambio covariable. ResNet utiliza la Conexión de Identidad, que ayuda a proteger la red del problema del gradiente desvaneciente. La Red Neuronal Profunda Residual utiliza el diseño de bloque residual de botella para aumentar el rendimiento de la red. (Ko Min-soo, 2023)

##### *Arquitectura ResNet50*

Teniendo en cuenta el contexto anterior, como se observa en la figura 2.24, la Arquitectura de ResNet50 consta de 4 etapas. La red puede tomar una imagen de entrada con altura y ancho múltiplos de 32 y un ancho de canal de 3 (He et al., 2015). Para fines explicativos, en Ko Min-soo (2023), se considera un tamaño de entrada de  $224 \times 224 \times 3$ . Cada arquitectura de ResNet realiza la convolución inicial y el max-pooling utilizando tamaños de kernel de  $7 \times 7$  y  $3 \times 3$  respectivamente. Posteriormente, se comienza la Etapa 1 de la red y consta de 3 bloques residuales que contienen 3 capas cada uno. El tamaño de los núcleos utilizados para realizar la operación de convolución en las 3 capas del bloque de la Etapa 1 es de 64, 64 y 128 respectivamente.

Para redes más profundas como ResNet50, ResNet152, etc., se utiliza el diseño de botella. Para cada función residual  $F$ , se apilan 3 capas una sobre la otra.

Las tres capas son convoluciones de  $1 \times 1$ ,  $3 \times 3$ ,  $1 \times 1$ . Las capas de convolución  $1 \times 1$  son responsables de reducir y luego restaurar las dimensiones. La capa de  $3 \times 3$  se mantiene como un cuello de botella con dimensiones de entrada/salida más pequeñas.

Finalmente, la red tiene una capa de Average Pooling seguida de una capa completamente conectada con 1000 neuronas (salida de clases de ImageNet) He et al. (2015).

## 5. Xception

Xception es una arquitectura profunda de redes neuronales convolucionales que involucra Convoluciones Separables por Profundidad (Depthwise Separable Convolutions) (Chollet, 2017). Fue desarrollada por investigadores de Google.

### *¿Cómo funciona Xception?*

Xception es una arquitectura eficiente que se basa en dos puntos principales:

- Convolución Separable por Profundidad
- Atajos entre bloques de convolución, como en ResNet

De acuerdo con Ko Min-soo (2023), las Convoluciones Separables por Profundidad son alternativas a las convoluciones clásicas que se supone son mucho más eficientes en cuanto al tiempo de cálculo.

#### a) **Convolución Separable por Profundidad Original**

Como se observa en la figura 2.25, la convolución separable por profundidad original consiste en la convolución por profundidad seguida de una convolución punto a punto.

- La convolución por profundidad es una convolución espacial de  $n \times n$  a nivel de canales. Supongamos que en la figura 2.25 tenemos 5 canales, entonces tendremos 5 convoluciones espaciales de  $n \times n$ .
- La convolución punto a punto en realidad es una convolución de  $1 \times 1$  para cambiar las dimensiones.

En comparación con la convolución convencional, no se necesita realizar convoluciones a través de todos los canales. Esto significa que hay menos conexiones y que el modelo es más ligero. (Ko Min-soo, 2023)

#### b) **Convolución Separable por Profundidad Modificada en Xception**

De acuerdo con la figura 2.26, la Convolución Separable por Profundidad Modificada consiste en una convolución punto a punto seguida de una convolución por profundidad. Esta modificación se inspira en el módulo Inception de Inception-v3 Chollet (2017), donde primero se realiza una convolución de  $1 \times 1$  antes de cualquier convolución espacial  $n \times n$ . Por lo tanto, es un poco diferente de la original ( $n = 3$  aquí, ya que en Inception-v3 se utilizan convoluciones espaciales de  $3 \times 3$ ). Ko Min-soo (2023)

Dos diferencias menores:

- Orden de operaciones: Como se mencionó, las convoluciones separables por profundidad originales, tal como suelen implementarse (por ejemplo, en TensorFlow), realizan primero una convolución espacial a nivel de canales y luego una convolución de  $1 \times 1$ , mientras que en la Convolución Separable por Profundidad Modificada se realiza primero una convolución de  $1 \times 1$  y luego una convolución espacial a nivel de canales. Se afirma que esto no es importante porque cuando se utiliza en un entorno apilado, solo aparecen pequeñas diferencias al principio y al final de todos los módulos Inception encadenados.
- Presencia/Ausencia de no linealidad: En el Módulo Inception original, hay no linealidad después de la primera operación. En Xception, la Convolución Separable por Profundidad Modificada, NO hay no linealidad intermedia.

### ***Arquitectura Xception***

Como se observa en la figura 2.27, en una red Xception los datos primero pasan por el flujo de entrada, luego por el flujo intermedio, el cual se repite ocho veces, y finalmente por el flujo de salida. Es importante destacar que todas las capas de convolución (Conv) y convolución separable (SeparableConv) están seguidas de una normalización por lotes (no se incluye en el diagrama). Chollet (2017)

El autor ha dividido toda la arquitectura Xception en 14 módulos, donde cada módulo es simplemente un conjunto de capas DSC (Depthwise Separable Convolution) y capas de pooling. Los 14 módulos se agrupan en tres grupos; el flujo de entrada, el flujo intermedio y el flujo de salida. Cada uno de los grupos tiene cuatro, ocho y dos módulos respectivamente. El último grupo, es decir, el flujo de salida, opcionalmente puede tener capas completamente conectadas al final Ko Min-soo (2023).

*Nota: Todas las capas DSC en la arquitectura utilizan un tamaño de filtro de  $3 \times 3$ , un paso de 1 y un relleno "same" (igual). Y todas las capas de MaxPooling utilizan un núcleo de  $3 \times 3$  y un paso de 2.*

### **2.3.11. Estándares de ergonomía para dispositivos IoT vestibles**

De acuerdo con *Ergonomía y Discapacidad* (2020), los dispositivos de IoT vestibles son aquellos que se usan directamente en el cuerpo o integrados en la ropa y accesorios, abren nuevas fronteras en términos de interacción y comunicación entre los individuos y su entorno tecnológico. Conforme a Esparza Maldonado et al. (2018) a medida que la línea entre el mundo digital y el mundo físico se desdibuja, la ergonomía se convierte en un factor crítico para garantizar que estos dispositivos no solo sean funcionales, sino también seguros, cómodos y eficaces para su uso. En este sentido, de acuerdo con "Ergonomics — General approach, principles and concepts" (2011), las Normas ISO surgen como una valiosa herramienta o guía esencial, ofreciendo un conjunto de principios y directrices para diseñar, desarrollar y crear productos y sistemas que satisfagan los requisitos de seguridad, eficiencia y comodidad en su uso. Estas normas se alinean tomando en consideración las capacidad

y necesidades de cada persona, incluyendo a aquellos que podrían tener discapacidades. A continuación, se analizan detalladamente las normas ISO dedicadas a la ergonomía y dispositivos de IoT, ofreciendo una perspectiva integral sobre su alcance y aplicación.

1. ISO 26800:2011: Según “Ergonomics — General approach, principles and concepts” (2011), la norma ISO 26800:2011 aborda el ámbito de la ergonomía, ofreciendo un enfoque general y estableciendo principios y conceptos fundamentales relacionados con esta disciplina. El propósito de esta norma es mejorar diversos aspectos de los resultados de diseño a lo largo del ciclo de vida, incluyendo seguridad, rendimiento, eficacia, eficiencia, confiabilidad, disponibilidad y mantenibilidad. Además, se busca salvaguardar y mejorar la salud, el bienestar y la satisfacción de todas las personas involucradas o afectadas por los elementos diseñados. Esta norma se dirige a diseñadores, ergonomistas y otros profesionales, y sirve como referencia para desarrolladores de regulaciones relacionadas con ergonomía. Fue publicada en agosto de 2011, con un alcance en el ámbito de la ergonomía y la protección de la salud.
2. ISO 9241-210:2019: En “Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems” (2019) menciona que, la norma ISO 9241-210:2019 brinda directrices esenciales para aplicar el diseño centrado en el ser humano en sistemas interactivos basados en computadora, asegurando una interacción optimizada entre humanos y sistemas a lo largo del ciclo de vida del diseño. La norma está dirigida a responsables de la planificación y gestión de proyectos de diseño y desarrollo de sistemas interactivos, así como a profesionales de factores humanos y usabilidad involucrados en el diseño centrado en el ser humano.
3. ISO/IEC 30141:2018: Por otro lado, en “Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems” (2019) menciona que la norma ISO/IEC 30141:2018 busca establecer un lenguaje común, diseños reutilizables y buenas prácticas en el ámbito del Internet de las Cosas (IoT). Su propósito es brindar una estructura coherente para comprender y desarrollar sistemas en el campo del IoT. Dirigida a profesionales como diseñadores, desarrolladores y arquitectos de sistemas en este ámbito, esta norma proporciona una arquitectura de referencia estandarizada. Esto facilita la comunicación, el diseño y la implementación de soluciones de IoT, promoviendo la coherencia y eficiencia en el desarrollo tecnológico en constante evolución.

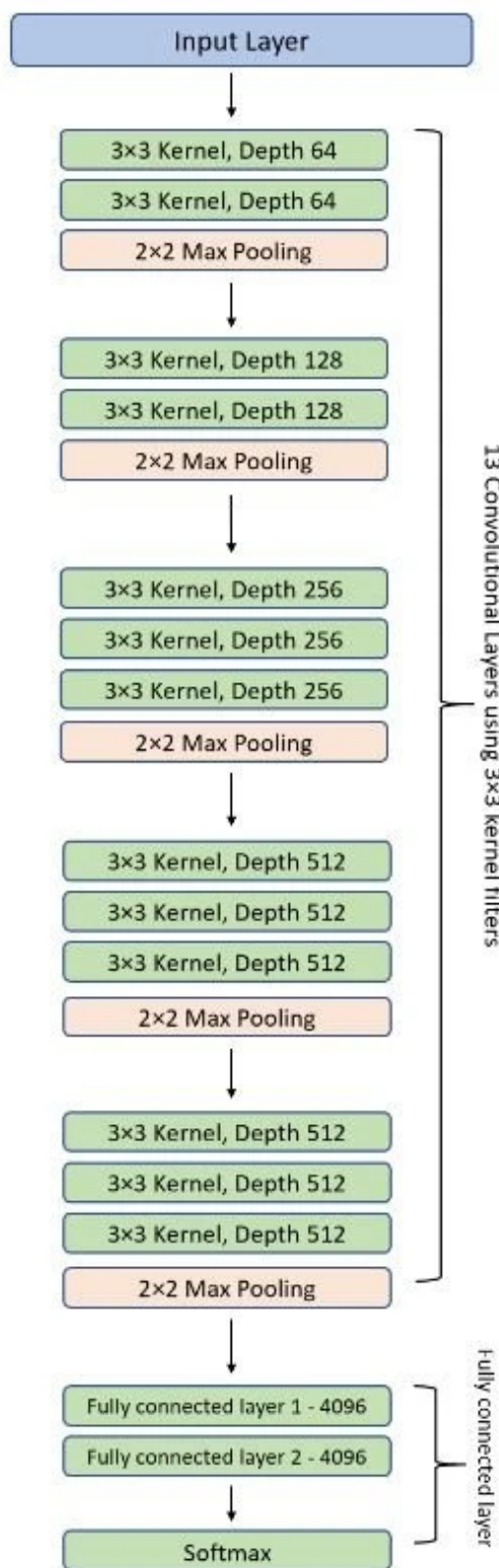
En síntesis, los dispositivos de IoT vestibles representan una innovación en la interacción entre individuos y tecnología al ser incorporados en el cuerpo o la vestimenta. La convergencia entre lo digital y lo físico demanda una ergonomía enfocada en la seguridad y comodidad de estos dispositivos, lo cual es abordado por las Normas ISO. La norma ISO 26800:2011 establece principios ergonómicos fundamentales, mientras que la norma ISO 9241-210:2019 promueve el diseño centrado en el ser humano en sistemas interactivos basados en computadora. Además, la norma ISO/IEC 30141:2018 brinda una estructura para sistemas IoT, asegurando coherencia y eficiencia en su desarrollo. Estas normas, al considerar capacidades y necesidades diversas, contribuyen a la creación de productos y sistemas accesibles y óptimamente diseñados.

### 2.3.12. Diseño inclusivo vs diseño universal

En el ámbito del diseño y la creación de productos, dos conceptos han emergido con fuerza en respuesta a la necesidad de abordar las diversas necesidades de las personas en la sociedad actual: el diseño universal y el diseño inclusivo. Conforme a Lemos (2015), Estas ideas marcan un cambio significativo en la manera en que concebimos y desarrollamos productos, centrándose en garantizar que sean accesibles para todos, sin importar las capacidades o características individuales de las personas. A medida que la conciencia sobre la diversidad y la inclusión se expande, la noción de accesibilidad cobra un valor primordial entre el diseño y la posibilidad de que todas las personas participen de forma activa respecto experiencia del producto. Conforme a Ontiveros et al. (2014), en la actualidad se han establecido estándares y normativas que proveen una serie de directrices esenciales a considerar en el proceso de diseño y fabricación de productos, especialmente aquellos destinados a personas que enfrentan diversas discapacidades, como la discapacidad visual. Las normas ISO desempeñan un papel fundamental al posibilitar la creación e implementación de una variedad de productos y servicios orientados a satisfacer las necesidades de accesibilidad de cada individuo, en particular de aquellos que presentan alguna forma de discapacidad. Estas normativas establecen directrices específicas para la elaboración de dispositivos diseñados para asistir a personas invidentes, además de promover la inclusión de personas de la tercera edad.

A modo de resumen, el diseño universal y el diseño inclusivo representan un cambio significativo en la concepción y desarrollo de productos, priorizando la accesibilidad para todas las personas. Las normas ISO desempeñan un papel fundamental al establecer directrices esenciales para la creación de productos inclusivos, promoviendo la participación y el bienestar de individuos con diversas capacidades y necesidades.

*Figura 2.18: Arquitectura VGG-16 de un modelo VGG16.*



*Fuente: Extraído de Ko Min-soo (2023).*

*Figura 2.19: Configuración de la CNN de VGGNet.*

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input ( $224 \times 224$ RGB image)					
conv3-64	conv3-64 <b>LRN</b>	conv3-64 <b>conv3-64</b>	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 <b>conv3-128</b>	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 <b>conv1-256</b>	conv3-256 conv3-256 <b>conv3-256</b>	conv3-256 conv3-256 conv3-256 <b>conv3-256</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

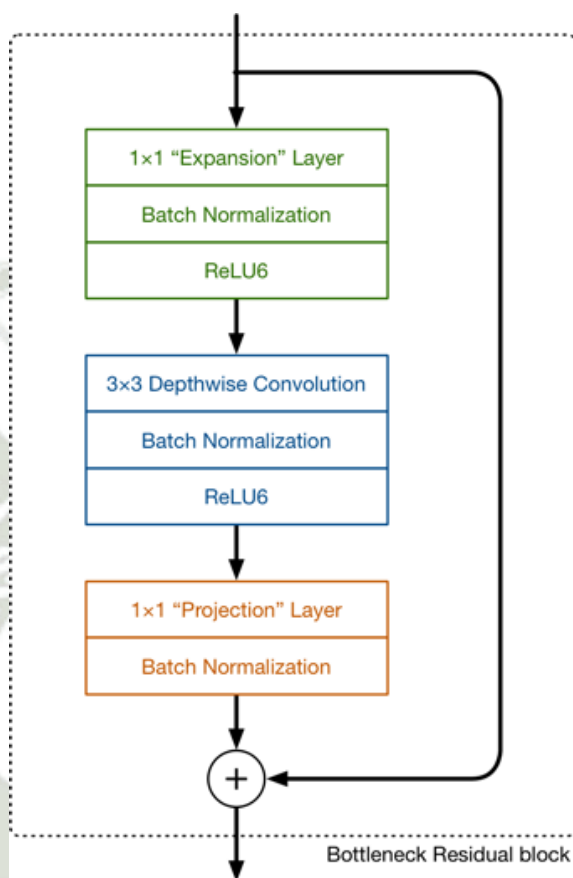
*Fuente: Extraído de Simonyan y Zisserman (2015).*

*Figura 2.20: Arquitectura MobileNet.*

Type / Stride	Filter Shape	Input Size	
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$	
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$	
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$	
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$	
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$	
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$	
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$	
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$	
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$	
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$	
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$	
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$	
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$	
5×	Conv dw / s1	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
	Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$	
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$	
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$	
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$	
Avg Pool / s1	Pool $7 \times 7$	$7 \times 7 \times 1024$	
FC / s1	$1024 \times 1000$	$1 \times 1 \times 1024$	
Softmax / s1	Classifier	$1 \times 1 \times 1000$	

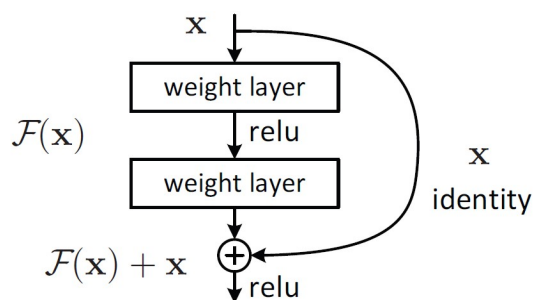
*Fuente: Extraído de Howard et al. (2017).*

**Figura 2.21:** Construcción de una arquitectura MobileNetV2.



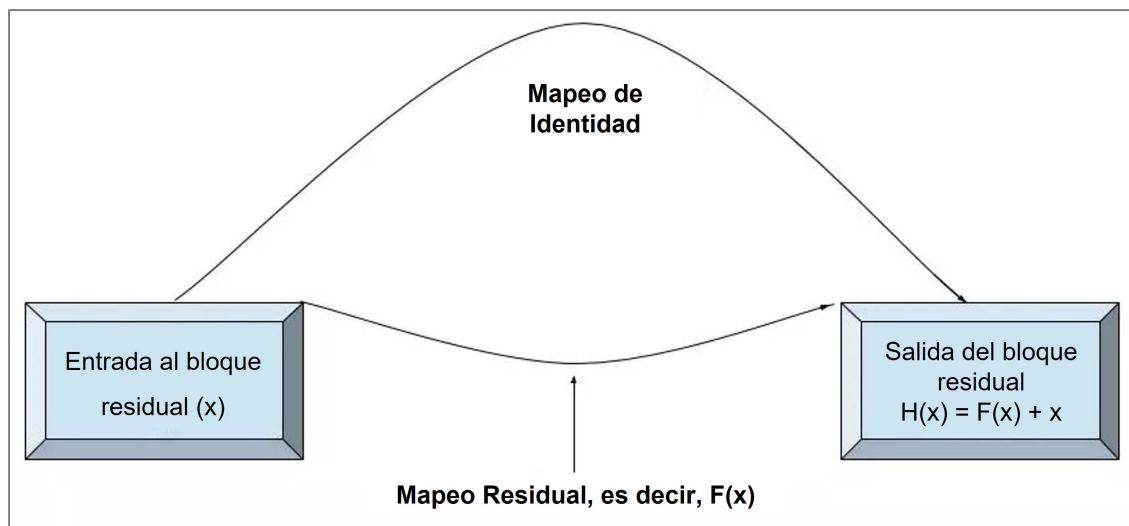
Fuente: Extraído de Ko Min-soo (2023).

**Figura 2.22:** Un Bloque Residual de una Red Neuronal Profunda Residual.



Fuente: Extraído de He et al. (2015).

**Figura 2.23:** Representación intuitiva de la Función Residual.



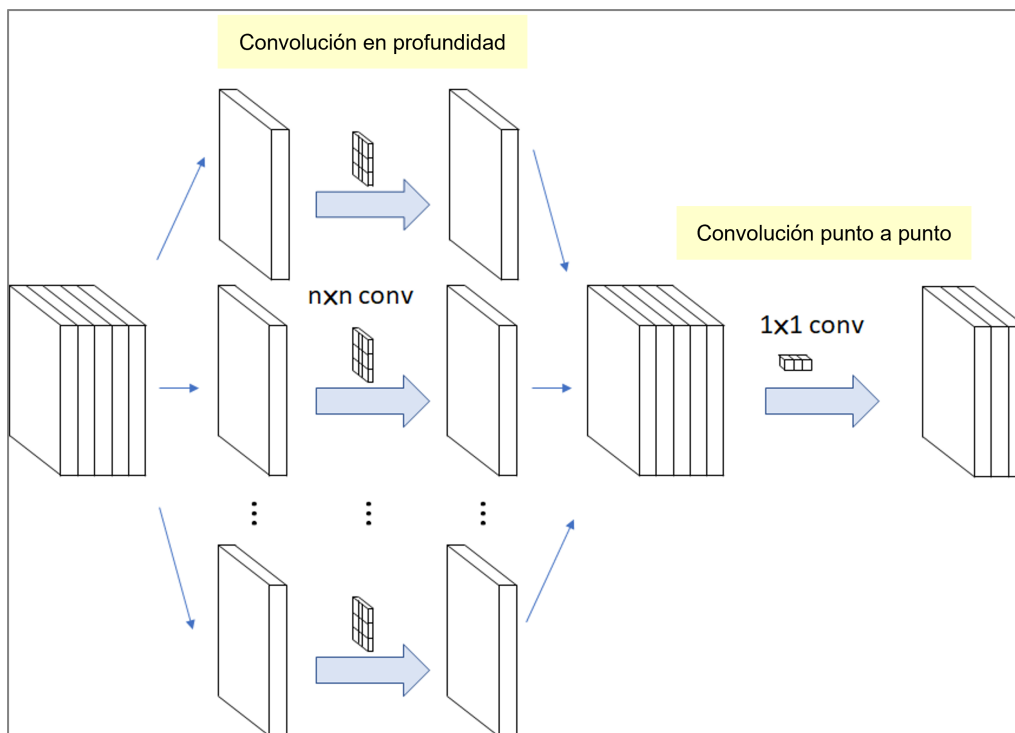
Fuente: Adaptado de Ko Min-soo (2023).

**Figura 2.24:** Arquitectura ResNet50.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		$1.8 \times 10^9$	$3.6 \times 10^9$	$3.8 \times 10^9$	$7.6 \times 10^9$	$11.3 \times 10^9$

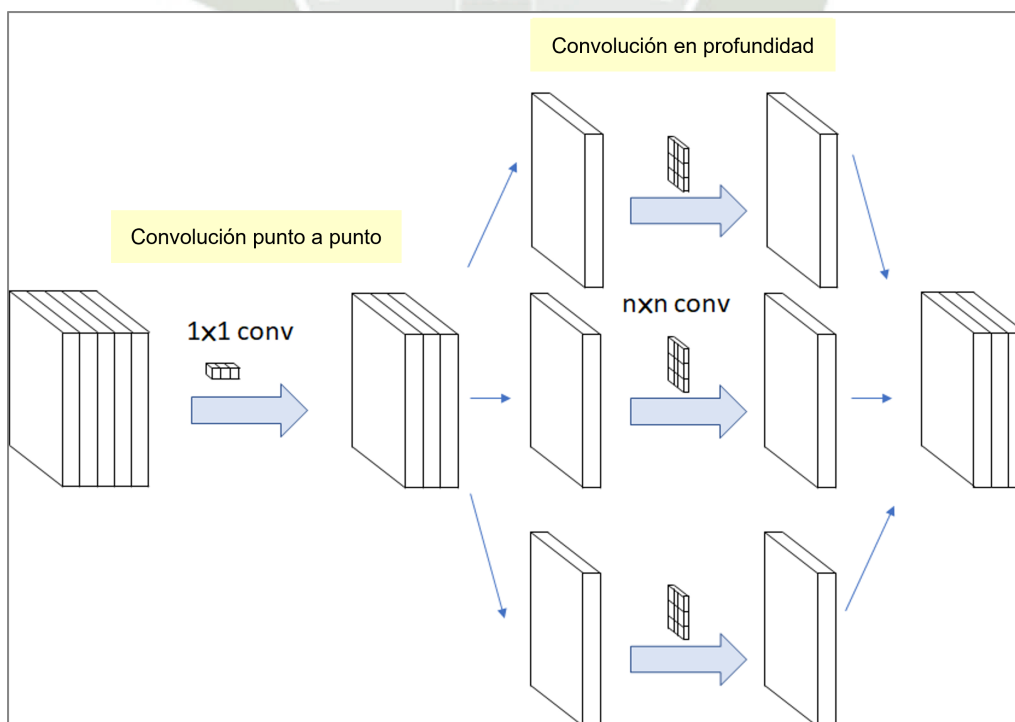
Fuente: Extraído de He et al. (2015).

**Figura 2.25:** *Convolution Separable por Profundidad Original.*



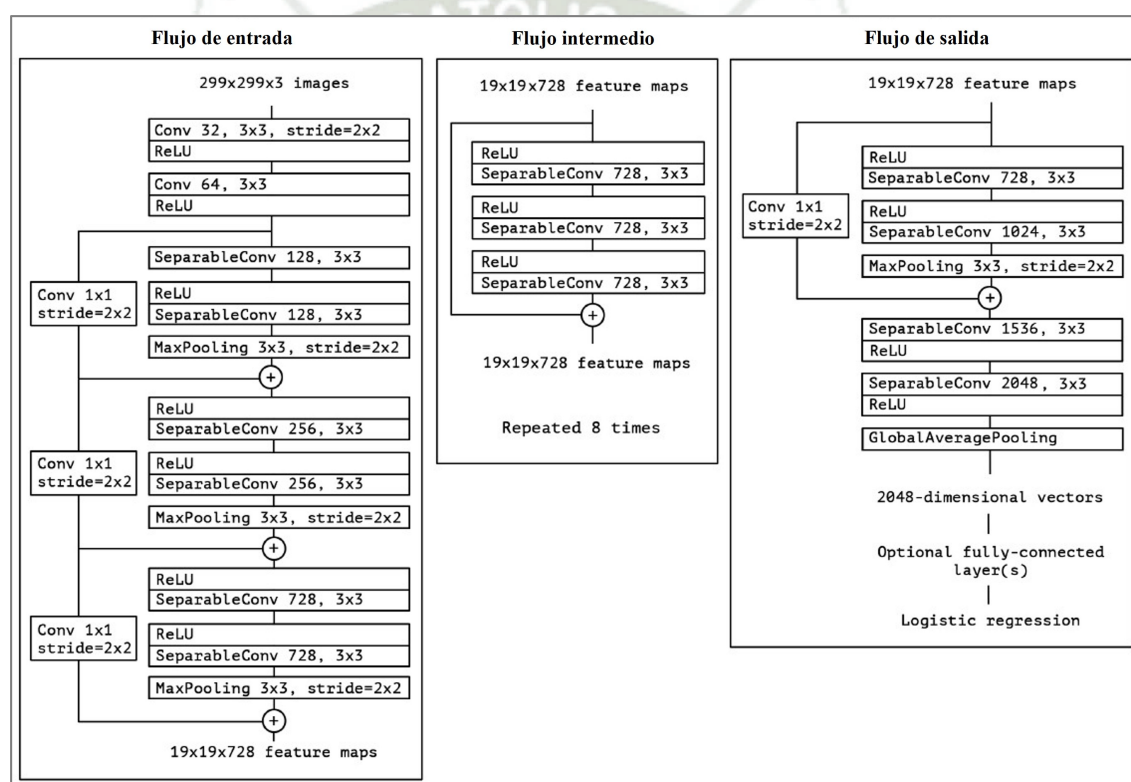
Fuente: Adaptado de Ko Min-soo (2023).

**Figura 2.26:** *Convolution Separable por Profundidad Modificada en Xception.*



Fuente: Adaptado de Ko Min-soo (2023).

*Figura 2.27: Arquitectura de una red Xception.*



*Fuente: Adaptado de Chollet (2017).*



## CAPITULO III

## MARCO METODOLÓGICO

### 3.1. Alcances y Limitaciones

Los alcances que se han identificado son:

- Se trabajará con una muestra real de 44 personas con discapacidad visual de la Asociación Unión de Ciegos de Arequipa en el periodo 2023.
- La captura de imágenes del dispositivo de asistencia se dará en tiempo real.
- Se utilizará la gama media de la placa electrónica de computadora Raspberry Pi.
- Para la creación del conjunto de datos de billetes se trabajará con imágenes en formato jpg.
- Se trabajará en un entorno colaborativo de desarrollo de Google Colab Pro+.
- No se divulgará información personal de las partes involucradas en esta investigación.

Algunas de las limitaciones que se han identificado son:

- El dispositivo reconocerá billetes peruanos de (10, 20, 50 y 100) comprendidos en los años 2011 hasta 2021.
- El dispositivo no podrá reconocer billetes que estén en un mal estado o que no sea legible.
- El dispositivo no podrá detectar si los billetes son falsos o no.
- La cantidad de GPUs que se necesitan para el entrenamiento de los modelos puede variar según el uso y la cantidad de pruebas que se realicen.

### 3.2. Tipo y Nivel de Investigación

Considerando a Hernández Sampieri y Mendoza Torres (2018), el tipo de investigación es Aplicada - Transversal, ya que el propósito de la investigación es dar solución a un problema que aqueja a las personas con discapacidad visual; asimismo, el nivel de investigación es correlacional, pues se busca conocer la relación que existe entre las variables (dependiente e independientes).

#### 3.2.1. Tipo de Investigación

Aplicada - Transversal

#### 3.2.2. Nivel de Investigación

Correlacional

### 3.3. Universo, Población y Muestra

#### 3.3.1. Universo

El universo de la investigación esta constituido por personas con discapacidad visual en la ciudad de Arequipa. "En la región Arequipa habría más de 30 mil personas con discapacidad visual" (Chávez, 2022).

#### 3.3.2. Población

La población objetivo de este estudio son las personas con discapacidad visual de la Asociación Unión de Ciegos de Arequipa. Se considerarán personas adultas cuyas edades oscilan entre 20 y 50 años con deterioro de la visión grave y ceguera. La población puede incluir tanto a usuarios de dispositivos de asistencia existentes como a aquellos que no los utilizan actualmente. La población objeto del estudio está constituida por:

- Representante de la organización: 1
- Personas invidentes: 50

#### 3.3.3. Muestra

La muestra se obtuvo haciendo uso de la fórmula del *procedimiento para estimar el tamaño de la muestra representativa para una población finita* definida por (Bernal, 2010):

$$n = \frac{Z^2 \times P \times Q \times N}{E^2(N - 1) + Z^2 \times P \times Q} \quad (3.1)$$

Donde:

- n: Tamaño de la muestra por estimar.  
Z: Nivel de confianza o margen de confiabilidad (95 %, es decir, Z=1.96).  
N: Tamaño de la población (Número).  
P: Probabilidad de que ocurra el evento estudiado. En este caso P = 50 % ó 0.5.  
Q = 1-P: Probabilidad de que no ocurra el evento estudiado. En este caso Q = 0.5.  
E: Error de estimación máximo (diferencia máxima entre la proporción muestral y la proporción proporcional. Se calcula en base al nivel de confianza definido para el estudio). En este caso E = 5 % ó 0.05.

Reemplazando los datos en la ecuación 3.1:

$$n = \frac{1,96^2 \times 0,5 \times 0,5 \times 50}{0,05^2(50 - 1) + 1,96^2 \times 0,5 \times 0,5} = 44 \quad (3.2)$$

Esta muestra nos permitió entrevistar a 44 personas con discapacidad visual de la Asociación Unión de Ciegos de Arequipa para extraer información sobre su experiencia con herramientas de asistencia a su alcance.

## 3.4. Métodos, Técnicas e Instrumentos de Recolección de Datos

### 3.4.1. Metodología de la investigación

El estudio utilizó una metodología iterativa “basada en la acción” y siguió el proceso de cinco pasos del enfoque Design Thinking (es decir, empatizar, definir, idear, prototipar y evaluar). (Geerts et al., 2023)

### 3.4.2. Técnicas de recolección de datos

Se consideran las siguientes técnicas de recolección de datos:

- Mapa de empatía (Ver apéndice 5.1).
- Entrevistas estructuradas y semiestructuradas con expertos y personas con discapacidad visual.
- Encuestas para recopilar datos cuantitativos sobre las necesidades y experiencias de las personas con discapacidad visual.
- Observación participante para obtener información sobre el uso de dispositivos de asistencia existentes.
- Observación participante para obtener información sobre el uso del dispositivo de asistencia propuesto en la presente investigación.
- Captura de imágenes para el entrenamiento del módulo de visión artificial.

### 3.4.3. Instrumentos de recolección de datos

Los instrumentos de recolección de datos incluirán:

- Guías de entrevistas para entrevistas con expertos y personas con discapacidad visual.
- Cuestionarios para las encuestas.
- Cámaras y dispositivos de captura de imágenes para la recopilación de datos visuales.

### 3.4.4. Técnicas de procesamiento de datos

El tipo de análisis que se empleará en la investigación es del tipo descriptivo e inferencial. Asimismo, se consideran las siguientes técnicas de procesamiento de datos:

- **Procesamiento de imágenes:** El procesamiento de imágenes es fundamental para extraer características visuales relevantes de los billetes. Se aplicarán técnicas de segmentación de imágenes, de extracción de características, de reconocimiento de patrones, y aprendizaje automático para identificar y clasificar los elementos de interés.

- **Aprendizaje profundo (Deep Learning):** Se emplearán técnicas de aprendizaje automático, con un enfoque en el aprendizaje profundo, para entrenar el modelo de reconocimiento de billetes. Se utilizarán algoritmos basados en modelos de aprendizaje profundo preentrenados, como las redes neuronales convolucionales (CNN), con el propósito de procesar y clasificar los billetes detectados de manera efectiva.
- **Modelos de aprendizaje profundo preentrenados:** La transferencia de aprendizaje implica utilizar un modelo preentrenado como punto de partida y adaptarlo para la tarea deseada. En este caso, se aprovechará el conocimiento previo de los modelos de aprendizaje profundo preentrenados (mediante la técnica de transferencia de *Transfer Learning*) y se ajustarán para el reconocimiento de billetes mediante el proceso de *Fine-Tuning*.
- **Técnicas de visión artificial:** Las técnicas de visión artificial se centran en el procesamiento de imágenes y la extracción de información visual. Se utilizarán técnicas como el análisis de contornos, reconocimiento de formas y la detección de billetes para identificar los elementos de interés en las imágenes de los billetes.

### 3.4.5. Herramientas para el procesamiento de datos

Dado que se trabajará en un entorno colaborativo de desarrollo, como es el caso de Google Colaboratory, se hará uso de las siguientes herramientas para el procesamiento de los datos:

- **OpenCV:** Una biblioteca ampliamente utilizada para el procesamiento de imágenes y visión artificial en Python.
- **TensorFlow y Keras:** Ambos frameworks de aprendizaje profundo que proporcionan herramientas y modelos preentrenados que serán reutilizados y adaptados para el reconocimiento de billetes.
- **Scikit-learn:** Biblioteca para aprendizaje automático en Python que ofrece algoritmos de clasificación y herramientas de evaluación de modelos.
- **NumPy y SciPy:** Bibliotecas para el procesamiento numérico y científico en Python, útiles para el procesamiento de señales y extracción de características.



## CAPITULO IV

## DESARROLLO DE LA PROPUESTA DE INVESTIGACIÓN

En este capítulo, en primer lugar, se detalla el proceso de creación del conjunto de datos de imágenes de billetes, un componente esencial para el desarrollo del sistema de asistencia. Este conjunto de datos es fundamental para entrenar y evaluar los modelos de reconocimiento de billetes.

### 4.1. Desarrollo en Visión Artificial

#### 4.1.1. Generación del Conjunto de Datos

##### 4.1.1.1. Configuración del Equipo de Captura de Imágenes

La creación del conjunto de datos comenzó con la configuración del equipo de captura de imágenes necesario para adquirir imágenes de alta calidad de billetes. Para este propósito, se utilizó un dispositivo móvil inteligente (smartphone) equipado con una cámara de alta resolución y configuraciones adecuadas para garantizar la calidad de las imágenes capturadas.

Tomando en cuenta lo anterior, en la tabla 4.1 se muestran las características del dispositivo móvil inteligente utilizado para la captura de imágenes de billetes:

**Tabla 4.1**

*Características del dispositivo de captura de imágenes billetes*

Característica	Descripción
Procesador	Snapdragon 685
RAM	8 GB LPDDR4X
Almacenamiento	128 GB
Cámara Frontal	13 megapíxeles f/2.45
Cámara Trasera	Principal: 50 Mpx f/1.8 Gran angular: 8 Mpx f/2.2 Macro: 2 Mpx f/2.4
Sistema Operativo	Android 12 + MIUI 14

##### 4.1.1.2. Procedimiento para la Captura de Imágenes

Se diseñó un procedimiento que aborda la captura de imágenes de manera estratégica, considerando diferentes aspectos para garantizar la efectividad del entrenamiento de los modelos.

#### 1. Planificación de la Captura de Imágenes

**Denominación de los Billetes:** Se realizó la captura de imágenes de billetes con denominaciones de S/ 10, S/ 20, S/ 50 y S/ 100, entre familias de billetes nuevas y antiguas.

**Diseño de un Plan de Captura Detallado:** Con la finalidad de garantizar el entrenamiento efectivo de los modelos de reconocimiento de billetes y evaluar su rendimiento de manera sólida, se ha diseñado un plan de captura

de imágenes estratégico que abarca tres conjuntos de datos: entrenamiento, validación y pruebas.

#### a) **Conjunto de Datos de Entrenamiento**

El objetivo de este procedimiento fue capturar imágenes significativas que permitan a los modelos aprender las características distintivas de los billetes y sus partes, tanto en el anverso como en el reverso, en fondos blancos y neutros (entiéndase a fondos neutros como aquellos fondos que son visualmente distintos y contrastantes con los billetes, permitiendo una clara separación entre el billete y su entorno), para aislar los billetes y facilitar su detección y reconocimiento.

##### ***Plan de Captura:***

- **Imagen de Billete Completo (Anverso y Reverso):** Capturar varias imágenes de cada denominación de billete (10, 20, 50 y 100 nuevos soles) en condiciones controladas con fondos blancos y neutros. Estas imágenes representarán los billetes en su totalidad y serán esenciales para que los modelos comprendan las características globales de los billetes. La imagen 4.1 muestra un ejemplo de imagen de billete completo.

**Figura 4.1:** *Captura de billetes completos (Anverso y Reverso).*



- **Detalles de billetes:** Capturar imágenes de partes específicas de los billetes, como detalles de los laterales, personajes y denominación del billete. Esto ayudará a los modelos a identificar y diferenciar características detalladas de los billetes. La imagen 4.2 muestra un ejemplo de detalles de billetes.

#### b) **Conjunto de Datos de Validación**

El objetivo de este procedimiento fue representar situaciones del mundo real donde se utilizará el dispositivo de asistencia. Estas imágenes se utilizarán para validar el rendimiento de los modelos en condiciones más diversas.

##### ***Plan de Captura:***

- **Entornos cotidianos:** Capturar imágenes de billetes en escenarios de la vida cotidiana, como habitaciones, parques, calles, etc. Estas imágenes reflejarán situaciones reales de uso del dispositivo de asistencia.

*Figura 4.2: Captura de detalles de billetes (Anverso y Reverso).*



- Distintas condiciones de iluminación:** Realizar capturas bajo diferentes condiciones de iluminación, incluyendo luz natural y artificial, para que los modelos puedan adaptarse a diversas situaciones lumínicas. La imagen 4.3 muestra la captura de billetes en entornos cotidianos y con distintas condiciones de iluminación.

*Figura 4.3: Captura de billetes en distintos entornos y condiciones de iluminación.*



### c) Conjunto de Datos de Pruebas

Este procedimiento tiene como objetivo utilizar parte de las imágenes de validación para evaluar el rendimiento de los modelos en condiciones diversas y verificar su capacidad de reconocimiento.

#### **Plan de Captura:**

- Selección aleatoria:** Seleccionar aleatoriamente un subconjunto de imágenes de validación para las pruebas de los modelos. Esto garantiza que las pruebas sean imparciales y representativas de las imágenes reales.

Este plan de captura de imágenes estratégico asegura que los modelos de reconocimiento de billetes se entrenen de manera completa y se validen de manera sólida en situaciones del mundo real. El uso de fondos para el conjunto de entrenamiento y la diversidad de escenarios en los conjuntos de validación y

pruebas contribuye al desarrollo del sistema de asistencia de manera precisa y confiable.

La tabla 4.2 muestra la cantidad de imágenes recolectadas por tipo de denominación.

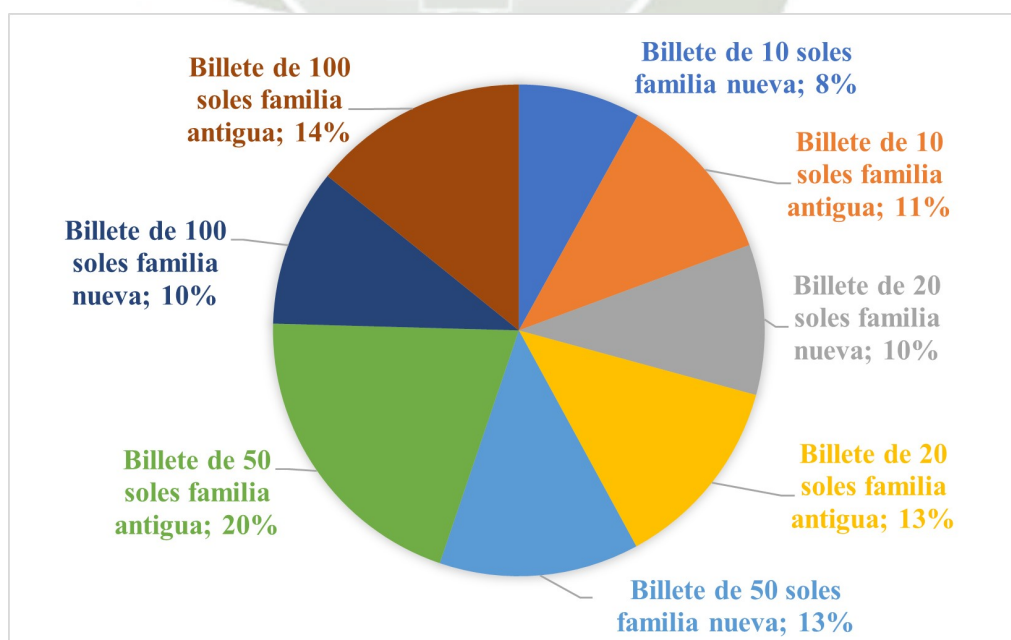
**Tabla 4.2**

*Tabla de Denominaciones Totales*

	Denominación				Total
	10	20	50	100	
Familia de billetes Nuevo	728	894	1190	939	3751
Familia de billetes Antiguo	1025	1162	1831	1285	5303
Total	1753	2056	3021	2224	9054

De acuerdo con la Tabla 4.2, la distribución de las denominaciones en el conjunto de datos se ilustra en la Figura 4.4. La Figura 4.4 muestra la representación de diferentes denominaciones de billetes. El billete de 10 Soles de la nueva familia constituye el 8% del conjunto de datos, mientras que el billete de 10 Soles de la antigua familia representa el 11%. El billete de 20 Soles de la nueva familia es el 10% del conjunto de datos, mientras que el billete de 20 Soles de la antigua familia representa el 12%. El billete de 50 Soles de la nueva familia constituye el 13% del conjunto de datos, mientras que el billete de 50 Soles de la antigua familia tiene la mayor participación con un 20%. El billete de 100 Soles de la nueva familia es el 10% del conjunto de datos, mientras que el billete de 100 Soles de la antigua familia es el 14%. (Cayturo-Silva et al., 2023)

**Figura 4.4:** Porcentaje de cada denominación de billete en el conjunto de datos.



## 2. Selección, Etiquetado y Preprocesamiento Manual de las Imágenes

A continuación, se detalla el proceso de selección, etiquetado y preprocesamiento manual de las imágenes adquiridas durante la fase de captura de datos. Este

paso es esencial para preparar las imágenes y utilizarlas en el entrenamiento y evaluación de los modelos de aprendizaje profundo.

#### a) Selección de Imágenes

Las imágenes capturadas se organizaron en 16 carpetas distintas, que servirán como clases para los modelos de aprendizaje profundo. Estas carpetas representan las diferentes combinaciones de denominaciones de billetes, caras (anverso, reverso) y familias de billetes (nuevo/antiguo). Las carpetas se nombran de la siguiente manera:

- billete10\_anverso\_antiguo
- billete10\_reverso\_antiguo
- billete10\_anverso\_nuevo
- billete10\_reverso\_nuevo
- billete20\_anverso\_antiguo
- billete20\_reverso\_antiguo
- billete20\_anverso\_nuevo
- billete20\_reverso\_nuevo
- billete50\_anverso\_antiguo
- billete50\_reverso\_antiguo
- billete50\_anverso\_nuevo
- billete50\_reverso\_nuevo
- billete100\_anverso\_antiguo
- billete100\_reverso\_antiguo
- billete100\_anverso\_nuevo
- billete100\_reverso\_nuevo

#### b) Etiquetado de Imágenes

Cada imagen se etiquetó siguiendo una nomenclatura específica que permite su identificación y categorización en el proceso de entrenamiento. La nomenclatura incluye:

- El tipo de conjunto al que pertenece la imagen (entrenamiento, validación o pruebas).
- La denominación del billete (10, 20, 50 o 100).
- La familia del billete (nuevo o antiguo).
- Una enumeración secuencial para identificar de manera única cada imagen.

Ejemplo de etiqueta: “entrenamiento\_10\_antiguo\_001”.

#### c) Preprocesamiento Manual de Imágenes

Inicialmente, las imágenes capturadas por el dispositivo móvil eran de alta calidad, con dimensiones de 3072 píxeles de ancho por 4080 píxeles de alto y un tamaño de archivo aproximado de 6.35 MB. Para adaptarlas al proceso de reconocimiento, se realizó un preprocesamiento manual de las imágenes utilizando la utilidad de “Cambio de tamaño de imágenes” de la herramienta Microsoft PowerToys (2023), disponible en la Microsoft Store.

Las imágenes fueron redimensionadas a un nuevo tamaño de 854 píxeles de ancho por 480 píxeles de alto, con un tamaño de archivo aproximado de 207 KB. Este proceso de redimensionamiento permitió que las imágenes fueran más adecuadas para el análisis y el entrenamiento de los modelos, manteniendo la calidad necesaria para el reconocimiento de billetes.

Este riguroso proceso de selección, etiquetado y preprocesamiento asegura que las imágenes estén listas para ser utilizadas en el entrenamiento y evaluación de los modelos de aprendizaje profundo.

Se puede acceder al conjunto de imágenes a través de los siguientes repositorios:

- **IEEE Dataport Cayturo Silva et al. (2023a):** <https://dx.doi.org/10.21227/1raa-mz71>
- **Mendeley Data Cayturo Silva et al. (2023b):** <https://data.mendeley.com/datasets/22jmt8xhpn/1>

## 4.1.2. Modelos de Aprendizaje Profundo

En esta etapa, se aplicaron modelos de aprendizaje profundo preentrenados en el lenguaje de programación Python para abordar la detección y el reconocimiento de los billetes peruanos. Cinco modelos preentrenados se utilizaron en particular: VGG16, VGG19, Xception, ResNet50 y MobileNetV2.

### 4.1.2.1. Entrenamiento de modelos de aprendizaje profundo preentrenados

#### Configuración del entorno de trabajo

Para llevar a cabo el entrenamiento de los modelos, se ha seleccionado un entorno colaborativo: Google Colab. Se cuenta con una suscripción mensual que brinda acceso a GPUs de alto rendimiento, en particular, la GPU V100. Esta elección de hardware permite acelerar significativamente el proceso de entrenamiento de los modelos.

#### Importación y Procesamiento de Imágenes

El proceso se inicia con la importación y el procesamiento de datos. La importación implica cargar un conjunto de datos de imágenes, mientras que el procesamiento es esencial para asegurar que los datos sean adecuados para el entrenamiento de los algoritmos. Como se mencionó previamente, cada algoritmo requiere imágenes con dimensiones específicas. Esto implica tareas como normalización, redimensionamiento y división de los datos en conjuntos de entrenamiento y prueba. Este proceso garantiza que los datos estén en el formato correcto y preparados para su uso en los modelos.

```
1 # Importando las librerías necesarias
2 import os
3 from keras.models import Model
4 from keras.optimizers import Adam, RMSprop
5 from keras.applications.vgg16 import VGG16, preprocess_input
6 from keras.preprocessing.image import ImageDataGenerator
7 from keras.callbacks import ModelCheckpoint, EarlyStopping
8 from keras.layers import Dense, Dropout, Flatten
```

```

9  from pathlib import Path
10 import numpy as np
11 from livelossplot.inputs.keras import PlotLossesCallback
12
13 # Aplicando la aumentacion y division de datos (Entrenamiento y
14   pruebas)
15 BATCH_SIZE = 64
16 train_generator = ImageDataGenerator(rotation_range=90,
17                                     brightness_range=[0.1, 0.7],
18                                     width_shift_range=0.5,
19                                     height_shift_range=0.5,
20                                     horizontal_flip=True,
21                                     vertical_flip=True,
22                                     validation_split=0.15,
23                                     # Preprocesamiento de VGG16
24                                     preprocessing_function=preprocess_input)
25
26 test_generator = ImageDataGenerator(preprocessing_function=
27   preprocess_input)
28
29 # Importando los datos
30 data_dir = Path('/content/drive/MyDrive/Tesis/New_Images_Dataset')
31
32 train_data_dir = data_dir/'training_set'
33 test_data_dir = data_dir/'validation_set'
34
35 # Creando generadores de flujo de datos para cargar lotes de
36   imagenes desde directorios
37 class_subset = sorted(os.listdir(data_dir/'training_set'))[:16]
38
39 traingen = train_generator.flow_from_directory(train_data_dir,
40   target_size=(224, 224),
41   class_mode='categorical',
42   classes=class_subset,
43   subset='training',
44   batch_size=BATCH_SIZE,
45   shuffle=True,
46   seed=42)
47
48 validgen = train_generator.flow_from_directory(train_data_dir,
49   target_size=(224, 224),
50   class_mode='categorical',
51   classes=class_subset,
52   subset='validation',
53   batch_size=BATCH_SIZE,
54   shuffle=True,
55   seed=42)
56
57 testgen = test_generator.flow_from_directory(test_data_dir,
58   target_size=(224, 224),
59   class_mode=None,
60   classes=class_subset,
61   batch_size=1,
62   shuffle=False,
63   seed=42)

```

Código 4.1: Importación de datos y procesamiento de imágenes

Donde:

1. **Importación de Bibliotecas y Parámetros:** El código comienza importando varias bibliotecas, incluyendo Keras, para construir y entrenar modelos, y configuraciones relacionadas con la operación del modelo, como la tasa de aprendizaje (utilizando los optimizadores Adam y RMSprop). También se define el tamaño de lote “`BATCH_SIZE`” que se utilizará durante el entrenamiento.
2. **ImageDataGenerator para Aumento de Datos y Preprocesamiento:** Se utilizan dos instancias de ImageDataGenerator. El primero “`train_generator`” que se utiliza para aumentar la cantidad de datos de entrenamiento y preprocesar las imágenes antes de alimentarlas al modelo. Esto incluye rotaciones aleatorias, ajuste de brillo, cambio de ancho y alto, y giro horizontal y vertical. También se reserva un 15 % de los datos para validación en cada época. El preprocesamiento de las imágenes se realiza de acuerdo con los requisitos de la arquitectura VGG16. El segundo generador es “`test_generator`”, el cual se utiliza para preprocesar las imágenes de prueba de acuerdo con la arquitectura del modelo.
3. **Importación de Datos:** Se especifica la ubicación de los datos de entrenamiento “`train_data_dir`” y los datos de validación “`test_data_dir`” en el sistema de archivos. Se obtiene una lista de las primeras 16 clases en los datos de entrenamiento, lo que limita el problema de clasificación a estas 16 clases. Esto se almacena en una lista de cadenas *class\_subset*.
4. **Creación de Generadores de Flujo de Datos:** Se crean tres generadores de flujo de datos *traingen*, *validgen* y *testgen* utilizando los datos de entrenamiento, validación y prueba respectivamente. Cada generador ajusta las imágenes a un tamaño de 224x224 píxeles y utiliza las clases especificadas en *class\_subset*. Los generadores para entrenamiento y validación se configuran para utilizar el modo ‘categórico’ ya que se trata de un problema de clasificación con etiquetas categóricas. Se especifica el tamaño de lote y se mezclan aleatoriamente los datos de entrenamiento. La semilla (seed) se utiliza para garantizar la reproducibilidad de la mezcla.
5. **Target Size Variable:** Este parámetro varía según el algoritmo. Esto se debe a que algunos modelos de redes neuronales, como VGG16, tienen requisitos específicos de tamaño de entrada. En este caso, VGG16 espera imágenes de 224x224 píxeles. Por lo tanto, el tamaño de la imagen se establece en 224x224 para coincidir con la entrada de esta arquitectura.

## Configuración de los modelos

Una vez que los datos están listos, se procede a la configuración de los modelos. Los modelos de aprendizaje profundo se importan haciendo uso de la biblioteca Keras. A continuación, se describen las configuraciones utilizadas. Los hiperparámetros son ajustados minuciosamente, equilibrando la precisión de los modelos con el tiempo de entrenamiento.

```
1 def create_model(input_shape, n_classes, optimizer='rmsprop',
2   fine_tune=0):
3     # Importar el modelo y excluir las ultimas capas
4     conv_base = VGG16(include_top=False,
5                       weights='imagenet',
6                       input_shape=input_shape)
7
8     # Se valida si se va a realizar el proceso de ajuste fino
9     if fine_tune > 0:
10        for layer in conv_base.layers[:-fine_tune]:
11            layer.trainable = False
12    else:
13        for layer in conv_base.layers:
14            layer.trainable = False
15
16    # Se anaden nuevas capas en la parte superior del modelo
17    top_model = conv_base.output
18    top_model = Flatten(name="flatten")(top_model)
19    top_model = Dense(4096, activation='relu')(top_model)
20    top_model = Dense(1072, activation='relu')(top_model)
21    top_model = Dropout(0.2)(top_model)
22    output_layer = Dense(n_classes, activation='softmax')(top_model)
23
24    model = Model(inputs=conv_base.input, outputs=output_layer)
25
26    # Se compila el modelo para el entrenamiento
27    model.compile(optimizer=optimizer,
28                 # Utilizado en problemas de clasificacion multiclase
29                 loss='categorical_crossentropy',
30                 metrics=['accuracy'])
31
32    return model
```

Código 4.2: Configuración del modelo

Donde, se define la función “create\_model”, la cual acepta varios parámetros, como el tamaño de entrada de las imágenes al modelo, el número de clases (en este caso, establecido en 16), el optimizador (con un valor predeterminado de 'rmsprop') y la bandera 'fine\_tune', que, en su estado inicial (0), determina que el proceso de entrenamiento se inicie sin ajuste fino, empleando la técnica de transferencia de aprendizaje. A continuación, se detalla el contenido y la funcionalidad de la función.

- **Importar el modelo y excluir las últimas capas:** La función comienza importando el modelo preentrenado, en este caso VGG16. Este modelo se carga con sus capas superiores excluidas (include\_top=False) porque se agregarán capas personalizadas para adaptar el modelo al reconocimiento de billetes.
- **Ajuste fino (fine-tuning):** El código permite realizar un proceso de ajuste fino (fine-tuning) si se especifica un valor mayor que cero para el parámetro

'fine\_tune'. Esto significa que algunas de las capas superiores del modelo preentrenado se harán entrenables para adaptarse a la nueva tarea. Si fine\_tune es cero o negativo, todas las capas del modelo preentrenado se mantendrán no entrenables. El ajuste fino se utiliza cuando se desea adaptar un modelo preentrenado a una tarea específica sin cambiar completamente sus representaciones.

- **Añadir nuevas capas:** Después de importar el modelo preentrenado, se agregan nuevas capas en la parte superior del mismo. Esto se hace mediante la definición de una secuencia de capas que procesan las características extraídas por el modelo base. En este caso, se agregan capas de aplanamiento (Flatten) para convertir las características en un vector unidimensional, capas densas (Dense) para la clasificación y una capa de eliminación (Dropout) para evitar el sobreajuste.
  - *Flatten*: Transforma las características extraídas en un vector 1D para alimentar las capas densas.
  - *Dense(4096, activation='relu')*: Una capa densa con 4096 canales y función de activación ReLU.
  - *Dense(1072, activation='relu')*: Otra capa densa con 1072 canales y función de activación ReLU.
  - *Dropout(0.2)*: Una capa de eliminación que desactiva aleatoriamente el 20% de las neuronas para evitar el sobreajuste.
  - *Dense(n\_classes, activation='softmax')*: La capa de salida con tantos canales como clases de salida y función de activación 'softmax' para la clasificación.
- **Compilar el modelo:** Finalmente, se compila el modelo. Durante el proceso de compilación, se configuran aspectos esenciales para el entrenamiento, como el optimizador, la función de pérdida ('categorical\_crossentropy' en este caso, adecuada para problemas de clasificación multiclase) y las métricas (en este caso, 'accuracy' para evaluar la precisión del modelo durante el entrenamiento).

### Proceso de entrenamiento y fine-tuning

Se aplicó la transferencia de aprendizaje para el entrenamiento del modelo con el nuevo conjunto de datos, seguido de un proceso de fine-tuning para adaptar el modelo al reconocimiento de dicho conjunto de datos.

#### 1. Transferencia de Aprendizaje (Transfer-learning)

```
1 # Definición de hiperparametros
2 input_shape = (224, 224, 3)
3 optim_1 = Adam(learning_rate=0.0001)
4 n_classes=16
5
6 n_steps = traingen.samples
7 n_val_steps = validgen.samples
8 n_epochs = 100
9
10 # Entrenamos el modelo sin fine-tuning
```

```
11 vgg_model = create_model(input_shape, n_classes, optim_1,
    fine_tune=0)
12
13 plot_loss_1 = PlotLossesCallback()
14
15 # Proceso de entrenamiento
16 # EarlyStopping
17 early_stop = EarlyStopping(monitor='val_loss',
    patience=10,
    restore_best_weights=True,
    mode='min')
21
22 %%time
23 vgg_history = vgg_model.fit(traingen,
    batch_size=BATCH_SIZE,
    epochs=n_epochs,
    validation_data=validgen,
    steps_per_epoch=n_steps,
    validation_steps=n_val_steps,
    callbacks=[tl_checkpoint_1, early_stop, plot_loss_1],
    verbose=1)
30
```

Código 4.3: Transferencia de Aprendizaje en el Entrenamiento

Donde:

- **Definición de Hiperparámetros:** Se inicia con la definición de los hiperparámetros. Se establece 'input\_shape' como (224, 224, 3), indicando que el modelo espera imágenes de entrada de  $224 \times 224$  píxeles con tres canales de color (RGB). Además, se configura un optimizador llamado 'optim\_1' utilizando el algoritmo Adam con una tasa de aprendizaje de 0.0001. La tasa de aprendizaje es crítica para determinar la velocidad de aprendizaje del modelo, influyendo en la estabilidad y la convergencia del entrenamiento. Se especifica el número de clases con 'n\_classes', que en este contexto es igual a 16, lo que señala que se está entrenando el modelo para clasificar datos en 16 categorías diferentes. Las variables 'n\_steps' y 'n\_val\_steps' se emplean para almacenar el número total de pasos que se realizarán durante el entrenamiento y la validación, respectivamente. La variable 'n\_epochs' indica cuántas épocas de entrenamiento se llevarán a cabo.
- **Creación del Modelo sin fine-tuning:** Posteriormente, se crea el modelo de red neuronal con los hiperparámetros definidos anteriormente. Se emplea la función 'create\_model' y se envían como parámetros las variables *input\_shape*, *n\_classes*, *optim\_1* y *fine\_tune* (se establece como 0).
- **Callback de EarlyStopping:** Este callback monitoriza la métrica de pérdida en el conjunto de validación durante el proceso de entrenamiento. Si esta métrica no muestra mejoras tras un número determinado de épocas (10 en este caso), el entrenamiento se detendrá anticipadamente. El parámetro 'restore\_best\_weights=True' garantiza que el modelo regrese a los mejores pesos guardados si el entrenamiento se interrumpe. Esta estrategia es utilizada para prevenir el sobreajuste del modelo durante el entrenamiento.

- **Entrenamiento del modelo con transfer-learning:** Finalmente, se entrena el modelo utilizando los conjuntos de entrenamiento y validación. Se emplea el callback 'PlotLossesCallback' para representar en tiempo real la pérdida del modelo durante el entrenamiento. Se mide el tiempo total de ejecución del proceso de entrenamiento utilizando '%time'. Esta medición ofrece una estimación de la duración del proceso en su totalidad.

## 2. Ajuste fino (Fine-tuning)

```
1 # Restablecer los generadores de datos de imágenes
2 traingen.reset()
3 validgen.reset()
4 testgen.reset()
5
6 # Iniciar con el proceso de fine-tuning
7 optim_2 = Adam(learning_rate=0.0001)
8
9 vgg_model_ft = create_model(input_shape, n_classes, optim_2,
    fine_tune=2)
10
11 %%time
12 plot_loss_2 = PlotLossesCallback()
13
14 # Retrain model with fine-tuning
15 vgg_ft_history = vgg_model_ft.fit(traingen,
16     batch_size=BATCH_SIZE,
17     epochs=n_epochs,
18     validation_data=validgen,
19     steps_per_epoch=n_steps,
20     validation_steps=n_val_steps,
21     callbacks=[tl_checkpoint_1, early_stop, plot_loss_2],
22     verbose=1)
```

Código 4.4: Ajuste Fino para la Adaptación de Datos

Donde:

- **Restablecer los generadores de datos de imágenes:** Antes de aplicar el ajuste fino, se restablecer los generadores de datos de imágenes. Al restablecer estos generadores, se asegura que se comience desde el principio del conjunto de datos, lo que es esencial para el proceso de fine-tuning.
- **Entrenamiento del modelo con fine-tuning:** Se comienza configurando un nuevo optimizador Adam con una tasa de aprendizaje de 0.0001. El fine-tuning implica ajustar las últimas capas del modelo para adaptarlo al conjunto de datos de imágenes de billetes. Se crea una instancia del modelo utilizando la función 'create\_model', y se establece la variable 'fine\_tune' en 2. Este valor de fine\_tune indica que se ajustarán las últimas dos capas del modelo.

### 4.1.3. Resultados obtenidos

#### 4.1.3.1. Configuración de hiperparámetros

En la tabla 4.3, se presentan los valores empleados en la configuración de los hiperparámetros de los modelos de aprendizaje profundo. Es importante destacar que todos los modelos se entrenaron bajo condiciones uniformes, lo que abarca la aplicación de los mismos hiperparámetros, la utilización del aumento de datos (Data Augmentation), un número equivalente de épocas y el preprocesamiento de las imágenes mediante la biblioteca Keras.

**Tabla 4.3**

*Configuración de hiperparámetros en los modelos*

N°	Nombre	Aumento de datos	N° de Épocas	Preprocesamiento con Keras	Hiperparámetros			
					Batch Size	Optimizador	Tasa de Aprendizaje	Función de Activación
1	VGG16	Si	100	Si	64	Adam	0.0001	ReLU
2	VGG19	Si	100	Si	64	Adam	0.0001	ReLU
3	Xception	Si	100	Si	64	Adam	0.0001	ReLU
4	ResNet50	Si	100	Si	64	Adam	0.0001	ReLU
5	MobileNetV2	Si	100	Si	64	Adam	0.0001	ReLU

Donde:

- **Aumento de datos:** Se aplicó la técnica de aumento de datos durante el entrenamiento, lo cual ayudó a incrementar la diversidad de los datos y mejorar la capacidad del modelo para generalizar a partir del conjunto de datos.
- **N° de Épocas:** El número de épocas se estableció en 100 para todos los modelos.
- **Preprocesamiento con Keras:** Indica que se realizó un preprocesamiento de datos específico con Keras. Esto incluye tareas como el redimensionamiento de imágenes y escalado de píxeles, y se aplicó a todos los modelos. Cabe mencionar que cada algoritmo posee su propio método de preprocesamiento (Véase el código 4.1 en el parámetro de 'preprocessing\_function').
- **Hiperparámetros:** Estas configuraciones afectan el proceso de entrenamiento. Se aplicaron los mismos valores a todos los modelos (Meshram et al., 2023):
  - **Batch size (Tamaño de lote):** Se estableció en 64, lo que significa que se utilizan lotes de 64 imágenes para cada iteración durante el entrenamiento.
  - **Optimizador (Optimizer):** Se utilizó el optimizador Adam en todos los modelos.
  - **Tasa de Aprendizaje (Learning Rate):** La tasa de aprendizaje se estableció en 0.0001 para todos los modelos. Luego de realizar pruebas, se definió que es la tasa de aprendizaje más adecuada para ajustar los pesos correctamente durante el entrenamiento del modelo.
  - **Función de Activación (Activation Function):** Se empleó la función de activación ReLU para todas las capas de los modelos.

#### 4.1.3.2. Duración del entrenamiento y precisión de los modelos

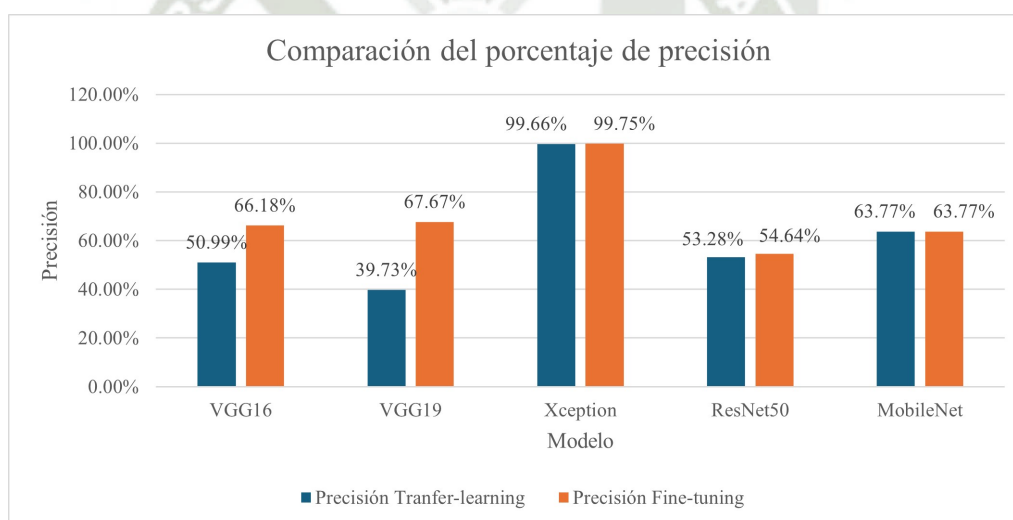
La tabla 4.4 ofrece una comparación entre el tiempo de entrenamiento y la precisión de los modelos tanto en el proceso de transferencia de aprendizaje como después de aplicar el ajuste fino. Esto nos permite evaluar cómo varía la eficiencia en términos de tiempo y el rendimiento de los modelos en cada fase del entrenamiento.

**Tabla 4.4**

*Duración del entrenamiento y precisión de los modelos*

N°	Nombre	Walk time		Precisión	
		Tranfer-learning	Fine-tuning	Tranfer-learning	Fine-tuning
1	VGG16	1h 47min 47s	1h 18min 7s	50.99%	66.18%
2	VGG19	1h 10min 22s	40min 18s	39.73%	67.67%
3	Xception	3h 12min 30s	1h 50min 56s	99.66%	99.75%
4	ResNet50	1h 40min 0s	1h 40min 0s	53.28%	54.64%
5	MobileNetV2	1h 56min 36s	1h 42min 21s	63.77%	63.77%

**Figura 4.5:** Comparación de la precisión entre transfer-learning y fine-tuning.



Donde:

- Tiempo de ejecución (Walk time):** Este valor representa el tiempo que llevó entrenar cada modelo en horas, minutos y segundos. Se observa que el tiempo de entrenamiento varía significativamente entre los modelos. Por ejemplo, VGG19 tiene uno de los tiempos de entrenamiento más cortos, con aproximadamente 1 hora y 10 minutos, mientras que Xception es el modelo que requiere más tiempo, con alrededor de 3 horas y 12 minutos.
- Precisión (Accuracy):** La precisión es un indicador crítico del rendimiento del modelo. Se mide en porcentaje y muestra cuántas de las predicciones del modelo son correctas. Se notan diferencias considerables en la precisión entre los modelos. Xception muestra la mayor precisión tanto en Transfer learning (99.66%) como en Fine-tuning (99.75%). Por otro lado, VGG19 tiene una precisión relativamente baja en Transfer learning (39.73%), que mejora significativamente con Fine-tuning (67.67%).

- **Comparación de Transfer learning y Fine-tuning:** Es importante destacar que, en la mayoría de los casos, Fine-tuning mejora la precisión en comparación con Transfer learning. Este resultado sugiere que Fine-tuning es una técnica efectiva para adaptar un modelo preentrenado a un nuevo conjunto de datos y mejorar su rendimiento. En algunos casos, como MobileNetV2, la precisión es la misma en ambas técnicas, lo que indica que Transfer learning es suficiente para lograr un buen rendimiento.
- **Equilibrio entre tiempo de ejecución y precisión:** VGG requiere menos tiempo con Fine-tuning, pero también logra una precisión notablemente alta. En contraste, Xception ofrece una alta precisión pero a expensas de un mayor tiempo de entrenamiento.

## 4.2. Desarrollo en IoT

### 4.2.1. Diseño del dispositivo IoT

Para el desarrollo del prototipo, se toman en cuenta los siguientes requisitos funcionales y no funcionales.

#### 4.2.1.1. Requisitos Funcionales

Las tablas 4.5, 4.6 y 4.7, detallan los requerimientos funcionales del dispositivo de asistencia.

**Tabla 4.5**

*Requerimiento Funcional 1: Captura de Billetes*

<b>Código del requerimiento</b>	RF01
<b>Nombre</b>	Captura de Billetes
<b>Propósito</b>	El dispositivo permitirá la captura de imágenes de billetes peruanos con denominaciones de 10, 20, 50 y 100 nuevos soles.
<b>Descripción</b>	El dispositivo estará equipado con una cámara de alta resolución que permitirá la captura de imágenes de billetes. El usuario activará la función de captura mediante un botón. Una vez que se ha capturado la imagen del billete, esta se almacenará en el Raspberry Pi 4B+ para su posterior procesamiento.
<b>Entrada</b>	La entrada principal para este requerimiento es la solicitud del usuario para capturar una imagen de un billete específico. Esto se logra a través de la interacción del usuario con la interfaz física del dispositivo, que incluye un botón de reconocimiento.
<b>Salida</b>	La salida de este requerimiento es la imagen capturada del billete, que se utilizará como entrada para el procesamiento y reconocimiento del billete mediante el Raspberry Pi 4B+.
<b>Prioridad</b>	Alta

Tabla 4.6

*Requerimiento Funcional 2: Identificación de Billetes*

<b>Código del requerimiento</b>	RF02
<b>Nombre</b>	Identificación de Billetes
<b>Propósito</b>	El dispositivo identificará de manera precisa y confiable el valor de los billetes capturados.
<b>Descripción</b>	El dispositivo estará equipado con un sistema de procesamiento de imágenes basado en visión artificial que procesará y generará la predicción sobre el billete capturado para determinar su valor. El sistema utilizará algoritmos de aprendizaje profundo previamente entrenados para llevar a cabo esta tarea. Tanto el procesamiento como la predicción se realizarán en el Raspberry Pi 4B+.
<b>Entrada</b>	La entrada para este requerimiento son las imágenes de los billetes capturados por la cámara del dispositivo.
<b>Salida</b>	La salida de este requerimiento es la clase predicha por el algoritmo de visión artificial. Esta clase servirá para proporcionar información auditiva al usuario, que indica el valor del billete capturado.
<b>Prioridad</b>	Alta

Tabla 4.7

*Requerimiento Funcional 3: Salida de Audio*

<b>Código del requerimiento</b>	RF03
<b>Nombre</b>	Salida de Audio
<b>Propósito</b>	El dispositivo proporcionará información auditiva sobre la denominación del billete reconocido. También representa una forma de comunicar de manera comprensible al usuario con discapacidad visual.
<b>Descripción</b>	Cuando el sistema de visión artificial identifica y clasifica los billetes a través de las imágenes capturadas, debe generar una salida en forma de audio que describa la identificación realizada. La voz generada será reproducida a través de altavoces conectados al prototipo o auriculares.
<b>Entrada</b>	La entrada de este requerimiento es la clase predicha por el algoritmo de visión artificial; que indica el valor del billete capturado.
<b>Salida</b>	La salida de este requerimiento es la información de voz generada a partir de los resultados de la identificación. La voz generada se reproduce a través de los altavoces conectados al prototipo o mediante auriculares conectados a este.
<b>Prioridad</b>	Alta

#### 4.2.1.2. Requisitos no Funcionales

La tabla 4.8 lista los requerimientos no funcionales del dispositivo de asistencia.

**Tabla 4.8**

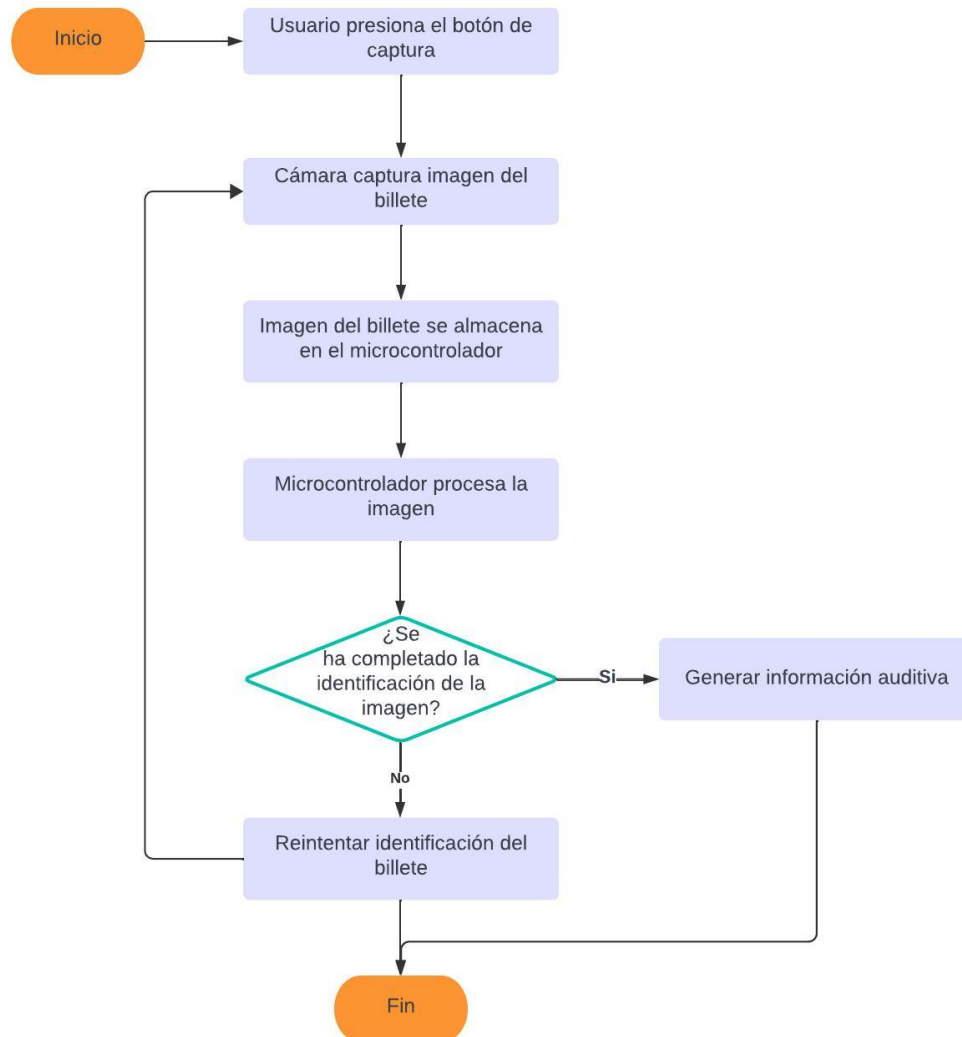
*Requerimientos no Funcionales*

<b>Usabilidad</b>	<p><b>Facilidad de Uso:</b> El diseño del prototipo se orientará hacia la simplicidad y claridad, garantizando una experiencia de uso intuitiva para personas con discapacidad visual. Los botones y la salida de audio serán accesibles y fácilmente comprensibles.</p> <p><b>Retroalimentación Clara:</b> La retroalimentación auditiva proporcionada por el prototipo será diseñada para ser clara y comprensible, asegurando que la información sobre la denominación del billete sea fácilmente entendida por el usuario.</p> <p><b>Ergonomía y Comodidad:</b> Se prestará especial atención al diseño ergonómico del prototipo, considerando el peso, tamaño y facilidad de transporte. Esto garantizará que el dispositivo sea cómodo de llevar y utilizar en el día a día.</p>
<b>Confiabilidad</b>	<p><b>Funcionamiento Confiable:</b> El prototipo se desarrollará para garantizar un funcionamiento constante y confiable. La precisión y velocidad del reconocimiento de billetes se mantendrán consistentes para evitar inconvenientes y frustraciones del usuario.</p> <p><b>Precisión:</b> Se establecerá un objetivo de precisión entre el 86 % y 90 %, minimizando al máximo los errores en el sistema de reconocimiento de billetes. La exactitud de la información es crucial para la utilidad del prototipo.</p>
<b>Portabilidad</b>	<p>Se diseñará el prototipo con la portabilidad en mente, permitiendo a las personas con discapacidad visual llevarlo consigo de manera fácil y cómoda.</p>
<b>Eficiencia</b>	<p><b>Eficiencia Energética:</b> El prototipo priorizará la eficiencia energética para maximizar la duración de la batería, apuntando a una autonomía de aproximadamente 3 horas. Esto asegurará un uso prolongado sin la necesidad constante de recarga.</p> <p><b>Procesamiento Rápido:</b> El sistema de reconocimiento de billetes se diseñará para un procesamiento rápido, con un tiempo objetivo de 2 segundos. Esto garantizará retroalimentación oportuna al usuario sin retrasos significativos.</p>
<b>Rendimiento</b>	<p><b>Detección en Variadas Condiciones:</b> El prototipo se probará y optimizará para ofrecer un rendimiento óptimo en la detección y reconocimiento de billetes en diversas condiciones ambientales y de iluminación.</p>
<b>Espacio</b>	<p>El prototipo se construirá en base a dimensiones aproximadas de 8 a 10 cm de ancho y 10 cm de largo. Esto permitirá su integración sin ocupar excesivo espacio y se adaptará a las preferencias individuales de los usuarios.</p>

#### 4.2.1.3. Flujo del funcionamiento del dispositivo de asistencia

En este sentido, en la figura 4.6 se muestra el flujo del funcionamiento del dispositivo de asistencia.

**Figura 4.6:** Diagrama de flujo del funcionamiento del dispositivo de asistencia.



El proceso inicia cuando el usuario presiona el botón de captura, activando la cámara Raspberry Pi para capturar una imagen del billete correspondiente. Posteriormente, la imagen se envía al Raspberry Pi 4B+, que realiza el procesamiento utilizando un sistema de visión artificial. El diagrama incorpora una lógica de verificación para determinar si la identificación del billete ha sido exitosa. Si es así, el flujo continúa hacia la generación de información auditiva basada en la denominación del billete identificado, proporcionando así retroalimentación al usuario. En caso contrario, se permite reintentar la identificación, volviendo al proceso de captura de la imagen. El proceso concluye después de generar la información auditiva.

#### 4.2.1.4. Diseño del Circuito

Por otro lado, para el diseño del circuito del dispositivo, se utilizaron los siguientes componentes, tomando en cuenta los requerimientos funcionales y no funcionales. Para visualizar el plan de costos del proyecto véase el apéndice 5.2.

- **Placa electrónica de computadora Raspberry Pi 4B+ 2GB de RAM:** Potente placa de computadora de bajo costo y bajo consumo energético. Equipada con un procesador de cuatro núcleos y una memoria RAM de 1GB, esta placa ofrece un rendimiento excepcional para realizar tareas de procesamiento de imágenes y generación de salida de audio en dispositivos IoT. En el anexo 5.6 se realiza una comparación del Raspberry Pi 4B+ con otras placas electrónicas del mercado.
- **Cámara Raspberry Pi de 5MP:** Componente diseñado específicamente para su uso con las placas Raspberry Pi. Esta cámara proporciona capacidades de captura de imágenes y video de alta calidad, lo que permite al dispositivo IoT capturar contenido visual con claridad y precisión.
- **Driver de amplificador de sonido Adafruit Max98357:** Componente que amplifica la señal de audio para mejorar la calidad del sonido emitido por el altavoz. Este driver de clase D es eficiente en términos de energía y es adecuado para aplicaciones de audio en dispositivos de bajo consumo.
- **Convertidor de Voltaje de 7.4V a 5V:** Dispositivo que convierte la energía eléctrica de 7.4V a 5V, proporcionando una fuente de alimentación adecuada para el Raspberry Pi 4B+. Dado que se está trabajando con dos baterías de 3.7V se hace necesario este dispositivo para disminuir el voltaje y amperaje.
- **Driver de carga para baterías con conexión tipo C:** Componente que permite cargar las baterías de litio del dispositivo de forma rápida y segura mediante una conexión USB tipo C.
- **Batería de Litio de 3.7V y 1000mA:** Fuente de energía portátil que proporciona alimentación al dispositivo. Esta batería recargable ofrece una capacidad de 1000mAh, lo que permite un funcionamiento prolongado del dispositivo.
- **Diodo de 2 amperios:** Componente utilizado para limitar la corriente en las conexiones eléctricas y garantizar un funcionamiento adecuado del dispositivo.
- **Resistencia de 330 ohmios:** Componente utilizado para limitar la corriente eléctrica en el circuito y proteger los componentes sensibles del dispositivo como leds.
- **Interruptor tipo botón:** Componente de entrada que permite al usuario encender y apagar el dispositivo con facilidad.
- **Pulsador:** Componente de entrada que permite al usuario interactuar con el dispositivo IoT, como iniciar una función específica o realizar una acción determinada, en particular la captura de imágenes.
- **Altavoz de 5W:** Componente de salida que emite sonidos y alertas necesarias, mejorando la interacción auditiva del dispositivo.

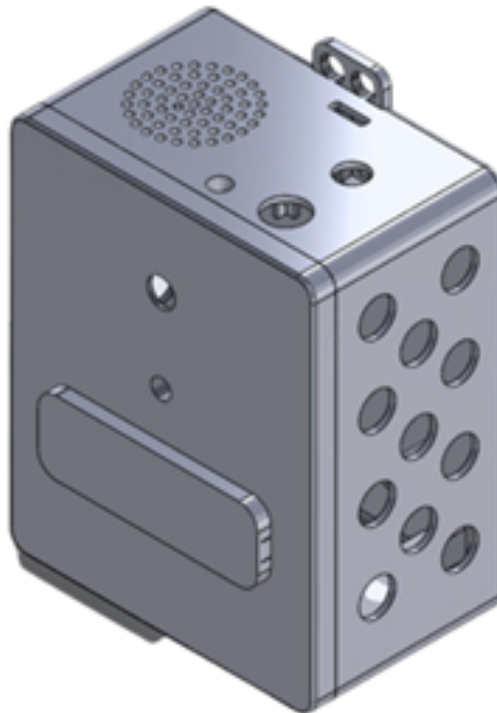
- **Potenciómetro lineal:** Componente que se utiliza para ajustar y controlar la intensidad de la señal de audio emitida por el altavoz.
- **Jumpers:** Conectores versátiles utilizados para establecer conexiones eléctricas entre los componentes.

## 4.2.2. Diseño e impresión en 3D

### 4.2.2.1. Diseño de la carcasa

La carcasa del dispositivo ha sido fabricada utilizando tecnología de impresión 3D, lo cual permite una construcción precisa y personalizada en un material manipulable (plástico). El diseño se ha concebido cuidadosamente para garantizar su comodidad y practicidad en su uso diario, siguiendo estándares de facilidad de uso y ergonomía (Ver 2.3.11). La imagen 4.7 muestra el diseño de la carcasa. Se utilizó un programa de Diseño Asistido por Computadora (CAD).

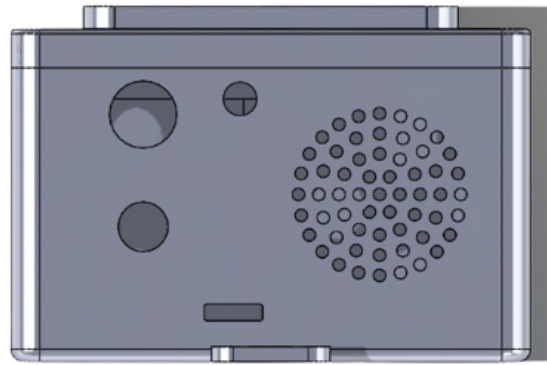
*Figura 4.7: Diseño de la carcasa del dispositivo.*



Para facilitar el transporte del dispositivo, se ha integrado un lanyard que se coloca alrededor del cuello del usuario, permitiendo llevar el dispositivo cómodamente en el pecho. En cuanto a la disposición de los elementos, la carcasa presenta una serie de características ergonómicas y funcionales. En su parte exterior, se han incluido orificios estratégicamente ubicados para el parlante, dirigidos hacia la cabeza del usuario para asegurar una óptima calidad acústica. En la parte superior, se encuentran las ranuras para el botón de encendido, el puerto de carga tipo C y un control manual de volumen del parlante, diseñados para una fácil accesibilidad y manejo (véase imagen 4.8).

En la parte frontal del dispositivo, se ha incorporado una ranura para un LED indicador del estado del equipo, un orificio para la cámara y un botón rectangular

*Figura 4.8: Disposición de orificios en la parte superior.*



de gran tamaño para el pulsador de activación de la cámara y el procesamiento de la imagen. Estos elementos están dispuestos de manera intuitiva para una interacción fluida y sencilla por parte del usuario (véase imagen4.9).

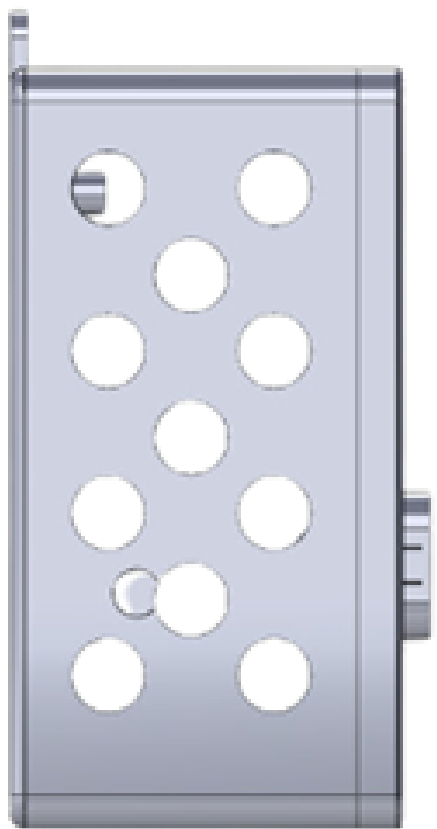
*Figura 4.9: Parte frontal del dispositivo.*



En los laterales, se han integrado orificios de ventilación para garantizar una adecuada disipación del calor generado por el Raspberry Pi 4B+, asegurando un óptimo funcionamiento del dispositivo en todo momento (véase imagen 4.10).

Finalmente, en la parte inferior, se ha diseñado un soporte ajustable para los componentes electrónicos, permitiendo una instalación precisa y segura mediante pernos y adhesivos, adaptándose así a las necesidades específicas de cada usuario. La carcasa del dispositivo está diseñada para una fácil instalación y mantenimiento del dispositivo, con la parte frontal unida a la parte trasera mediante un sistema de presión que garantiza una conexión segura y estable.

*Figura 4.10: Parte lateral del dispositivo.*



#### 4.2.2.2. Dispositivo final

La carcasa del dispositivo fue diseñada utilizando un programa de Diseño Asistido por Computadora (CAD) y posteriormente se exportó el archivo en formato STL. Seguidamente, este archivo se segmentó en el programa Ultimaker Cura en su versión 5.0.4 para proceder con la impresión 3D. Se utilizaron materiales de alta calidad, como PLA+ para la parte frontal y PETG para el botón y el cuerpo, garantizando así una carcasa resistente y duradera. A continuación, se muestra el prototipo impreso en 3D. La imagen 4.11 muestra el dispositivo con todos los componentes integrados y en funcionamiento.

*Figura 4.11: Dispositivo final en funcionamiento.*



### 4.3. Integración de Visión Artificial e IoT

Una vez completado el proceso de entrenamiento de los modelos, escoger el modelo con mayor precisión, y construir el dispositivo de IoT. Se procede con la integración de los mismos, es decir, unir todos componentes para que trabajen de manera conjunta.

Para ello, en primer lugar, se inicia con la comunicación entre el ordenador y el módulo Raspberry Pi 4B+. Se hace uso del programa WinSCP en su versión 6.2.3 RC para iniciar la comunicación vía SFTP (Secure File Transfer Protocol) <sup>1</sup>. La imagen 4.12 muestra la interfaz de WinSCP para iniciar una nueva conexión con el Raspberry Pi 4B+ <sup>2</sup>.

En segundo lugar y, una vez iniciada la comunicación. Se utiliza WinSCP para cargar los archivos de nuestro ordenador hacia el Raspberry Pi 4B+ (Véase imagen 4.13).

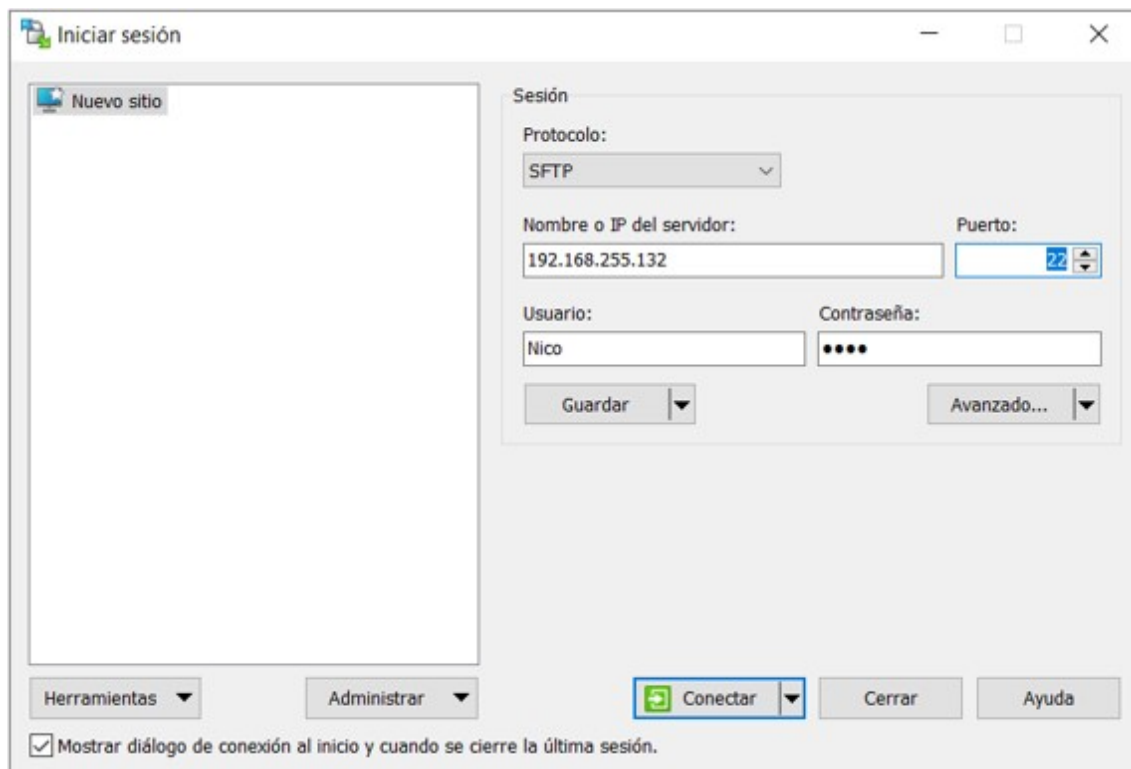
Finalmente, los archivos que se cargaron al Raspberry Pi 4B+ son los siguientes:

- **xception\_model.hdf5**: Almacena todos los pesos del modelo Xception, utilizado para el reconocimiento de billetes.
- **billete10soles.mp3**: Reproduce el audio “Billete de 10 soles.”. Indica que el billete capturado es de 10 soles.
- **billete20soles.mp3**: Reproduce el audio “Billete de 20 soles.”. Indica que el billete capturado es de 20 soles.

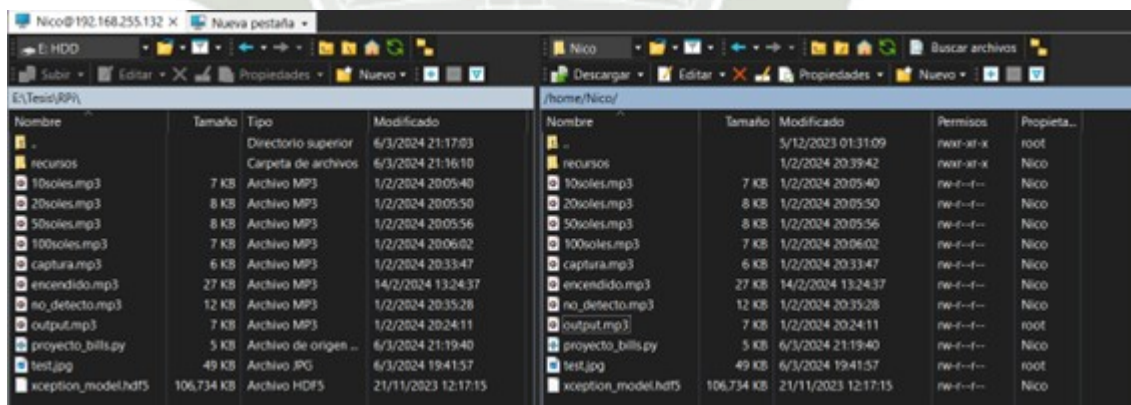
<sup>1</sup>Protocolo seguro utilizado para transferir archivos de manera remota entre sistemas. SFTP se basa en SSH (Secure Shell) para proporcionar un entorno seguro y cifrado para la transferencia de archivos.

<sup>2</sup>Para realizar la conexión con el Raspberry Pi 4B+ es necesario conocer la IP que se le asigna dentro de la red local, así como el Usuario y Contraseña establecidos en la configuración del mismo.

*Figura 4.12: Interfaz WinSCP para inicializar nueva conexión con Raspberry Pi 4B+.*



*Figura 4.13: Interfaz WinSCP para cargar archivos a Raspberry Pi 4B+ desde ordenador.*



- **billete50soles.mp3:** Reproduce el audio “Billete de 50 soles.”. Indica que el billete capturado es de 50 soles.
- **billete100soles.mp3:** Reproduce el audio “Billete de 100 soles.”. Indica que el billete capturado es de 100 soles.
- **pressbutton.mp3:** Reproduce el audio “Presione el botón.”. Indica que el usuario deberá presionar el botón de captura de imagen.
- **captura.mp3:** Reproduce el audio “Capturando imagen.”. Indica que se está

capturando imagen.

- **encendido.mp3**: Reproduce el audio “Dispositivo encendido.”. Indica que el dispositivo se ha encendido correctamente.
- **no\_detecto.mp3**: Reproduce el audio “No detectó nada, intente nuevamente.”. Indica que la imagen capturada no corresponde a un billete.
- **proyecto\_bills.py**: Contiene el código para el funcionamiento del dispositivo.

A continuación, se muestra extractos de código que se utilizan para realizar la predicción y reproducir los audios respectivamente. Para visualizar el código completo véase el apéndice 5.3.

### 1. Código para realizar la predicción

Se utiliza la función `capture_image(channel)` para capturar la imagen.

```
1 # Funcion para capturar la imagen cuando se presiona el boton
2 def capture_image(channel):
3     # Reproducir el audio de captura
4     os.system('mpg321 -o alsa captura.mp3')
5
6     # Inicializar camara y Capturar imagen
7     picam2.start()
8     picam2.capture_file("test.jpg")
9
10    #Realizar la prediccion en la imagen capturada
11    xception_preds_single = predict_image('test.jpg')
12
13    # Obtener la clase predicha (indice)
14    predicted_class_index = np.argmax(xception_preds_single,
15    axis=1)
16
17    # Verificar si la prediccion esta dentro del rango de
18    # clases validas
19    if predicted_class_index[0] in class_indices.values():
20        predicted_class_label = [key for key, value in
21        class_indices.items() if value == predicted_class_index
22        [0]][0]
23        # Reproducir el tipo de billete detectado
24        play_audio_mp3(predicted_class_label)
25        os.system('mpg321 -o alsa recursos/pressbutton.mp3')
26
27    return
```

Código 4.5: Función para capturar imagen

Donde:

- Recibe como parámetro *channel*, que indica el número del pin GPIO al que está asociado el evento que ha sido detectado, en este sentido, este parámetro hace referencia a que el pin del Raspberry Pi 4B+ asociado al botón ha sido activado para captura la imagen.
- Se inicia reproduciendo el audio de captura de imagen, se inicializa la cámara y se captura la imagen. La imagen capturada se almacena en un archivo temporal llamado “test.jpg”, el cual se almacena en la carpeta root del Raspberry Pi 4B+.

- Se realiza la predicción de la imagen llamando a la función *predict\_image(img\_path)* y enviando como parámetro la ruta de la imagen, en este caso “test.jpg”. La predicción se almacena en una variable denominada *xception\_preds\_single*.
- Finalmente se obtiene el índice de la clase predicha y se verifica si esta se encuentra dentro del rango de clases validas, en caso se encuentre dentro del rango, se hace un llamado a la función *play\_audio\_mp3(message)*, enviando como parámetro la etiqueta de la clase predicha y también se reproduce el audio que indica al usuario que presione nuevamente el botón en caso quiera capturar nuevamente la imagen.

También, se utiliza la función *predict\_image(img\_path)* para realizar la predicción de la imagen capturada.

```

1 # Funcion para realizar la prediccion en una imagen dada
2 def predict_image(img_path):
3     img = image.load_img(img_path, target_size=(299, 299))
4     img = image.img_to_array(img)
5     img = np.expand_dims(img, axis=0)
6     return xception_model.predict(img)

```

Código 4.6: Función para realizar la predicción

Donde:

- Se utiliza la función *image.load\_img()* del módulo **keras.preprocessing.image** para cargar la imagen desde la ruta que se recibe como parámetro. Además, la imagen se redimensiona a un tamaño de 299x299 píxeles.
- En seguida, la imagen se convierte en un array numpy utilizando la función *img\_to\_array()*. Esto convierte la imagen en una matriz numérica que puede ser procesada por el modelo.
- La función *expand\_dims()* se utiliza para expandir las dimensiones del array de la imagen. Esto se hace para que coincida con las dimensiones esperadas por el modelo, el cual requiere un tensor de entrada con una dimensión adicional para representar el lote de imágenes.
- Finalmente, se retorna la predicción del modelo, utilizando el método *predict()*, y enviando como parámetro la imagen preprocesada anteriormente.

## 2. Código para reproducir los audios

Se utiliza la función *play\_audio\_mp3(message)* para reproducir el audio que indica la denominación de la imagen capturada.

```

1 # Funcion para reproducir audio mp3
2 def play_audio_mp3(message):
3     if message == "billete10_anverso_antiguo" or "
4         billete10_anverso_nuevo"
5         or "billete10_reverso_antiguo" or "
6         billete10_reverso_nuevo":
7         os.system('mpg321 -o alsa 10soles.mp3')
8     if message == "billete20_anverso_antiguo" or "
9         billete20_anverso_nuevo"

```

```
7         or "billete20_reverso_antiguo" or "  
billete20_reverso_nuevo":  
8         os.system('mpg321 -o alsa 20soles.mp3')  
9         if message == "billete50_anverso_antiguo" or "  
billete50_anverso_nuevo"  
10            or "billete50_reverso_antiguo" or "  
billete50_reverso_nuevo":  
11            os.system('mpg321 -o alsa 50soles.mp3')  
12            if message == "billete100_anverso_antiguo" or "  
billete100_anverso_nuevo"  
13                or "billete100_reverso_antiguo" or "  
billete100_reverso_nuevo":  
14                os.system('mpg321 -o alsa 100soles.mp3')
```

Código 4.7: Función para reproducir audios

Donde:

- Recibe como parámetro la etiqueta de la clase predicha.
- Se realizan validaciones en base a esta etiqueta, de modo que imprime un determinado audio para cada tipo de billete (10, 20, 50 y 100 soles).
- Se utiliza el comando “os.system('mpg321 -o alsa {archivo\_audio}')” para reproducir el audio.

## 4.4. Pruebas y validación del dispositivo

Se realizaron pruebas del dispositivo con 44 personas con discapacidad visual de la Asociación Unión de Ciegos de Arequipa (Definición de la muestra en el capítulo 3.3.3), donde se aplicaron las pruebas estandarizadas de UEQ (User Experience Questionnaire) y SUS (System Usability Scale).

### 4.4.1. UEQ - Cuestionario sobre la experiencia de los Usuarios

De acuerdo con Schrepp et al. (2014), para crear productos o servicios exitosos, es necesario asegurarse de que el producto tenga una experiencia de usuario lo suficientemente alta. Un método eficiente y económico para llevar a cabo estas mediciones es el uso de cuestionarios estandarizados/validados.

En este sentido, la experiencia del usuario se ve como un concepto holístico que incluye todo tipo de reacciones emocionales, cognitivas o físicas relacionadas con el uso concreto o incluso solo el uso presumido de un producto. (Schrepp et al., 2014)

De acuerdo con Laugwitz et al. (2006) y Laugwitz et al. (2008) el objetivo principal del UEQ es permitir una medición rápida e inmediata de la experiencia del usuario. El UEQ considera aspectos de calidad pragmática y hedónica.

La versión original en alemán del UEQ fue creada en 2005 mediante un enfoque analítico de datos para garantizar la relevancia práctica de las escalas construidas, que corresponden a aspectos de calidad distintos. Se creó un conjunto inicial de elementos de 229 posibles ítems relacionados con la experiencia del usuario en sesiones de lluvia de ideas con expertos en usabilidad. Este conjunto de elementos se redujo

a una versión inicial de 80 ítems del cuestionario mediante una evaluación experta. (Schrepp et al., 2014)

El cuestionario de experiencia del usuario contiene 6 escalas con 26 ítems (Laugwitz et al., 2008) (Véase el apéndice 5.4 para visualizar el UEQ aplicado):

- **Atractivo:** Impresión general del producto. ¿Les gusta o no a los usuarios?  
**Ítems:** incómodo / cómodo, bueno / malo, desagradable / agradable, atractivo / feo, amigable / no amigable.
- **Perspicuidad:** ¿Es fácil familiarizarse con el producto?  
**Ítems:** no comprensible / comprensible, fácil de aprender / difícil de aprender, complicado / fácil, claro / confuso.
- **Eficiencia:** ¿Pueden los usuarios resolver sus tareas sin esfuerzo innecesario?  
**Ítems:** rápido / lento, ineficiente / eficiente, no pragmático / pragmático, ordenado / sobrecargado.
- **Confiabilidad:** ¿Siente el usuario que tiene control sobre la interacción?  
**Ítems:** impredecible / predecible, obstructivo / impulsor de apoyo, seguro / inseguro, cubre expectativas / no cubre expectativas.
- **Estimulación:** ¿Es emocionante y motivador usar el producto?  
**Ítems:** valioso / de poco valor, aburrido / emocionante, no interesante / interesante, activante / adormecedor.
- **Novedad:** ¿Es el producto innovador y creativo?  
**Ítems:** creativo / sin imaginación, original / convencional, creativo / sin imaginación, conservador / innovador.

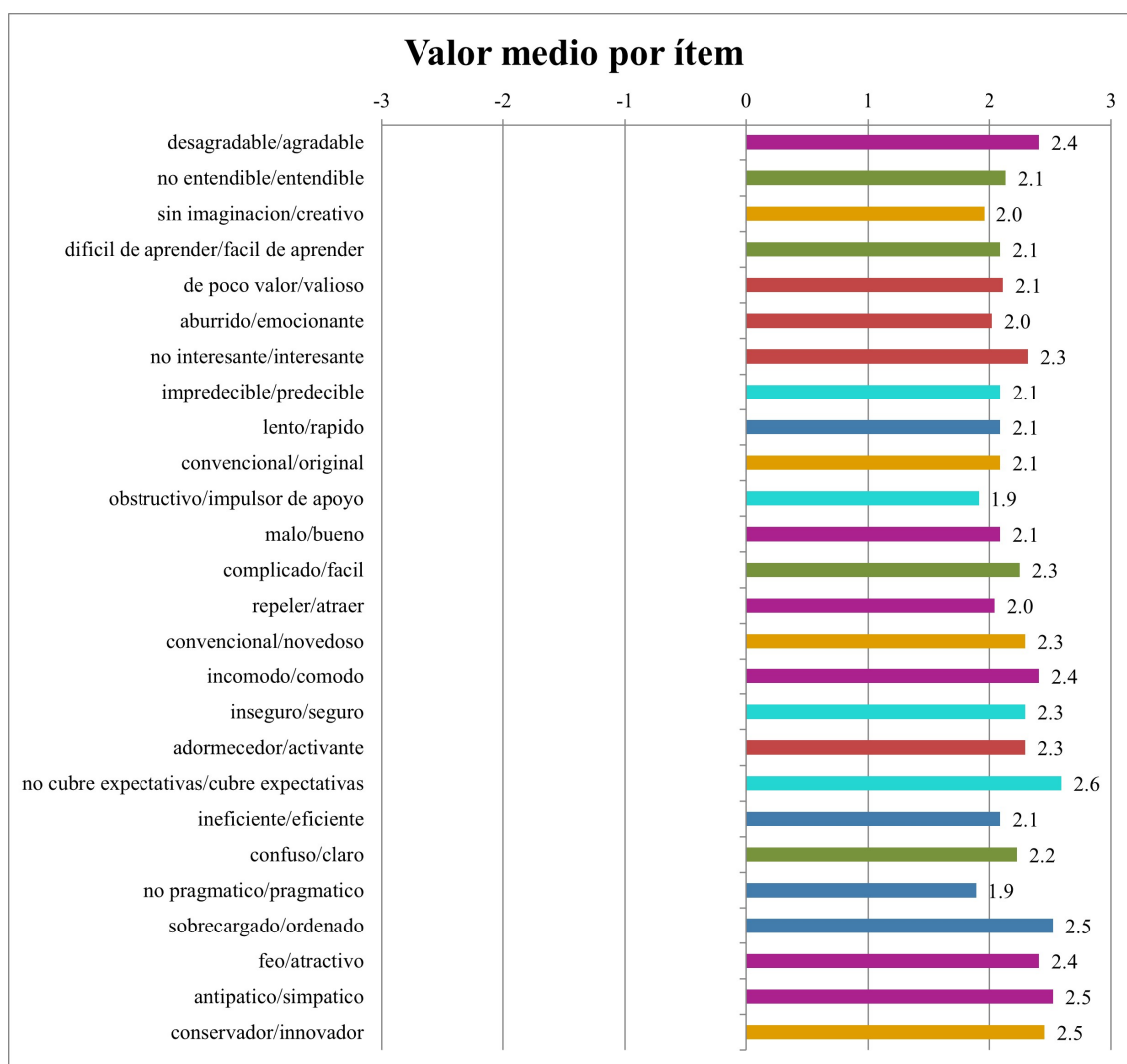
El tiempo estimado que se tomó por cada usuario para responder este cuestionario fue de 3-5 minutos (Schrepp et al., 2014). En el apéndice 5.4 se encuentra el enlace al repositorio para visualizar las encuestas aplicadas.

En el gráfico 4.14 se observan los resultados de la aplicación del test UEQ. El rango de las escalas oscila entre -3 (terriblemente malo) y 3 (extremadamente bueno). El gráfico muestra una tendencia positiva hacia el dispositivo por parte de las 44 personas encuestadas. Destacando 3 aspectos con puntuaciones superiores a la media: su simpatía con una puntuación de 2.5, novedad con 2.5 y las expectativas que cubre con 2.6. Estos resultados sugieren una percepción positiva y una alta valoración de estos atributos por parte de los encuestados. Por otro lado, se identifican áreas que reciben puntuaciones con un promedio de 1.9 o 2.0, ello indica que el dispositivo continúa cubriendo las expectativas de los usuarios, aunque se plantean posibles áreas de mejora.

La gráfica 4.15 indica que el dispositivo de asistencia está mayormente en línea con las expectativas de los usuarios, reflejando valoraciones positivas en las diferentes escalas, como se destacan el atractivo con un promedio de 2.3 y la transparencia, controlabilidad, estimulación y novedad con 2.2. No obstante, también se identifican áreas susceptibles de mejora, como la eficiencia, calificado con 2.1. Estos resultados sugieren todavía existen oportunidades para optimizar su rendimiento, especialmente en términos de eficiencia.

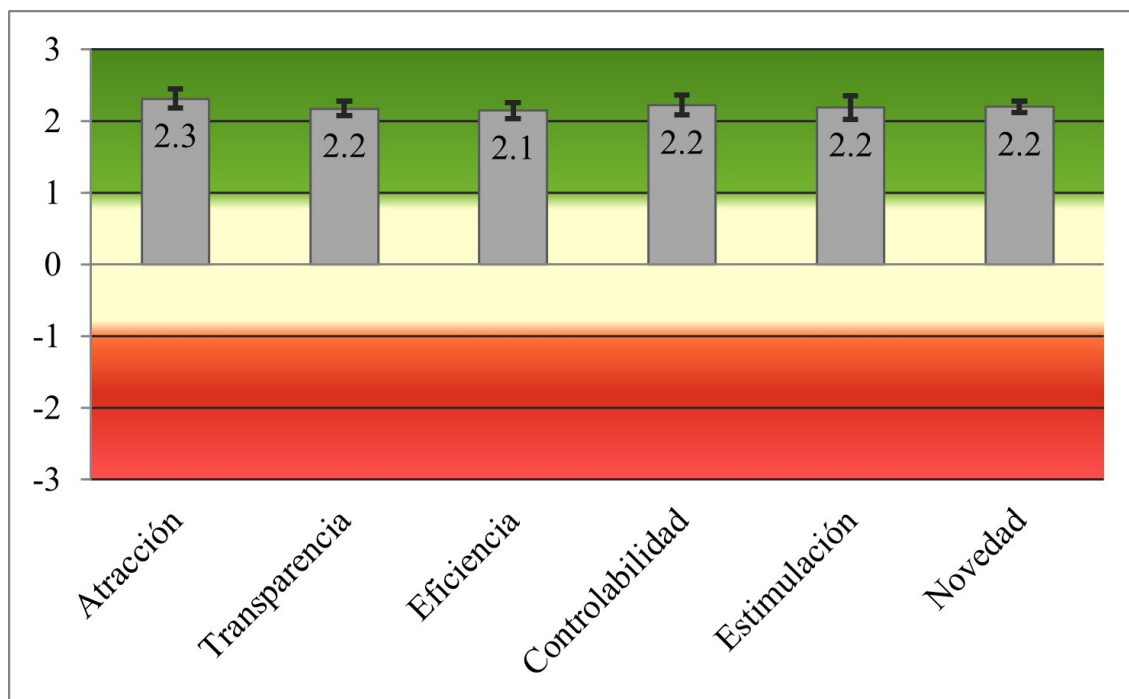
Una de las facilidades que ofrece UEQ, es que se puede realizar la comparación del producto con valores existentes de un conjunto de datos de referencia. Este

*Figura 4.14: Valor medio por ítem.*

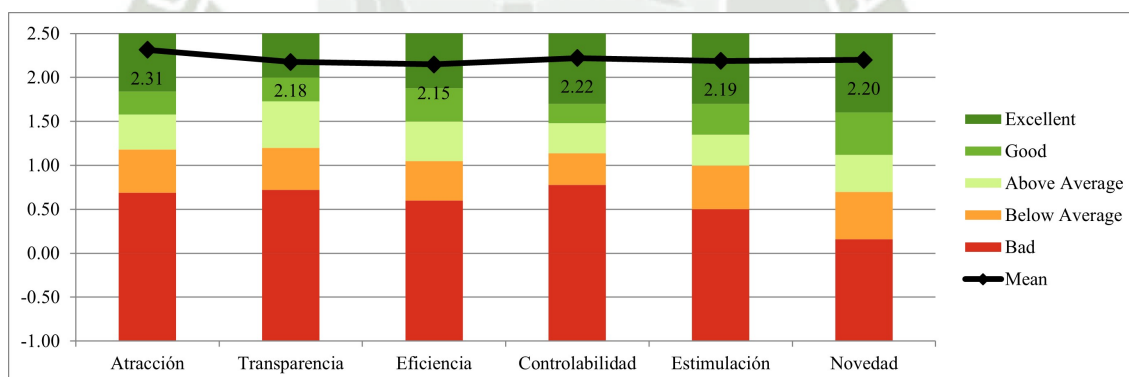


conjunto de datos contiene datos de 21175 personas procedentes de 468 estudios relativos a diferentes productos. La comparación de los resultados del producto evaluado con los datos del gráfico de referencia permite extraer conclusiones sobre la calidad relativa del producto evaluado en comparación con otros productos. (Hinderks et al., 2018) En este sentido, en el gráfico 4.16, se aprecia que, en comparación con otros productos, el dispositivo posee valoraciones positivas, llegando a cubrir las expectativas del usuario. Además, sugiere que podría ingresar fácilmente en un mercado más amplio.

**Figura 4.15:** Valoración por escala.



**Figura 4.16:** Comparación con otros productos.



#### 4.4.2. SUS - Escala de usabilidad del sistema

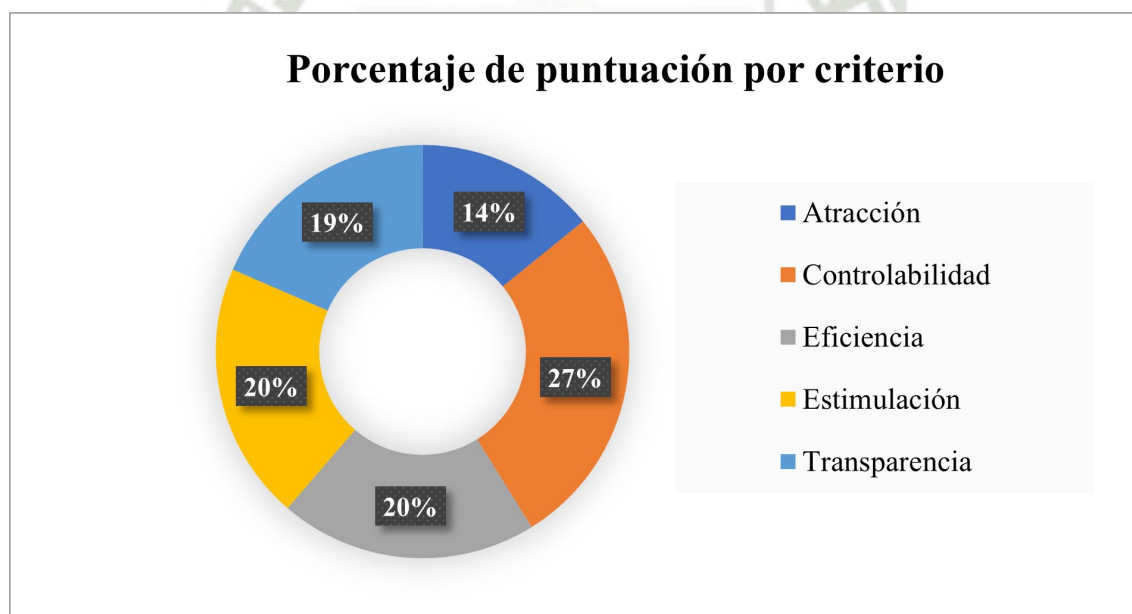
La Escala de Usabilidad del Sistema (SUS) propuesta por Brooke (1995) es una de las encuestas que se pueden utilizar para evaluar la usabilidad de una variedad de productos o servicios. De acuerdo con Bangor et al. (2009), hay varias características del SUS que hacen que su uso sea atractivo. Primero, está compuesta por solo diez afirmaciones, por lo que es relativamente rápido y fácil de completar para los participantes del estudio y para los administradores para calificar. Segundo, no es propietaria, por lo que es rentable de usar y se puede calificar muy rápidamente, inmediatamente después de completarse. Tercero, el SUS es tecnológicamente agnóstico, lo que significa que puede ser utilizado por un amplio grupo de profesionales de la usabilidad para evaluar casi cualquier tipo de interfaz de usuario, incluidos sitios web, teléfonos celulares, sistemas de respuesta interactiva de voz (IVR) (tanto

táctiles como de voz), aplicaciones de televisión y más. Por último, el resultado de la encuesta es una puntuación única, que varía de 0 a 100, y es relativamente fácil de entender por una amplia gama de personas de otras disciplinas.

El tiempo estimado que se tomó por cada usuario para responder este cuestionario fue de 2-4 minutos (Schrepp et al., 2014). En el apéndice 5.4 se encuentra el cuestionario SUS aplicado y el enlace al repositorio con las encuestas.

La gráfica 4.17 muestra una distribución porcentual de los criterios evaluados en la prueba de usabilidad, donde se identifica a la controlabilidad como el aspecto más resaltante de la prueba, con un 27%; seguido de la estimulación y la eficiencia con un 20%. Este resultado indica que el 67% de los participantes se sintieron capacitados para manejar el dispositivo con facilidad, lo que sugiere que la interfaz del dispositivo posee la intuitividad y accesibilidad como características inherentes. Por otro lado, se observa que el 14 y 19% de las personas encuestadas resaltaron la atracción y transparencia como aspectos con menor porcentaje. Este resultado indica la necesidad de poder enfocarse en aspectos como el diseño del dispositivo y la disposición de elementos que permitan aumentar tanto la comprensión y la claridad de la interfaz del dispositivo.

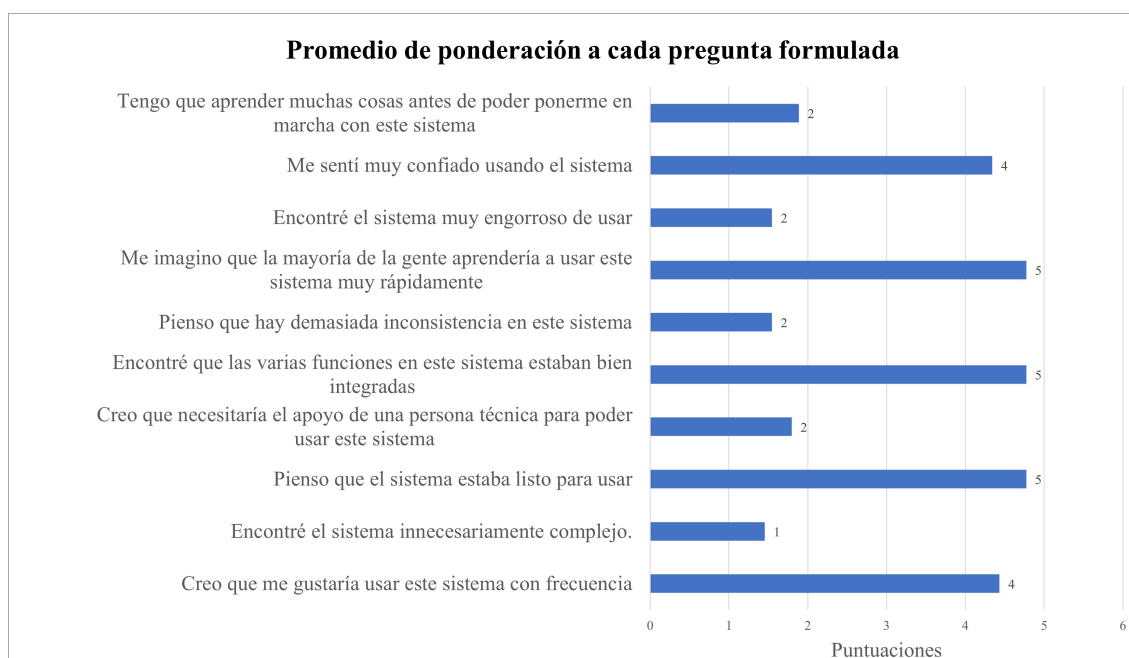
*Figura 4.17: Porcentaje de puntuación por criterio.*



De acuerdo con la gráfica 4.18, las respuestas reflejan una alta confianza en la capacidad de uso del sistema, con una puntuación promedio de 4, lo que indica una sensación de seguridad entre los usuarios al utilizarlo. Además, la percepción de que las funciones del sistema se encuentran bien integradas, con una puntuación promedio de 5, sugiere una experiencia satisfactoria y fluida durante la interacción con el dispositivo.

Por otro lado, la tendencia positiva también se ve respaldada por la afirmación de que la mayoría de las personas aprenderían rápidamente a usar el sistema, con una puntuación promedio de 5, ya que refuerza la idea de que el dispositivo es intuitivo y fácil de dominar para la mayoría de los usuarios. Estas evaluaciones favorables coinciden con la observación general de la gráfica, donde aspectos como

**Figura 4.18:** Promedio de ponderación a cada pregunta formulada.



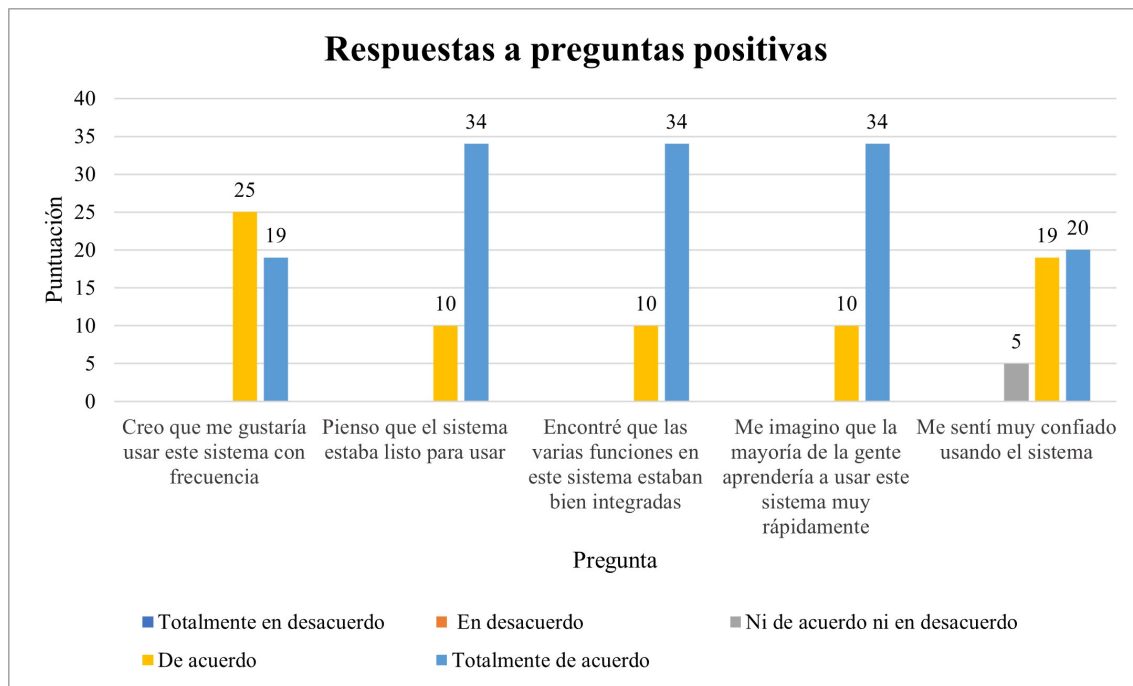
la controlabilidad y la facilidad de uso del dispositivo son destacados.

En línea con lo mencionado anteriormente, se procede a realizar un análisis por grupos de preguntas. En este sentido, las preguntas se han dividido en dos categorías: preguntas positivas y preguntas negativas. Las preguntas positivas buscan destacar los aspectos favorables del dispositivo, mientras que las preguntas negativas buscan identificar posibles áreas de mejora o debilidades del mismo.

Como se observa en la figura 4.19, se identificaron 5 preguntas como positivas, donde se observa que los usuarios generalmente se encuentran “Totalmente de acuerdo” y “De acuerdo” con las afirmaciones. Por ejemplo, en la primera afirmación “Creo que me gustaría usar este sistema con frecuencia”, 25 (57 %) usuarios se encuentran de acuerdo, mientras que 19 (43 %) usuarios se encuentran totalmente de acuerdo, lo cual sugiere que el dispositivo satisface las necesidades y expectativas de los usuarios en términos de funcionalidad, utilidad y accesibilidad. Asimismo, en la afirmación “Me sentí muy confiado usando el sistema”, 19 (43 %) usuarios se encuentran de acuerdo y 20 (46 %) usuarios se encuentran totalmente de acuerdo, estos resultados sugieren que no se requiere una preparación profunda o extensa para hacer uso del dispositivo de asistencia, ya que es muy intuitivo y amigable. 5 (11 %) usuarios no poseen una opinión clara al respecto.

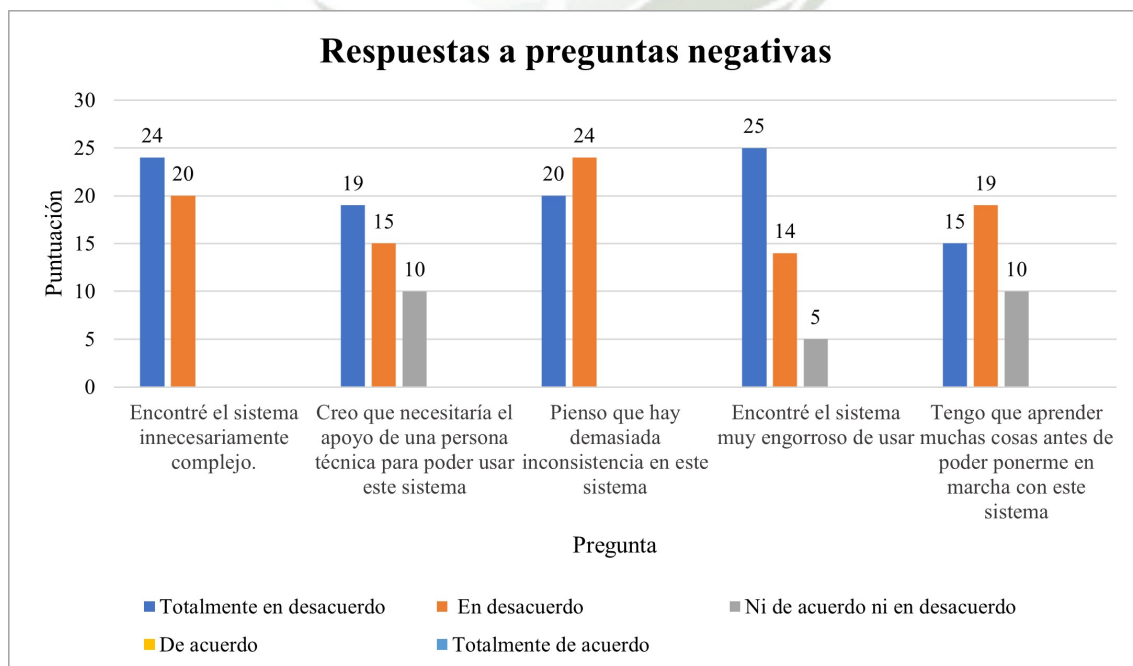
Por otro lado, en la figura 4.20 se observan las preguntas identificadas como negativas, se puede apreciar que los usuarios generalmente se encuentran “Totalmente en desacuerdo” y “Desacuerdo”, lo cual indica que los usuarios tienen una percepción positiva con respecto al uso del dispositivo. Por ejemplo en la afirmación “Encontré el sistema innecesariamente complejo”, 24 (55 %) usuarios se encuentran totalmente en desacuerdo, mientras que 20 (45 %) usuarios se encuentran en desacuerdo, lo cual sugiere que la mayoría de las personas invidentes no presentan dificultades al usar el dispositivo, sino por el contrario, lo consideran como accesible, cómodo y fácil de usar. Asimismo, en la afirmación “Tengo que aprender muchas cosas antes de poder

*Figura 4.19: Respuestas a preguntas positivas*



ponerme en marcha con este sistema”, 19 (43%) usuarios se encuentran en desacuerdo y 15 (34%) usuarios se encuentran totalmente en desacuerdo, estos resultados indican que los usuarios no requieren de una preparación profunda o extensa para hacer uso del dispositivo de asistencia. 10 (23%) usuarios no poseen una opinión clara al respecto.

*Figura 4.20: Respuestas a preguntas negativas*



En síntesis, se evidencia que las 44 personas invidentes encuestadas poseen una percepción positiva con respecto a la usabilidad del dispositivo de asistencia, percibiéndolo con características inherentes como la accesibilidad, facilidad en su uso, facilidad de aprender a utilizar, confiabilidad y la integración efectiva de sus funciones.

#### 4.4.3. Tiempo de procesamiento y respuesta del dispositivo

En las pruebas realizadas, el dispositivo demostró un buen rendimiento, con una duración de 4 horas en constante funcionamiento. Asimismo, el promedio de duración de la batería es de 8 horas con intervalos de tiempo en funcionamiento y apagado.

Por otro lado, el tiempo de encendido del dispositivo es de 35 segundos  $\pm$  3 segundos. El tiempo de procesamiento es de 6 segundos  $\pm$  1 segundo (3 segundos  $\pm$  1 segundo para reproducir los audios y 3 segundos  $\pm$  1 segundo para reproducir la denominación del billete).





## CAPITULO V

## ANÁLISIS Y VALIDACIÓN DE RESULTADOS

En el presente capítulo, se detallan los resultados obtenidos durante la investigación, acompañados de un análisis exhaustivo. Se inicia con la exposición de los métodos de validación empleados para evaluar la precisión de los algoritmos de aprendizaje profundo mediante los indicadores descritos en la tabla 1.1. Posteriormente, se aborda la integración del sistema, seguida de pruebas prácticas del funcionamiento del dispositivo en colaboración con miembros de la Asociación Unión de Ciegos de Arequipa.

### 5.1. Resultados

#### 5.1.1. Resultados en Visión Artificial

Como se detalla en el capítulo 4.1, se evaluaron cinco modelos de aprendizaje profundo preentrenados, entre los cuales Xception exhibió un rendimiento destacado, alcanzando una precisión superior tanto en Transfer learning (99.66 %) como en Fine-tuning (99.75 %), como se presenta en la tabla 4.4. En esta sección, se presentan los resultados específicos de las experimentaciones realizadas con este modelo.

##### 5.1.1.1. Tamaño del modelo (parámetros y memoria requerida)

Para obtener el tamaño del modelo (número de parámetros y la memoria requerida), ejecutamos el siguiente código en el entorno de trabajo: `xception_model.summary()`.

Como se muestra en la figura 5.1, se muestra una tabla que resume las capas del modelo, incluyendo el nombre de la capa, el tipo de capa, la forma de salida y el número de parámetros.

Al final del resumen, se muestra el tamaño total del modelo en términos de parámetros. El modelo tiene un total de 22,976,056 parámetros, lo que equivale a 87.65 megabytes (MB) en memoria. De estos, 2,114,576 son parámetros entrenables, y los restantes 20,861,480 son parámetros no entrenables (generalmente relacionados con BatchNormalization).

##### 5.1.1.2. Relación entre precisión y tiempo de predicción

De acuerdo con la tabla 4.4, Xception ofrece una alta precisión pero a expensas de un mayor tiempo de entrenamiento. (3h y 12min para Transfer Learning y 1h 50min para Fine-Tuning).

#### 5.1.2. Resultados en IoT

Como se detalla en el capítulo 4.2, se logró el desarrollo del dispositivo de asistencia para personas con discapacidad visual, para ello se siguieron los siguientes pasos:

##### 1. Diseño del dispositivo IoT:

Se comenzó con el diseño del dispositivo a nivel conceptual, es decir, recopilar los requerimientos de los usuarios, y en base a estos realizar las especificaciones en cuanto a los requerimientos funcionales y no funcionales del dispositivo.

*Figura 5.1: Tamaño del modelo (parámetros y memoria requerida).*

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 299, 299, 3)]	0	[]
block1_conv1 (Conv2D)	(None, 149, 149, 32)	864	['input_1[0][0]']
block1_conv1_bn (Batch Normalization)	(None, 149, 149, 32)	128	['block1_conv1[0][0]']
block1_conv1_act (Activation)	(None, 149, 149, 32)	0	['block1_conv1_bn[0][0]']
block1_conv2 (Conv2D)	(None, 147, 147, 64)	18432	['block1_conv1_act[0][0]']
block1_conv2_bn (Batch Normalization)	(None, 147, 147, 64)	256	['block1_conv2[0][0]']
block1_conv2_act (Activation)	(None, 147, 147, 64)	0	['block1_conv2_bn[0][0]']
block2_sepconv1 (Separable Conv2D)	(None, 147, 147, 128)	8768	['block1_conv2_act[0][0]']
...			
global_average_pooling2d (GlobalAveragePooling2D)	(None, 2048)	0	['block14_sepconv2_act[0][0]']
dense (Dense)	(None, 1024)	2098176	['global_average_pooling2d[0][0]']
dropout (Dropout)	(None, 1024)	0	['dense[0][0]']
dense_1 (Dense)	(None, 16)	16400	['dropout[0][0]']
=====			
Total params: 22976056 (87.65 MB)			
Trainable params: 2114576 (8.07 MB)			
Non-trainable params: 20861480 (79.58 MB)			

También, se creó el diagrama de flujo del funcionamiento del dispositivo de asistencia, como se muestra en la figura 4.6.

## 2. Diseño del circuito:

Se avanza con la etapa de diseño del circuito, iniciando las pruebas funcionales del dispositivo, las cuales se basan en los requerimientos tanto funcionales como no funcionales establecidos por los usuarios. Se considera cuidadosamente la selección de la placa electrónica de computadora y demás componentes, garantizando así que el funcionamiento del dispositivo cumpla con las expectativas previstas.

## 3. Diseño e impresión 3D:

Después de completar el diseño del circuito, se inicia el proceso de diseño de la carcasa del dispositivo, considerando los estándares de ergonomía y los requisitos específicos proporcionados por los usuarios de la Asociación Unión de Ciegos de Arequipa. Una vez finalizado el diseño, se procede a imprimir la carcasa utilizando materiales de alta calidad, como PLA+ para la parte

frontal y PETG para el botón y el cuerpo, lo que asegura una carcasa robusta y duradera. Finalmente, se lleva a cabo la unión del circuito con la carcasa impresa.

#### 4. Integración con Visión Artificial:

Por último, se realiza la integración de la visión artificial y el IoT. La imagen 4.11 ilustra el dispositivo final en pleno funcionamiento durante la tarea de reconocimiento de billetes.

## 5.2. Discusión

### 5.2.1. Principales hallazgos

#### 5.2.1.1. Software

1. **Usabilidad del Sistema:** Según el Cuestionario sobre la experiencia de los Usuarios (UEQ), el dispositivo de asistencia está mayormente en línea con las expectativas de los usuarios, reflejando valoraciones positivas en sus diferentes escalas, como se destaca el atractivo, transparencia, controlabilidad, estimulación y novedad. No obstante, también se identifican áreas susceptibles de mejora, como la eficiencia. Estos resultados sugieren que existen oportunidades de mejora para optimizar el rendimiento del dispositivo, especialmente en términos de eficiencia.

Asimismo, UEQ nos permitió comparar la propuesta con otros productos del mercado, donde, el dispositivo posee valoraciones positivas, llegando a cubrir las expectativas de los usuarios. Además, sugiere que podría ingresar fácilmente en un mercado más amplio.

Por otro lado, de acuerdo con la Escala de Usabilidad del Sistema (SUS), el software obtuvo una puntuación promedio del 80%, lo que indica una buena usabilidad general. Los resultados indican que las 44 personas invidentes encuestadas poseen una percepción positiva con respecto a la usabilidad del dispositivo de asistencia, denotándolo con características inherentes como la accesibilidad, facilidad en su uso, facilidad de aprender a utilizar, confiabilidad y la integración efectiva de sus funciones.

2. **Eficiencia Operativa:** La eficiencia operativa del dispositivo en términos de reconocimiento de billetes es notable, alcanzando un 99.7% de precisión. Además, se destaca el enfoque ergonómico y de bajo costo, lo cual posiciona al dispositivo como una solución innovadora y efectiva para mejorar la calidad de vida de las personas con discapacidad visual en la ciudad de Arequipa.

Asimismo, en términos de rendimiento (tiempo de procesamiento y respuesta del dispositivo), el dispositivo se mantuvo en funcionamiento constante durante 4 horas, destacando su rendimiento óptimo. Además, se registró un promedio de duración de la batería de 8 horas, considerando intervalos de tiempo entre su uso y periodos de inactividad.

3. **Fiabilidad y Estabilidad:** El software demostró una alta fiabilidad y estabilidad durante las pruebas, con una tasa mínima de errores y fallas del sistema. Los usuarios informaron una experiencia fluida y sin interrupciones durante su uso diario, lo que sugiere una sólida robustez del software.

#### 5.2.1.2. Hardware

1. **Diseño Ergonómico:** El dispositivo posee un diseño ergonómico que le permite colgar del cuello del usuario, lo que facilita su manipulación y reduce la fatiga durante un uso prolongado.

Los usuarios destacaron que la disposición de los botones y componentes del dispositivo es cómoda, lo cual resalta su accesibilidad y facilidad de uso.

2. **Durabilidad y Resistencia:** Se observó una buena durabilidad y resistencia en el hardware, incluso bajo condiciones de uso intensivo.

Los materiales de construcción de calidad garantizan una vida útil prolongada del dispositivo, minimizando la necesidad de mantenimiento y reparaciones considerablemente.

3. **Selección de Componentes y Rendimiento:** Gracias a la buena selección de materiales, el dispositivo demostró tener un buen rendimiento, no solo en términos de rendimiento, sino que también en la tarea de reconocimiento de billetes, gracias al uso del Raspberry Pi 4B+ de 1GB de memoria RAM, se logró reducir considerablemente el tiempo de procesamiento de las imágenes, así como la reproducción de audios.

### 5.2.2. Comparación con trabajos previos

#### 5.2.2.1. Software

En el estudio realizado por Meshram et al. (2023), se evaluaron cinco modelos preentrenados: VGG16, VGG19, Xception, ResNet152V2 e InceptionV3, con el objetivo de clasificar billetes tailandeses (THB). Los resultados revelaron que VGG16 obtuvo la mayor precisión, alcanzando un 87.50 %. La evaluación se llevó a cabo utilizando un conjunto de 3200 imágenes de billetes. Para cada modelo, se emplearon las técnicas de transferencia de aprendizaje y ajuste fino para mejorar su rendimiento.

En Sufri et al. (2019), se propone un sistema de reconocimiento automático de billetes malayos (MYR) utilizando técnicas de aprendizaje automático y aprendizaje profundo. Se realizaron dos experimentos utilizando imágenes de billetes en diferentes regiones y orientaciones. En el primer experimento, se extrajeron características de color de las imágenes y se utilizaron algoritmos de clasificación de aprendizaje automático para reconocer los billetes. Los resultados mostraron que los algoritmos de aprendizaje automático lograron una precisión del 99,7 %. En el segundo experimento, se utilizó una red neuronal convolucional preentrenada AlexNet para el reconocimiento de imágenes en diferentes orientaciones. Los resultados mostraron que la red neuronal convolucional tuvo un rendimiento mejor en las imágenes con la misma orientación, logrando una precisión del 100 %. Esta investigación también resalta la importancia de la aplicación de modelos preentrenados en la clasificación de billetes.

Del mismo modo, la investigación realizada por Kulkarni et al. (2023) presenta similitudes directas con nuestra propuesta, ya que aplican ambas técnicas, de transferencia de aprendizaje y ajuste fino para la clasificación de billetes. En el estudio se evaluaron 4 modelos preentrenados: ResNet, MobileNet, EfficientNet e InceptionNet, con el objetivo de clasificar billetes Indios (Rs). De estos modelos, ResNet superó y logró una alta precisión del 95.00 %, destacándose entre todos los demás modelos. Finalmente, los autores indican que el aprendizaje por transferencia mejora significativamente la precisión y eficiencia de los modelos de clasificación de billetes. Los modelos ajustados finamente superaron consistentemente a los modelos entrenados desde cero, resaltando la importancia de aprovechar el conocimiento preexistente.

Nuestra propuesta se destaca principalmente en comparación con estas investigaciones, ya que en el contexto peruano no se contaba con conjuntos de datos que incluyeran imágenes de billetes nacionales. En respuesta a esta carencia, se creó un nuevo conjunto de imágenes de billetes peruanos debidamente anotadas (Cayturo-Silva et al., 2023). Se aplicaron modelos de aprendizaje profundo preentrenados, tales como VGG16, Xception, ResNet50 y MobileNetV2, sobre este conjunto de datos, logrando alcanzar una precisión del 99.75 %. Este logro representa un punto de partida significativo para futuras investigaciones en esta área.

#### 5.2.2.2. Hardware

La investigación Duarte y Rojas (2021), se centró en el desarrollo de un dispositivo lector de billetes colombianos para personas con discapacidad visual, mediante un sensor de color y Arduino Nano como componentes principales del hardware. Esta elección resalta un enfoque cuidadoso en la accesibilidad y el costo, al mismo tiempo manteniendo la efectividad en la identificación de billetes colombianos. El sensor de color se utilizó para capturar las características visuales distintivas de los billetes, mientras que Arduino Nano proporcionó la capacidad de procesamiento y control. Los resultados de las pruebas mostraron que el dispositivo logró una identificación precisa de las denominaciones de billetes colombianos, además brindando a los usuarios una experiencia cómoda y práctica.

En contraste, la investigación realizada por Solano (2018) optó por utilizar Raspberry Pi como el componente principal para implementar un dispositivo portátil de identificación de billetes de dólares. Esta elección de hardware proporcionó un mayor poder de procesamiento y versatilidad en la identificación de billetes de diferentes denominaciones. El dispositivo desarrollado logró identificar billetes de dólares en ambas caras y en diferentes condiciones. Además, que consiguió un tiempo de identificación de 2,41 minutos a una distancia de 16 a 29 cm, de esta manera resaltando la eficiencia y practicidad del dispositivo en entornos reales.

Por último, la investigación realizada por Tipán (2019) se centro en combinar tecnologías avanzadas como visión artificial, tecnología IoT y Raspberry Pi para implementar un dispositivo identificador de billetes orientado a personas no videntes. Esta combinación de hardware y tecnologías permitió desarrollar un dispositivo efectivo y portable, adaptado para mejorar la accesibilidad y funcionalidad del dispositivo en situaciones cotidianas. La integración de visión artificial facilitó la identificación precisa de billetes, mientras que la tecnología IoT como el Raspberry Pi, proporcionó accesibilidad en la web, facilidad de uso, costo económico y capacidad para aplicaciones de visión artificial, que requieren procesamiento de vídeo, memoria RAM externa y almacenamiento de datos. Dentro de los resultados, se demuestra el

dispositivo como eficiente, versátil y adaptable a diferentes escenarios de uso para usuarios no videntes.

En síntesis, cada investigación se distinguió por la selección y uso estratégico de diferentes componentes de hardware, lo que resultó en dispositivos efectivos y adaptados a las necesidades específicas de usuarios con discapacidad visual. Resaltando que la integración de tecnología IoT, visión artificial y sensores, permiten un mejor resultado en cuanto a costo, eficiencia y practicidad en su uso.

### 5.2.3. Fortalezas y limitaciones de la propuesta

#### 5.2.3.1. Software

##### 1. Fortalezas

- **Flexibilidad y adaptabilidad:** El software del dispositivo se destacó por su capacidad para adaptarse a diferentes entornos y necesidades de los usuarios en la detección de billetes. Esta flexibilidad permitió su implementación en una variedad de situaciones, desde entornos domésticos hasta comerciales.
- **Actualizaciones y Mejoras Continuas:** La arquitectura del software fue diseñada de manera modular, lo cual facilitó la implementación de actualizaciones y mejoras de forma rápida y eficiente. Esto garantiza que el dispositivo pueda mantenerse al día con las últimas tendencias tecnológicas y las necesidades cambiantes de los usuarios.
- **Interfaz Intuitiva:** A pesar de las limitaciones identificadas, la interfaz de usuario del software fue reconocida por su simplicidad y facilidad de uso. Los usuarios encontraron que la navegación era intuitiva, lo que contribuyó a una experiencia general positiva.
- **Independencia:** El dispositivo funciona sin necesidad de conexión a internet, lo que amplía su utilidad, garantiza la privacidad de los datos, mejora la velocidad y eficiencia, y prolonga la duración de la batería.

##### 2. Limitaciones

- **Interfaz de Usuario Limitada:** Aunque la interfaz de usuario es intuitiva, se observaron algunas limitaciones en cuanto a la personalización y la capacidad de adaptarse a las preferencias individuales de los usuarios. Esto podría afectar la experiencia del usuario, especialmente para aquellos con necesidades específicas o habilidades técnicas limitadas.
- **Compatibilidad Limitada:** Dado que se trata de un software desarrollado a medida, el sistema está restringido a las funciones específicas para las que fue diseñado, lo que podría restringir su accesibilidad para algunos usuarios y limitar su adopción en ciertos entornos.

### 5.2.3.2. Hardware

#### 1. Fortalezas

- **Durabilidad y Resistencia:** Durante las pruebas, el hardware del dispositivo demostró una notable durabilidad y resistencia. Esto sugiere una larga vida útil y un rendimiento consistente en diversas condiciones de uso, lo que aumenta la confiabilidad del dispositivo.

Además, dado que el hardware fue fabricado mediante impresión 3D, en caso de que el dispositivo sufra algún daño o fractura, resulta fácilmente reemplazable.

- **Eficiencia Energética:** Se observó una eficiencia energética destacada en el hardware, con un bajo consumo de batería y la capacidad de funcionar durante periodos prolongados sin necesidad de recarga. Esto es especialmente importante en aplicaciones donde la disponibilidad de energía puede ser limitada.

#### 2. Limitaciones

- **Tamaño y Portabilidad:** Aunque el hardware es robusto y resistente, su tamaño y peso pueden representar una limitación en términos de portabilidad. Esto puede dificultar su transporte y manejo en ciertos escenarios de uso, especialmente aquellos que requieran movilidad constante o espacio limitado.
- **Nuevas Funcionalidades:** Aunque el diseño del dispositivo es versátil, la incorporación de nuevas funcionalidades puede requerir modificaciones tanto en el diseño en sí como en la estructura de la carcasa en algunos casos.

### 5.2.4. Análisis Costo/Beneficio Del Proyecto

Como se detalla en el apéndice 5.2 la inversión inicial para la construcción del dispositivo de asistencia ergonómico y de bajo costo para personas invidentes es de S/269,95. De esta manera, se realizó el cálculo tanto de VAN y TIR del proyecto considerando una tasa de 0,081, que se presentaran a continuación:

**Tabla 5.1**

*Valor Actual Neto (VAN)*

Periodo	Flujo de Dinero (S/)
0	269.95
1	3,200.00
2	3,200.00
3	3,200.00
4	3,200.00
5	3,200.00
<b>VAN</b>	<b>S/13,013.02</b>

**Tabla 5.2**

*Tasa Interna de Retorno (TIR)*

Periodo	Flujo de Dinero (S/)
0	-269.95
1	3,200.00
2	3,200.00
3	3,200.00
4	3,200.00
5	3,200.00
<b>TIR</b>	<b>11.85</b>

La VAN de S/13,013.02 indica que el proyecto es viable económicamente, ya que el retorno esperado es mayor al costo inicial de la inversión. Además, una TIR de 11.85 % indica que el proyecto es altamente rentable, superando la tasa de descuento del 8.10 %.

Para el cálculo del precio de venta del dispositivo se consideró tener como ganancia deseada de S/. 50.00 soles y un margen de ganancia de 10.02 % soles por cada producto, para ello se realizó los siguientes cálculos:

$$\text{PreciodeVenta} = \text{Costodeproducción} + \text{Gananciadeseada} \quad (5.1)$$

$$\text{PreciodeVenta} = 269,95 + 50 = 319,95 = 320 \quad (5.2)$$

$$\text{Margendeganancia} = \left( \frac{(\text{PreciodeVenta} - \text{Costodeproducción})}{\text{PreciodeVenta}} \right) \times 100 \quad (5.3)$$

$$\text{Margendeganancia} = \left( \frac{(320 - 269,95)}{320} \right) \times 100 = (0,100166) \times 100 = 10,02 \% \quad (5.4)$$

El margen de ganancia de 10.02 % indica que se tiene una buena competitividad en el mercado y una capacidad positiva para generar beneficios a largo plazo, respecto al costo de producción del dispositivo. El costo unitario de S/ 320.00 del dispositivo, evidencia que es totalmente coherentes, ya que permite cubrir con los costos de producción.

En comparación a otros dispositivos del mercado, la propuesta presentada se destaca por su enfoque ergonómico y de bajo costo, así como por la integración de tecnologías de visión artificial con IoT, que contribuyeron a satisfacer las necesidades de las personas con discapacidad visual de la Asociación Unión de Arequipa. Las características como la facilidad de uso, transparencia, controlabilidad, estimulación y novedad, lo diferencian significativamente de los productos existentes en el mercado y lo posicionan como una solución innovadora y eficiente para abordar la tarea de reconocimiento de billetes. La tabla 5.3 muestra la comparativa de dispositivos similares en el mercado a la propuesta.

**Tabla 5.3**

*Comparación con otros productos del mercado.*

Dispositivo	Componentes	Costo de Producción (S/)
Dispositivo de asistencia propuesto	Placa electrónica y Drivers, Cámara Raspberry Pi de 5MP, Batería de Litio de 3.7V y 1000mA	S/ 269.95
Tarjeta escalonada de billetes (González, 2013)	Material de plástico	S/ 1.01 (aprox.)
Tablilla de billetes (Banxico, 2024)	Material de plástico	S/1.12 (aprox.)
Dispositivo lector de billetes colombianos (Duarte y Rojas, 2021)	Modulo sensor color RGB TCS34725, Tarjeta Arduino nano, Micro-motor vibrador 1027, Carcasa del dispositivo, Código fuente, Fuente de alimentación	S/ 733.09
Dispositivo portátil para la identificación de billetes (Solano, 2018)	Amplificador de audio, Altavoz, Pi Camera v2, Pulsador, Raspberry pi 3B	S/ 5,884.16
Dispositivo Orcam MyEye (Infobae, 2017)	Cámara inteligente, Leds, Deslizador, Mini altavoz HD incorporado, Puerto de UDB, Botón de encendido, Imanes incrustados	S/ 7,957.34 (aprox.)

Como se observa en la tabla 5.3 el dispositivo de asistencia se diferencia de otros productos en el mercado por su enfoque ergonómico, su tecnología avanzada y el bajo costo que posee, lo que lo posiciona como una opción atractiva y deseable para las personas invidentes de la Asociación Unión de Ciegos de Arequipa. Además, el costo de producción de S/269.95 que posee va de acuerdo con los beneficios que proporciona. Por otro lado, el precio de venta de S/320.00 genera una ganancia significativa de 50 soles por unidad, demostrando así viabilidad económica y su potencial para generar beneficios a largo plazo. En conclusión, se demuestra que el dispositivo de asistencia es una solución rentable que ofrece beneficios tangibles tanto para los usuarios y posibles inversores. Su capacidad para satisfacer una necesidad importante en la comunidad como la detección de billetes y su enfoque en la innovación, accesibilidad e inclusión financiera para las personas con discapacidad visual, lo convierten en una inversión con un impacto positivo en la sociedad.

## Conclusiones

- PRIMERO. Fue posible realizar la construcción de un dispositivo de asistencia ergonómico y de bajo costo para el reconocimiento de billetes para personas con discapacidad visual, gracias a la integración de técnicas de visión artificial con la tecnología de IoT.
- SEGUNDO. Se realizó el análisis de la situación actual de las personas con discapacidad visual, teniendo como resultado una mejor comprensión de sus necesidades y dificultades, por ejemplo, en tareas como el reconocimiento de billetes, movilidad y desplazamiento seguro. Por otro lado, se realizó un análisis exhaustivo de las herramientas de asistencia actuales, de esta manera identificando las principales limitaciones de las propuestas planteadas dentro de la región de Arequipa, como la falta de ergonomía, altos costos y la detección de billetes en tiempo real.
- TERCERO. Se elaboró un conjunto de datos robusto y diverso de imágenes de billetes peruanos de 10, 20, 50 y 100 nuevos soles, los cuales se usaron como base y entrada para el entrenamiento y validación de los modelos de visión artificial que son Xception, ResNet50, MobileNetV2, VGG16 y VGG19, siendo este paso fundamental para garantizar la correcta identificación de la denominación de billetes peruanos en distintas condiciones ambientales y de iluminación.
- CUARTO. La implementación del módulo de visión artificial contribuye al correcto funcionamiento del dispositivo de asistencia, evidenciándose en las tareas de captura, procesamiento y salida de audio, las cuales fueron aprobadas después de realizar las pruebas de integración, donde se evaluó el paso a paso de su funcionamiento con todos los componentes, cámara de Raspberry Pi de 5MP, Adafruit Max98357 y una placa electrónica de computadora Raspberry Pi 4B+.
- QUINTO. Se realizó una integración exitosa del módulo de IoT con el modelo de visión artificial Xception, puesto que tiene un nivel de precisión del 99.75 % y un tiempo promedio de 4s para el reconocimiento de imágenes en tiempo real.
- SEXTO. Se llevó a cabo la validación del sistema a través de pruebas funcionales y no funcionales, incluyendo pruebas orientadas al usuario en diversas condiciones y entornos. Estas pruebas abarcaron situaciones como billetes arrugados y doblados, escenarios cerrados, abiertos, con presencia de ruido y sin ruido. Los resultados demostraron de manera concluyente que el sistema efectivamente ayuda a los usuarios invidentes en la tarea de reconocer las denominaciones de los billetes peruanos.

## Recomendaciones

- Se recomienda mejorar la infraestructura del dispositivo de asistencia considerando aspectos ergonómicos como peso, tamaño y volumen, además se debería de explorar otros componentes IoT de Arduino y Raspberry, que tengan mayor procesamiento y rendimiento.
- Se recomienda abordar otras carencias identificadas en el análisis de la situación actual de las personas con discapacidad visual y de herramientas de asistencia, como la detección de billetes falsos y reconocimiento de texto para la lectura de documentos físicos y/o digitales, siendo también necesidades que afectan a las personas con discapacidad visual. Así mismo, para abordar este desafío, se debería revisar la literatura respecto a arquitecturas de modelos de visión artificial para la detección de billetes falsos y reconocimiento de textos en tiempo real.
- Se sugiere utilizar modelos de visión artificial con mayores números de capas de profundidad, para aumentar la precisión en el reconocimiento de billetes, ya que la literatura respalda que mientras más capas tenga el modelo tendrá mayor capacidad de precisión y respuesta para el reconocimiento de imágenes.
- Para enriquecer el funcionamiento del dispositivo, se sugiere considerar la adición de nuevas funcionalidades, como el reconocimiento de objetos, señales de tránsito, colores y billetes falsos. Esta ampliación no solo diversificará las capacidades del dispositivo, sino que también se alinearán más estrechamente con las necesidades y expectativas de los usuarios invidentes. Además, la colaboración continua con la Asociación Unión de Ciegos de Arequipa será crucial para identificar estas necesidades específicas.
- Se sugiere mejorar la precisión del modelo de visión artificial, especialmente el modelo Xception, mediante la expansión y diversificación del conjunto de datos. La literatura respalda que un conjunto de datos más amplio y variado contribuirá a mejorar la precisión en el reconocimiento de billetes.
- Se recomienda explorar y realizar pruebas con nuevas placas electrónicas y dispositivos de captura de imágenes. La exploración deberá tener como objetivo mejorar la velocidad de procesamiento, la resolución de las imágenes y la capacidad de respuesta del dispositivo, ya que la incorporación de tecnologías más avanzadas llevará a mejoras significativas en el rendimiento general del sistema.
- Se recomienda centrarse en la experiencia del usuario, priorizando sus necesidades ergonómicas y de usabilidad. Esta mejora no solo optimizará la interacción del usuario con el dispositivo, sino que también proporcionará información valiosa sobre las preferencias y desafíos específicos que los usuarios pueden enfrentar al utilizar el prototipo.

## Trabajos Futuros

- Se explorarán nuevas funcionalidades para el dispositivo, como la capacidad de reconocer otros tipos de objetos además de billetes, como tarjetas de identificación, productos alimenticios, entre otros. Esto requerirá la expansión del software y la integración de algoritmos de reconocimiento adicionales.
- Mejorar la diversidad y cantidad de datos de entrenamiento es crucial para fortalecer el rendimiento del modelo de aprendizaje profundo. Además, se explorarán otros modelos más eficientes en términos de tamaño y recursos computacionales para optimizar el desempeño del sistema y su aplicabilidad en una variedad de contextos. Esto permitirá aprovechar el potencial del modelo en aplicaciones más amplias y garantizar su viabilidad en diferentes escenarios.
- Se realizarán ajustes en el diseño del dispositivo para mejorar su ergonomía y comodidad de uso para una variedad más amplia de usuarios, considerando factores como el tamaño, peso y disposición de los componentes.
- Se buscará mejorar la accesibilidad del dispositivo para personas con discapacidades adicionales, como la pérdida de audición, mediante la integración de opciones de entrada y salida alternativas, como comandos de voz y retroalimentación táctil.
- Se llevarán a cabo estudios para evaluar el impacto social y económico del dispositivo en la calidad de vida de las personas con discapacidad visual, así como su viabilidad en diferentes entornos y comunidades. Esto incluirá la recopilación de datos sobre la satisfacción del usuario, la eficacia del dispositivo y su adopción a largo plazo.

## Referencias

- Affif, M., Ayachi, R., Said, Y., y Atri, M. (2021). A transfer learning approach for indoor object identification. *SN Computer Science*, 2, 424. doi: 10.1007/s42979-021-00790-7
- Aguado, I. (2020). *Deep learning: Redes neuronales convolucionales en r* (Tesis de Master no publicada). Universidad de Sevilla.
- Avila, G. (2017). *Construcción de un bastón electrónico para las personas no videntes del cantón la maná* (Tesis de Master no publicada). Ingeniería en Informática y Sistemas Computacionales.
- Awad, S. R., Sharef, B. T., Salih, A. M., y Malallah, F. L. (2022). Deep learning-based iraqi banknotes classification system for blind people. *Eastern-European Journal of Enterprise Technologies*, 1(2 (115)), 31–38. doi: 10.15587/1729-4061.2022.248642
- Ayala, E. (2011). *Análisis, diseño y desarrollo de un sistema integral de soporte en la movilización y acceso interactivo a recursos de información en internet para personas no videntes, basado en tecnologías móviles* (Tesis de Master no publicada). Ingeniería Electrónico.
- Banco Central de Reserva del Perú. (2020). *Familia de billetes - nuevos billetes*. Descargado 18 August 2023, de <https://www.bcrp.gob.pe/billetes-y-monedas/familia-de-billetes.html>
- Bangor, A., Kortum, P., y Miller, J. (2009, 04). Determining what individual sus scores mean: Adding an adjective rating scale. *J. Usability Stud.*, 4, 114-123.
- Banxico. (2024). *Tablilla de billetes para personas con discapacidad visual*. Descargado de <https://www.banxico.org.mx/servicios/tablilla-billetes-discapacida.html> (Recuperado de <https://www.banxico.org.mx/servicios/tablilla-billetes-discapacida.html>)
- Basterrechea, M., y et al. (2011). Discapacidad visual y autonomía personal. En *Organización nacional de ciegos españoles* (p. 1-900).
- Beheshti, M., Naeimi, T., Hudson, T. E., Feng, C., Mongkolwat, P., Riewpaiboon, W., ... Rizzo, J.-R. (2023). A smart service system for spatial intelligence and onboard navigation for individuals with visual impairment (vis4ion thailand): study protocol of a randomized controlled trial of visually impaired students at the ratchasuda college, thailand [Article]. *Trials*, 24(1). Descargado de <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85149999175&doi=10.1186%2fs13063-023-07173-8&partnerID=40&md5=dad225d95f8a5be0bc14b077c156986c> doi: 10.1186/s13063-023-07173-8
- Bernal, C. A. (2010). *Metodología de la investigación científica: administración, economía, humanidades y ciencias sociales*. PEARSON EDUCACIÓN.

- Brooke, J. (1995, 11). Sus: A quick and dirty usability scale. *Usability Eval. Ind.*, 189.
- Castelar, S., Banegas, L. A., Mendoza, D. A., Soto, J. C., y Davila, K. (2022). Automated honduran banknote image classification using machine learning. En *2022 IEEE 40th Central America and Panama Convention (CONCAPAN)* (p. 1-6). doi: 10.1109/CONCAPAN48024.2022.9997703
- Cayturo Silva, N. E., Peña Alejandro, J. M., y Castro Gutierrez, E. G. (2023a). *Dataset of peruvian banknotes*. IEEE Dataport. Descargado de <https://dx.doi.org/10.21227/1raa-mz71> doi: 10.21227/1raa-mz71
- Cayturo Silva, N. E., Peña Alejandro, J. M., y Castro Gutierrez, E. G. (2023b). *Peruvian banknotes*. Mendeley Data. Descargado de <https://data.mendeley.com/datasets/22jmt8xhpn/1> doi: 10.17632/22jmt8xhpn.1
- Cayturo-Silva, N. E., Peña-Alejandro, J. M., Castro-Gutierrez, E. G., Sulla-Torres, J., y Maraza-Quispe, B. (2023). Annotated peruvian banknote dataset for currency recognition and classification. *Data in Brief*, 51, 109715. Descargado de <https://www.sciencedirect.com/science/article/pii/S2352340923007862> doi: <https://doi.org/10.1016/j.dib.2023.109715>
- Chollet, F. (2017). *Xception: Deep learning with depthwise separable convolutions*.
- Chávez, A. L. (2022, Oct). *En arequipa habría más de 30 mil personas con discapacidad visual*. Descargado de <http://radiosanmartin.pe/2022/10/14/en-arequipa-habria-mas-de-30-mil-personas-con-discapacidad-visual/>
- Cocconcelli, F., Matrella, G., Mora, N., Casu, I., Vargas Godoy, D. A., y Ciampolini, P. (2023). Iot smart flooring supporting active and healthy lifestyles. *Sensors*, 23(6). doi: 10.3390/s23063162
- Croce, D., Giarré, L., Pascucci, F., Tinnirello, I., Galioto, G. E., Garlisi, D., y Lo Valvo, A. (2019). An indoor and outdoor navigation system for visually impaired people. *IEEE Access*, 7, 170406-170418. doi: 10.1109/ACCESS.2019.2955046
- Das, U., Namboodiri, V., y He, H. (2021). Pathlookup: A deep learning-based framework to assist visually impaired in outdoor wayfinding. En *2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other affiliated events (PerCom Workshops)* (p. 111-116). doi: 10.1109/PerComWorkshops51409.2021.9431007
- De Frías, C. (2019). *Visión artificial aplicada en la identificación de objetos y su parametrización geométrica*. Departamento de Ingeniería de Sistemas y Automática, Grado en Ingeniería Electrónica Industrial y Automática, Universidad Carlos III de Madrid, Leganés, Madrid. (Trabajo Fin de Grado)

- Duarte, W., y Rojas, D. (2021). *Desarrollo de un dispositivo lector de billetes colombianos para la población con discapacidad visual* (Tesis Doctoral no publicada). Universidad Pedagógica Nacional.
- Dávila, B. (2016). *Vix - bastón inteligente para personas con discapacidad visual* (Tesis de Master no publicada). Diseñador Comunicacional.
- El-taher, F., Taha, A., Courtney, J., y Mckeever, S. (2021). A systematic review of urban navigation systems for visually impaired people. *Sensors*, 21(9). Descargado de <https://www.mdpi.com/1424-8220/21/9/3103> doi: 10.3390/s21093103
- Ergonomics of human-system interaction — part 210: Human-centred design for interactive systems [Manual de software informático]. (2019). (ISO 9241-210:2019)
- Ergonomics — general approach, principles and concepts [Manual de software informático]. (2011). (ISO 26800:2011)
- Ergonomía y discapacidad*. (2020). Instituto de Biomecánica de Valencia (IBV).
- Esparza Maldonado, A., Margain-Fuentes, L., Rodriguez, F., y Benitez-Guerrero, E. (2018). Desarrollo y evaluación de un sistema interactivo para personas con discapacidad visual. *TecnoLógicas*, 21. doi: 10.22430/22565337.733
- Fan, C.-B., Aseffa, D. T., Kalla, H., y Mishra, S. (2022). Ethiopian banknote recognition using convolutional neural network and its prototype development using embedded platform. *Journal of Sensors*, 2022, 4505089. doi: 10.1155/2022/4505089
- Fernandez, R. (2021). *Bastón sensorial geolocalizador inteligente para apoyar en el desplazamiento de personas invidentes en la organización regional de ciegos del Perú – chiclayo* (Tesis Doctoral no publicada). Universidad Católica Santo Toribio de Mogrovejo.
- Forsyth, D., y Ponce, J. (2012). *Computer vision: A modern approach*. Prentice Hall.
- Garnica, P., Gómez-Carreño, L. E., y Pérez-García, N. (2018). *Estimación del módulo de resiliencia usando redes neuronales*.
- Geerts, P., Eijsink, J., Moser, A., ter Horst, P., Boersma, C., y Postma, M. (2023). Rationale and development of an e-health application to deliver patient-centered care during treatment for recently diagnosed multiple myeloma patients: pilot study of the mm e-coach. *Pilot and Feasibility Studies*, 9(1).
- Godfrey, A., Hetherington, V., Shum, H., Bonato, P., Lovell, N. H., y Stuart, S. (2018). From a to z: Wearable technology explained. *Maturitas*, 113, 40-47. doi: 10.1016/j.maturitas.2018.04.012
- Gonzalez, R., y Woods, R. (2008). *Digital image processing*. Prentice Hall.

- González, A. (2013). 18.000 no videntes contarán con tarjeta para identificar billetes. *El País*. (Recuperado de <https://www.nacion.com/el-pais/18-000-no-videntes-contaran-con-tarjeta-para-identificar-billetes/ZZ7HW6BT4JCT3K5D3NUQ6YSRCU/story/>)
- Gunupudi, R. K., Achanta, S. R., D P, D. C., Ganga, H. R., Vipparthy, N., y Annaluri, S. R. (2021). Scene recognition and narration from video using deep learning techniques. En *The 7th international conference on engineering amp; mis 2021*. New York, NY, USA: Association for Computing Machinery. Descargado de <https://doi.org/10.1145/3492547.3492566> doi: 10.1145/3492547.3492566
- Gutiérrez, C. (2019). *Desarrollo de un prototipo electrónico que guié a personas invidentes en la recolección de objetos mediante audio y visión artificial* (Tesis Doctoral no publicada). Ingeniería Electrónica.
- Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and tensorflow*. O'Reilly Media, Inc.
- He, K., Zhang, X., Ren, S., y Sun, J. (2015). *Deep residual learning for image recognition*.
- Hernandez, C., Pedraza, L., y López, D. (2011, 10). Dispositivo tecnológico para la optimización del tiempo de aprendizaje del lenguaje braille en personas invidentes. , *13*, 1351-1358.
- Hernández, A. (2022). *Sistema de asistencia para personas con discapacidad visual basado en detección de objetos con deep learning* (Tesis de Master no publicada). Universidad de Chile.
- Hernández Sampieri, R., y Mendoza Torres, C. P. (2018). *Research methodology: The routes: Quantitative and qualitative and mixed*. Mc Graw Hill Education.
- Hersh, M., y Johnson, M. (2008). *Assistive technology for visually impaired and blind people* (1.<sup>a</sup> ed.). Springer.
- Hinderks, A., Schrepp, M., y Thomaschewski, J. (2018). *User experience questionnaire*. Descargado 10 March 2024, de <https://www.ueq-online.org/>
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... Adam, H. (2017). *Mobilenets: Efficient convolutional neural networks for mobile vision applications*.
- Huang, X., y Gai, S. (2020). Banknote classification based on convolutional neural network in quaternion wavelet domain. *IEEE Access*, *8*, 162141-162148. doi: 10.1109/ACCESS.2020.3021181
- Infobae. (2017). Orcam myeye, la novedosa cámara inteligente para ciegos. *INFOBAE*. (Publicado el 19 de junio de 2017)

- Instituto Nacional de Estadística e Informática. (2013). *En el Perú 1 millón 575 mil personas presentan algún tipo de discapacidad*. Descargado 29 March 2023, de <https://m.inei.gob.pe/prensa/noticias/en-el-peru-1-millon-575-mil-personas-presentan-alg/#:~:text=El%20Instituto%20Nacional%20de%20Estad%C3%ADstica%20e%20Inform%C3%A1tica%20inform%C3%B3%20que%20801,para%20ver%20incluso%20usando%20anteojos>
- Instituto Nacional de Estadística e Informática. (2014). *Primera encuesta nacional especializada sobre discapacidad 2012* (Inf. Téc.). Instituto Nacional de Estadística e Informática.
- Ipland, J., y Parra, C. (2009). La formación de ciegos y discapacitados visuales: visión histórica de un proceso de inclusión. , 1, 453-454.
- Jarraya, S., Al-Shehri, W., y Ali, M. (2020). Deep multi-layer perceptron-based obstacle classification method from partial visual information: Application to the assistance of visually impaired people. *IEEE Access*, 8, 26612-26622. doi: 10.1109/ACCESS.2020.2970979
- Keras. (2023). *Keras applications*. Descargado 17 October 2023, de <https://keras.io/api/applications/>
- Khalil, A., Yaghi, M., Basmaji, T., Faizal, M., Farhan, Z., Ali, A., y Ghazal, M. (2022). Mobile deep classification of UAE banknotes for the visually challenged. En *2022 9th international conference on future internet of things and cloud (fiCloud)* (p. 321-325). doi: 10.1109/FiCloud57274.2022.00053
- Ko Min-soo, D.-h. (2023). *Deep learning bible - 2. classification*. Descargado de <https://wikidocs.net/165427>
- Kulkarni, A., Shende, S., y Sasturkar, I. (2023). Banknote classification using transfer learning. *EPRA International Journal of Multidisciplinary Research (IJMR)*, 9(5), 382-388. doi: <https://doi.org/10.36713/epra2013>
- Laugwitz, B., Held, T., y Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. En A. Holzinger (Ed.), *Hci and usability for education and work* (pp. 63–76). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Laugwitz, B., Schrepp, M., y Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. En *Message Understanding Conference*. Descargado de <https://api.semanticscholar.org/CorpusID:35420403>
- Lemos, M. (2015). *Sistema de orientación para personas en situación de discapacidad visual (invidentes) en las estaciones tipo parada del sistema integrado m.i.o. (yo-guío)*. Universidad del Valle, Santiago de Cali. (Trabajo Fin de Grado)
- Lizárraga, C. (2018). *Propuesta para el diseño de un bastón electrónico para personas invidentes que mejorara la calidad de su desplazamiento diario* (Tesis de Master no publicada). Ingeniería Industrial.

- López, O., Sanabria, L., y Ibañez, J. (2006). Desarrollo de la cognición espacial en invidentes congénitos con apoyo de dispositivos tecnológicos. *LOM Ediciones 2006*, 2, 5-10.
- Matich, D. J. (2001). *Redes neuronales: Conceptos básicos y aplicaciones* (C. A. Ruiz y M. S. Basualdo, Eds.).
- Merchan, F., Poveda, M., Cáceres-Hernández, D., y Sanchez Galan F, J. (2021, 01). Indoor navigation aid systems for the blind and visually impaired based on depth sensors. , 187-223. doi: 10.4018/978-1-7998-6522-3.ch007
- Meshram, V., Patil, K., y Meshram, V. (2023). Evaluation of top pretrained models using transfer learning on banknote dataset with quality parameter. *Ingénierie des Systèmes d'Information*, 28(3), 693-701. doi: <https://doi.org/10.18280/isi.280319>
- Mettouris, C., Yeratziotis, A., Fixl, L., Parker, S., Starosta-Sztuczka, J., Clarke, V., ... Papadopoulos, G. A. (2023). esticky: An advanced remote reminder system for people with early dementia. *SN Computer Science*, 4(4). doi: 10.1007/s42979-023-01768-3
- Microsoft PowerToys. (2023). *Microsoft powertoys: utilidades para personalizar windows*. Descargado 06 September 2023, de <https://learn.microsoft.com/es-es/windows/powertoys/>
- Ministerio de Salud. (2014). *En el Perú cerca de 160,000 personas son invidentes y unas 600,000 sufren de alguna discapacidad visual*. Descargado 29 March 2023, de <https://www.gob.pe/institucion/minsa/noticias/31145-en-el-peru-cerca-de-160-000-personas-son-invidentes-y-unas-600-000-sufren-de-alguna-discapacidad-visual>
- Morparia, K., Kanabar, A., Bhadoriya, S., y Kadam, P. (2022). Face recognition with audio output: An aid for the visually impaired. En *2022 4th international conference on inventive research in computing applications (icirca)* (p. 1519-1524). doi: 10.1109/ICIRCA54612.2022.9985550
- Mouna Afif, R., y de Antonio, A. (2021, 11). A transfer learning approach for indoor object identification. *SN Computer Science*. doi: 10.1007/s42979-021-00790-7
- Mukhiddinov, M., Abdusalomov, A., y Cho, J. (2022, 04). Automatic fire detection and notification system based on improved yolov4 for the blind and visually impaired. *Sensors*, 22, 3307. doi: 10.3390/s22093307
- Ontiveros, S., Rojas, D., y Martínez, J. (2014). Diseño y construcción de una bastón blanco electrónico para personas invidentes. *Instituto Politécnico Nacional*, 18.
- Organización de las Naciones Unidas. (2014). *Convención sobre los derechos de las personas con discapacidad* (Inf. Téc.). Organización de las Naciones Unidas.

- Organización Mundial de la Salud. (2010). *Informe mundial sobre la visión*. Descargado de <https://apps.who.int/iris/bitstream/handle/10665/331423/9789240000346-spa.pdf>
- Organización Mundial de la Salud. (2018). *International statistical classification of diseases and related health problems (icd)*. Descargado 03 May 2023, de <https://www.who.int/classifications/classification-of-diseases>
- Organización Mundial de la Salud. (2019). *Ceguera y discapacidad visual*. Descargado 08 May 2023, de <https://www.who.int/es/news/item/08-10-2019-who-launches-first-world-report-on-vision>
- Organización Mundial de la Salud. (2022). *Ceguera y discapacidad visual*. Descargado 03 May 2023, de <https://www.who.int/es/news-room/fact-sheets/detail/blindness-and-visual-impairment>
- Organización Nacional de Ciegos Españoles. (2011). *Discapacidad visual y autonomía personal (1.ª ed.)* [Manual de software informático].
- Ortiz, L. (2012, 11). Sistema de identificación de objetos para personas invidentes usando la tecnología rfid. , 38-44.
- Pascolini, D., y Mariotti, S. P. (2012). Global estimates of visual impairment: 2010. *The British journal of ophthalmology*, 96(5), 614–618. doi: 10.1136/bjophthalmol-2011-300539
- Paz, M., y Sandoval, C. (2013). *Aporte en el uso de las tics, para las personas con discapacidad visual a través de la implementación del programa jaws y magnificador de pantalla en los cyber de la parroquia esmeraldas del cantón y provincia del mismo nombre, durante el año 2011; y evaluación de los resultados del proyecto piloto* (Tesis de Master no publicada). Educación Especial.
- Peralta, M., y Urmendiz, J. (2014). *Sistema de asistencia y guía para personas invidentes* (Tesis de Master no publicada). Ingeniería Biomédica.
- Piculo-Dos Santos, A. D., Grota-Suzuki, A. H., Orsi-Medola, F., y Vaezipour, A. (2021). A systematic review of wearable devices for orientation and mobility of adults with visual impairment and blindness. *IEEE Access*, 9, 162306-162324. doi: 10.1109/ACCESS.2021.3132887
- Portilla, G. (2022). *El csic desarrolla un perro-robot para guiar a personas dependientes o con discapacidad*.
- Prince, S. (2012). *Computer vision: Models, learning, and inference*. Cambridge University Press.
- Rajab Awad, S., Sharef, B., Mohammed, A., y Malallah, F. (2022, 02). Deep learning-based iraqi banknotes classification system for blind people. *Eastern-European Journal of Enterprise Technologies*, 1, 31-38. doi: 10.15587/1729-4061.2022.248642

- Rusli, S., Liawatimena, S., y Trisetyarso, A. (2022, 12). Context aware navigation system for assisting blind people. , *16*, 1351-1358. doi: 10.24507/icicel.16.12.1351
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., y Chen, L.-C. (2019). *Mobile-netv2: Inverted residuals and linear bottlenecks*.
- Schrepp, M., Hinderks, A., y Thomaschewski, J. (2014). Applying the user experience questionnaire (ueq) in different evaluation scenarios. En A. Marcus (Ed.), *Design, user experience, and usability. theories, methods, and tools for designing the user experience* (pp. 383–392). Cham: Springer International Publishing.
- Shah, H., Amin, M., Dadwani, K., Desai, N., y Chatiwala, A. (2023). Real-time object detection system with voice feedback for the blind people. En Y.-D. Zhang, T. Senjyu, C. So-In, y A. Joshi (Eds.), *Smart trends in computing and communications* (p. 683-690). Singapore: Springer Nature Singapore.
- Shahira, K. C., y Lijiya, A. (2021). Towards assisting the visually impaired: A review on techniques for decoding the visual data from chart images. *IEEE Access*, *9*, 52926-52943. doi: 10.1109/ACCESS.2021.3069205
- Simonyan, K., y Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition*.
- Solano, G. (2018). *Diseño y construcción de un dispositivo portátil para la identificación de billetes orientado a personas con discapacidad visual mediante el uso de visión artificial* (Tesis Doctoral no publicada). Escuela Superior Politécnica de Chimborazo.
- Soderback, I. (2015). *International handbook of occupational therapy interventions*. Springer. Descargado de <https://worldcat.org/title/903047936>
- Strategic Policy Forum on Digital Entrepreneurship. (2016). *A digital compass for decision makers: toolkit on disruptive technologies, impact and areas for action*. Descargado de <http://ec.europa.eu/DocsRoom/documents/17924>
- Sufri, N. A. J., Rahmad, N. A., Ghazali, N. F., Shahar, N., y As'ari, M. A. (2019). Vision based system for banknote recognition using different machine learning and deep learning approach. En *2019 iee 10th control and system graduate research colloquium (icsgrc)* (p. 5-8). doi: 10.1109/ICSGRC.2019.8837068
- Szeliski, R. (2010). *Computer vision: Algorithms and applications*. Springer.
- Tamayo Zapata, K. J. (2018). *Sistema de reconocimiento de billetes para personas con discapacidad visual mediante visión artificial* . Universidad EIA, Envigado. (Trabajo de grado para optar al título de Ingeniera Mecatrónica)
- Thomas, H. (2008, 01). Tecnologías para la inclusión social y políticas públicas en américa latina.

- Tipán, D. (2019). *Implementación de un dispositivo identificador de billetes para personas no videntes* (Tesis Doctoral no publicada). Escuela Politécnica Nacional.
- Torres, C., y Rodríguez, C. (2012). *Bastón blanco para prevenir obstáculos* (Tesis de Master no publicada). Ingeniería en control y automatización.
- Torres-Carazo, M. I., Rodríguez-Fórtiz, M. J., y Torres, M. V. H. (2021). Development of an ontology for the inclusion of app users with visual impairments. *IEEE Access*, 9, 44339-44353. doi: 10.1109/ACCESS.2021.3065274
- Unión Mundial de Ciegos. (2023). *Unión mundial de ciegos*. Descargado 29 March 2023, de <https://worldblindunion.org/es/>
- Valipoor, M. M., y de Antonio, A. (2022). Recent trends in computer vision-driven scene understanding for vi/blind users: a systematic mapping. *Universal Access in the Information Society*. doi: 10.1007/s10209-022-00868-w
- Vargas, S. (2020). *Prototipo de gafas para la detección de obstáculos mediante vibraciones para personas invidentes basado en la tecnología arduino* (Tesis Doctoral no publicada). Informática.
- Vasquez Salazar, R., y Cardona Mesa, A. (2019, jun.). Dispositivos de asistencia para la movilidad en personas con discapacidad visual: una revisión bibliográfica. *Revista Politécnica*, 15(28), 107–116. Descargado de <https://revistas.elpoli.edu.co/index.php/pol/article/view/1534> doi: 10.33571/rpolitec.v15n28a10
- Vela, E. (2019). *Diseño e implementación de un bastón ergonómico con sistema de posicionamiento global para mejorar el desplazamiento de personas invidentes en el centro “la unión nacional de ciegos del Perú”* (Tesis de Master no publicada). Ingeniería Electrónica.
- Vilchez, M. (2021). *Bastón sensorial geolocalizador inteligente para apoyar en el desplazamiento de personas invidentes en la organización regional de ciegos del Perú – Chiclayo* (Tesis de Master no publicada). Ingeniería de Sistemas y Computación.
- Vorapatratorn, S., Suchato, A., y Punyabukkana, P. (2021, Jul.). Fast obstacle detection system for the blind using depth image and machine learning. *Engineering and Applied Science Research*, 48(5), 593–603. Descargado de <https://ph01.tci-thaijo.org/index.php/easr/article/view/242952>
- World Health Assembly. (2001). *International classification of functioning, disability and health (icf)*. Descargado 03 May 2023, de <https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health>
- Yuan, Z., Azzino, T., Hao, Y., Lyu, Y., Pei, H., Boldini, A., ... Rizzo, J.-R. (2022). Network-aware 5g edge computing for object detection: Augmenting wearables to “see” more, farther and faster. *IEEE Access*, 10, 29612-29632. doi: 10.1109/ACCESS.2022.3157876

Zeinullin, M., y Hersh, M. (2022). Tactile audio responsive intelligent system. *IEEE Access*, 10, 122074-122091. doi: 10.1109/ACCESS.2022.3223099



# APÉNDICES

## 5.1. Mapa de empatía

Dado que estamos empleando la metodología de Design Thinking, se llevó a cabo un ejercicio de empatía profunda utilizando un mapa de empatía para comprender las necesidades, deseos, frustraciones y comportamientos de los usuarios finales. Este proceso nos permitió identificar de manera más clara las oportunidades de diseño y enfocar nuestros esfuerzos en crear una solución que realmente aborde las necesidades de los usuarios, en este caso, el reconocimiento de la denominación de los billetes.

*Figura 5.2: Mapa de empatía.*



## 5.2. Plan de Costos del Proyecto

**Tabla 5.4**

*Plan de Costos del Proyecto*

Descripción	Cantidad (Unidad)	Costo
Equipos y Bienes		
Placa electrónica y Drivers	1	S/. 199,75.00
Cámara Raspberry Pi de 5MP	1	S/. 40.00
Batería de Litio de 3.7V y 1000mA	2	S/. 14.00
Otros componentes	1	S/. 16,20.00
Servicios Tecnológicos		
Impresión en 3D	1	S/. 0.00
Total		S/. 269.95

## 5.3. Código para el funcionamiento del dispositivo IoT

```
1 # Importar librerías
2 import time
3 import numpy as np
4 import tensorflow as tf
5 from keras.preprocessing import image
6 import RPi.GPIO as GPIO
7 from gtts import gTTS
8 from picamera2 import MappedArray, Picamera2
9 import os
10
11 # Configuración del botón
12 BUTTON_PIN = 13
13 # Define el número de pin GPIO que está utilizando
14 LED_PIN = 6
15
16 # Configura el pin del LED como salida
17 GPIO.setmode(GPIO.BCM)
18 GPIO.setup(LED_PIN, GPIO.OUT)
19
20 GPIO.setup(BUTTON_PIN, GPIO.IN, pull_up_down=GPIO.PUD_UP)
21
22 # Configuración de la comunicación con la bocina
23 SPEAKER_PIN = 18
24 GPIO.setup(SPEAKER_PIN, GPIO.OUT)
25
26 # Cargar el modelo entrenado
27 xception_model = tf.keras.models.load_model('xception_model.hdf5')
28
29 # Configurar la cámara de la Raspberry Pi con libCamera
30 picam2 = Picamera2()
31
32 # Definir un diccionario que mapee clases a índices
33 class_indices = {
34     'billete100_anverso_antiguo': 0,
35     'billete100_anverso_nuevo': 1,
36     'billete100_reverso_antiguo': 2,
37     'billete100_reverso_nuevo': 3,
38     'billete10_anverso_antiguo': 4,
39     'billete10_anverso_nuevo': 5,
40     'billete10_reverso_antiguo': 6,
41     'billete10_reverso_nuevo': 7,
42     'billete20_anverso_antiguo': 8,
43     'billete20_anverso_nuevo': 9,
44     'billete20_reverso_antiguo': 10,
45     'billete20_reverso_nuevo': 11,
46     'billete50_anverso_antiguo': 12,
47     'billete50_anverso_nuevo': 13,
48     'billete50_reverso_antiguo': 14,
49     'billete50_reverso_nuevo': 15
50 }
51
52 # Función para realizar la predicción en una imagen dada
53 def predict_image(img_path):
```

```

54     img = image.load_img(img_path, target_size=(299, 299))
55     img = image.img_to_array(img)
56     img = np.expand_dims(img, axis=0)
57     return xception_model.predict(img)
58
59 # Funci n para reproducir audio mp3
60 def play_audio_mp3(message):
61     if message == "billete10_anverso_antiguo" or "
billete10_anverso_nuevo"
62         or "billete10_reverso_antiguo" or "
billete10_reverso_nuevo":
63         os.system('mpg321 -o alsa 10soles.mp3')
64     if message == "billete20_anverso_antiguo" or "
billete20_anverso_nuevo"
65         or "billete20_reverso_antiguo" or "
billete20_reverso_nuevo":
66         os.system('mpg321 -o alsa 20soles.mp3')
67     if message == "billete50_anverso_antiguo" or "
billete50_anverso_nuevo"
68         or "billete50_reverso_antiguo" or "
billete50_reverso_nuevo":
69         os.system('mpg321 -o alsa 50soles.mp3')
70     if message == "billete100_anverso_antiguo" or "
billete100_anverso_nuevo"
71         or "billete100_reverso_antiguo" or "
billete100_reverso_nuevo":
72         os.system('mpg321 -o alsa 100soles.mp3')
73
74 # Funci n para capturar la imagen cuando se presiona el bot n
75 def capture_image(channel):
76     # Reproducir el audio de captura
77     os.system('mpg321 -o alsa captura.mp3')
78
79     # Inicializar c mara y Capturar imagen
80     picam2.start()
81     picam2.capture_file("test.jpg")
82
83     #Realizar la predicci n en la imagen capturada
84     xception_preds_single = predict_image('test.jpg')
85
86     # Obtener la clase predicha ( ndice )
87     predicted_class_index = np.argmax(xception_preds_single, axis
=1)
88
89     # Verificar si la predicci n est  dentro del rango de clases
v lidas
90     if predicted_class_index[0] in class_indices.values():
91         predicted_class_label = [key for key, value in
class_indices.items() if value == predicted_class_index[0]][0]
92         # Reproducir el tipo de billete detectado
93         play_audio_mp3(predicted_class_label)
94         os.system('mpg321 -o alsa recursos/pressbutton.mp3')
95
96     return
97
98 GPIO.setmode(GPIO.BCM)
99 GPIO.setup(BUTTON_PIN, GPIO.IN, pull_up_down=GPIO.PUD_UP)
100

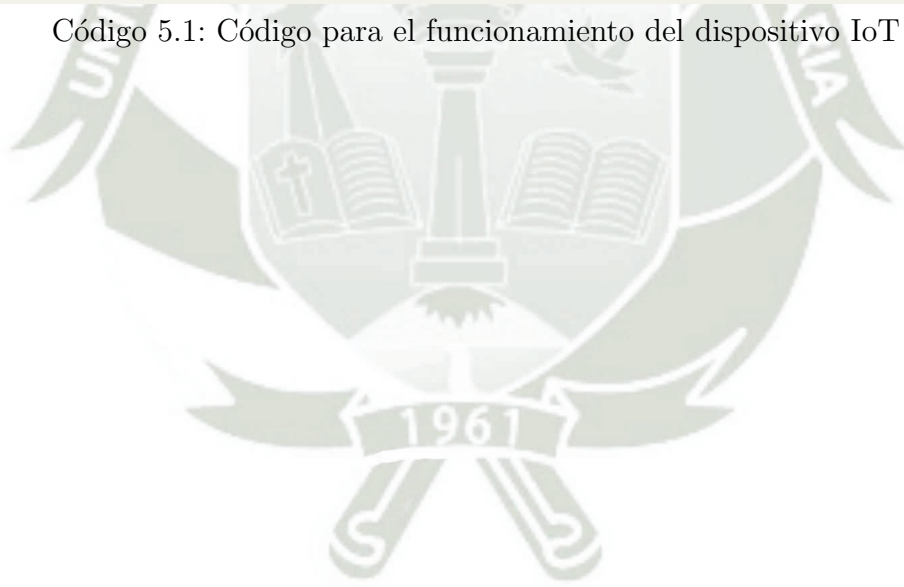
```

```

101 # Maneja el evento del bot n con debouncing
102 def button_event(channel):
103     if GPIO.input(channel) == GPIO.LOW:
104         capture_image(channel)
105         # Espera un tiempo breve para evitar rebotes
106         time.sleep(0.2) # Puedes ajustar este valor seg n sea
                             necesario
107
108 # Agrega el detector de eventos con debouncing
109 GPIO.add_event_detect(BUTTON_PIN, GPIO.FALLING, callback=
    button_event, bouncetime=300)
110
111 try:
112     os.system('mpg321 -o alsa encendido.mp3')
113
114     os.system('mpg321 -o alsa recursos/pressbutton.mp3')
115     GPIO.output(LED_PIN, GPIO.HIGH)
116     print("Presione el boton")
117     while True:
118         time.sleep(2)
119 except KeyboardInterrupt:
120     GPIO.cleanup()
121     print("Saliendo del programa...")
122 finally:
123     GPIO.cleanup()

```

Código 5.1: Código para el funcionamiento del dispositivo IoT



## 5.4. Cuestionarios aplicados

### UEQ - Cuestionario sobre la experiencia de los Usuarios

*Figura 5.3: UEQ - Cuestionario sobre la experiencia de los Usuarios.*

**Evaluación de Usabilidad de Software para personas ciegas**

La presente Pauta tiene por objetivo evaluar la usabilidad de un dispositivo para personas con discapacidad visual.

**Antecedentes**

Nombre del dispositivo

Nombre del usuario Edad Sexo

Nivel de evaluador	Resto Visual
<input type="checkbox"/> Aprendiz <input type="checkbox"/> Normal <input type="checkbox"/> Avanzado	<input type="checkbox"/> Si <input type="checkbox"/> No

	1	2	3	4	5	6	7		
desagradable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	agradable	1
no entendible	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	entendible	2
creativo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	sin imaginación	3
fácil de aprender	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	difícil de aprender	4
valioso	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	de poco valor	5
aburrido	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	emocionante	6
no interesante	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	interesante	7
impredecible	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	predecible	8
rápido	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	lento	9
original	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	convencional	10
obstrutivo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	impulsor de apoyo	11
bueno	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	malo	12
complicado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	fácil	13
repeler	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	atraer	14
convencional	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	novedoso	15
incómodo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	cómodo	16
seguro	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	inseguro	17
activante	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	adormecedor	18
cubre expectativas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	no cubre expectativas	19
ineficiente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	eficiente	20
claro	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	confuso	21
no pragmático	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	pragmático	22
ordenado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	sobrecargado	23
atractivo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	feo	24
simpático	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	antipático	25
conservador	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	innovador	26

Enlace de repositorio: [https://bit.ly/cuestionarios\\_ueq](https://bit.ly/cuestionarios_ueq)

# SUS - Escala de usabilidad del sistema

*Figura 5.4: SUS - Escala de usabilidad del sistema.*



## Prueba de usabilidad y experiencia de usuario (SUS)

Creo que me gustaría usar este sistema con frecuencia	Pienso que hay demasiada inconsistencia en este sistema
<input type="radio"/> (1) Totalmente en desacuerdo	<input type="radio"/> (1) Totalmente en desacuerdo
<input type="radio"/> (2) En desacuerdo	<input type="radio"/> (2) En desacuerdo
<input type="radio"/> (3) Ni de acuerdo ni en desacuerdo	<input type="radio"/> (3) Ni de acuerdo ni en desacuerdo
<input type="radio"/> (4) De acuerdo	<input type="radio"/> (4) De acuerdo
<input type="radio"/> (5) Totalmente de acuerdo	<input type="radio"/> (5) Totalmente de acuerdo
Encontré el sistema innecesariamente complejo.	Me imagino que la mayoría de la gente aprendería a usar este sistema muy rápidamente
<input type="radio"/> (1) Totalmente en desacuerdo	<input type="radio"/> (1) Totalmente en desacuerdo
<input type="radio"/> (2) En desacuerdo	<input type="radio"/> (2) En desacuerdo
<input type="radio"/> (3) Ni de acuerdo ni en desacuerdo	<input type="radio"/> (3) Ni de acuerdo ni en desacuerdo
<input type="radio"/> (4) De acuerdo	<input type="radio"/> (4) De acuerdo
<input type="radio"/> (5) Totalmente de acuerdo	<input type="radio"/> (5) Totalmente de acuerdo
Pienso que el sistema estaba listo para usar	Encontré el sistema muy engorroso de usar
<input type="radio"/> (1) Totalmente en desacuerdo	<input type="radio"/> (1) Totalmente en desacuerdo
<input type="radio"/> (2) En desacuerdo	<input type="radio"/> (2) En desacuerdo
<input type="radio"/> (3) Ni de acuerdo ni en desacuerdo	<input type="radio"/> (3) Ni de acuerdo ni en desacuerdo
<input type="radio"/> (4) De acuerdo	<input type="radio"/> (4) De acuerdo
<input type="radio"/> (5) Totalmente de acuerdo	<input type="radio"/> (5) Totalmente de acuerdo
Creo que necesitaría el apoyo de una persona técnica para poder usar este sistema	Me sentí muy confiado usando el sistema
<input type="radio"/> (1) Totalmente en desacuerdo	<input type="radio"/> (1) Totalmente en desacuerdo
<input type="radio"/> (2) En desacuerdo	<input type="radio"/> (2) En desacuerdo
<input type="radio"/> (3) Ni de acuerdo ni en desacuerdo	<input type="radio"/> (3) Ni de acuerdo ni en desacuerdo
<input type="radio"/> (4) De acuerdo	<input type="radio"/> (4) De acuerdo
<input type="radio"/> (5) Totalmente de acuerdo	<input type="radio"/> (5) Totalmente de acuerdo
Encontré que las varias funciones en este sistema estaban bien integradas	Tengo que aprender muchas cosas antes de poder ponerme en marcha con este sistema
<input type="radio"/> (1) Totalmente en desacuerdo	<input type="radio"/> (1) Totalmente en desacuerdo
<input type="radio"/> (2) En desacuerdo	<input type="radio"/> (2) En desacuerdo
<input type="radio"/> (3) Ni de acuerdo ni en desacuerdo	<input type="radio"/> (3) Ni de acuerdo ni en desacuerdo
<input type="radio"/> (4) De acuerdo	<input type="radio"/> (4) De acuerdo
<input type="radio"/> (5) Totalmente de acuerdo	<input type="radio"/> (5) Totalmente de acuerdo

Enlace de repositorio: [https://bit.ly/cuestionario\\_sus](https://bit.ly/cuestionario_sus)

## Expresión de consentimiento informado

Figura 5.5: Expresión de consentimiento informado.

 **Universidad Católica de Santa María** 

**EXPRESIÓN DE CONSENTIMIENTO INFORMADO**


Fecha: 10 de Febrero del 2024 Hora: 16:13

Yo, Cesar Avendaño Mamani con DNI N° 43628423, en calidad de Presidente de la Asociación Unión de Ciegos de Arequipa declaro lo siguiente:

- Que doy mi consentimiento para que los miembros de la Asociación Unión de Ciegos de Arequipa sean parte del proyecto de investigación (tesis) titulado "Desarrollo de un dispositivo de asistencia ergonómico y de bajo costo para reconocimiento de billetes basado en visión artificial e IoT para personas con discapacidad visual en la ciudad de Arequipa".
- Conozco los derechos y responsabilidades como miembros del proyecto de investigación.

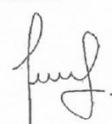
En este sentido, he sido informado(a) de los beneficios, ventajas y consecuencias del uso del dispositivo y, resueltas todas las preguntas y dudas al respecto, consciente de mis derechos y en forma voluntaria, en cumplimiento de la normativa vigente; *si* doy mi consentimiento para ser parte del proyecto de investigación.


De la parte interesada:

  
César Avendaño Mamani  
PRESIDENTE

**Nombres y Apellidos:** Cesar Mamani Avendaño  
**DNI:** 43628423

De los estudiantes:

  
**Nombres y Apellidos:** Jackeline Melady Peña Alejandro  
**DNI:** 72525180

  
**Nombres y Apellidos:** Nicolás Esleyder Cayturo Silva  
**DNI:** 70369522

## 5.5. Comparación con otras placas electrónicas

**Tabla 5.5**

*Comparación con otras placas electrónicas*

Característica	Arduino Nano	Raspberry Pi 3B+	Raspberry Pi 4B+
Microprocesador	ATmega328P	Broadcom BCM2837B0	Broadcom BCM2711
Arquitectura	AVR	ARM	ARM
Potencia de Procesamiento	Baja	Media	Alta
Memoria RAM	2KB	1 GB	Disponible en modelos de 2 GB, 4 GB y 8 GB.
Almacenamiento	Depende del modelo específico de la tarjeta SD o EEPROM que se utilice.	Mediante tarjeta microSD.	Mediante tarjeta microSD, pero también tiene la opción de almacenamiento a través de puertos USB.
Conectividad	Generalmente se programa y se conecta a través de un cable USB. Puede interactuar con una variedad de sensores y dispositivos electrónicos.	Tiene puertos HDMI, USB, Ethernet, Wi-Fi y Bluetooth integrados. También tiene pines GPIO para interactuar con dispositivos externos.	Similar al Raspberry Pi 3B+, pero con puertos USB 3.0, un puerto USB-C para la alimentación y un puerto micro HDMI para video. También tiene opciones de conectividad Wi-Fi y Bluetooth mejoradas.
Propósito	Diseñado para aplicaciones de electrónica de bajo nivel y proyectos embebidos.	Ideal para proyectos de computación de bajo costo, servidores domésticos, centros multimedia, estaciones de trabajo ligeras y proyectos de IoT.	Ofrece un rendimiento mejorado sobre el Raspberry Pi 3B+ y es adecuado para una variedad de aplicaciones.

## 5.6. Artículo publicado en Scopus

El artículo fue publicado en una revista cuartil 2 (Q2) de Scopus (Cayturo-Silva et al., 2023). El nombre de la revista es Data in Brief.





## Data Article

# Annotated Peruvian banknote dataset for currency recognition and classification



Nicolás Esleyder Cayturo-Silva<sup>a,\*</sup>,  
Jackeline Melady Peña-Alejandro<sup>a</sup>, Eveling Gloria Castro-Gutierrez<sup>a</sup>,  
Jose Sulla-Torres<sup>a</sup>, Benjamin Maraza-Quispe<sup>b</sup>

<sup>a</sup> Universidad Católica de Santa María, Arequipa Peru

<sup>b</sup> Universidad Nacional de San Agustín de Arequipa, Arequipa Peru

## ARTICLE INFO

*Article history:*

Received 15 September 2023

Revised 11 October 2023

Accepted 19 October 2023

Available online 24 October 2023

Dataset link: [Peruvian Banknotes \(Original data\)](#)

*Keywords:*

Banknote recognition

Peruvian banknote

Machine learning

Annotated dataset

## ABSTRACT

The real-time detection of multinational banknotes remains an ongoing research challenge within the academic community. Numerous studies have been conducted to address the need for rapid and accurate banknote recognition, counterfeit detection, and identification of damaged banknotes [1–3]. State-of-the-art techniques, such as machine learning (ML) and deep learning (DL), have supplanted traditional digital image processing methods in banknote recognition and classification. However, the success of ML or DL projects critically hinges on the size and comprehensiveness of the datasets employed. Existing datasets suffer from several limitations. Firstly, there is a notable absence of a Peruvian banknote dataset suitable for training ML or DL models. Second, the lack of annotated data with specific labels and metadata for Peruvian currency hinders the development of effective supervised learning models for banknote recognition and classification. Lastly, datasets from different regions may not align with the unique characteristics, design, and security features of Peruvian banknotes, limiting the accuracy and applicability of models in a Peruvian context [4]. To address these limitations, we have meticulously curated a compre-

\* Corresponding author.

E-mail addresses: [nicolas.cayturo@ucsm.edu.pe](mailto:nicolas.cayturo@ucsm.edu.pe) (N.E. Cayturo-Silva), [jackeline.pena@ucsm.edu.pe](mailto:jackeline.pena@ucsm.edu.pe) (J.M. Peña-Alejandro), [ecastrog@ucsm.edu.pe](mailto:ecastrog@ucsm.edu.pe) (E.G. Castro-Gutierrez), [jsullato@ucsm.edu.pe](mailto:jsullato@ucsm.edu.pe) (J. Sulla-Torres), [bmaraza@unsa.edu.pe](mailto:bmaraza@unsa.edu.pe) (B. Maraza-Quispe).

Social media: [@cayturo\\_silva](#) (N.E. Cayturo-Silva)

hensive dataset comprising a total of 9,315 images of Peruvian banknotes, encompassing both old and new denominations from 2011 (old) and 2019 (new) [5]. The Peruvian banknote dataset includes denominations of 10, 20, 50, and 100 Peruvian soles. Importantly, as indicated by [5], both the 2011 and 2019 families of banknotes are currently in circulation, further enhancing the dataset's relevance for real-world applications in currency recognition and verification.

This dataset serves as a vital resource for addressing the challenges in real-time multinational banknote detection. By offering a comprehensive collection of images of Peruvian banknotes, both old and new, this dataset fills a critical gap in the field of banknote recognition. Researchers can utilize it to train and evaluate advanced machine learning and deep learning models, ultimately enhancing the accuracy of banknote processing systems.

© 2023 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

## Specifications Table

Subject	Machine Learning
Specific subject area	Banknote detection and identification
Data format	Raw Labelled
Type of data	The Peruvian Banknote dataset comprises .jpg images with dimensions of 640 × 480 pixels and a resolution of 72 DPI.
Data collection	The images of Peruvian banknotes were captured using a high-resolution smartphone camera. The original banknote images in .jpg format have dimensions of 3072 × 3072 pixels. These images were resized to 640 × 480 pixels. There are a total of 16 classes of Peruvian banknotes, including denominations of 10, 20, 50, and 100 Peruvian soles (considered as separate classes for both the obverse and reverse sides of the banknotes). The banknote images were taken under various environmental conditions, including dark backgrounds, illuminated backgrounds, cluttered environments, and with folded banknotes, both within and outside the laboratory.
Data source location	Catholic University of Santa Maria De Arequipa Urb. San José, San Jose s/n, Yanahuara Arequipa, Peru
Data accessibility	Repository name: Peruvian Banknotes [6] Data identification number(doi): <a href="https://doi.org/10.17632/22jmt8xhpn.1">10.17632/22jmt8xhpn.1</a> Direct URL to data: <a href="https://data.mendeley.com/datasets/22jmt8xhpn/1">https://data.mendeley.com/datasets/22jmt8xhpn/1</a>

## 1. Value of the Data

- This dataset represents a significant and valuable addition to the available resources on the internet, comprising a total of 9,315 high-quality images spanning 16 distinct categories.
- Notably, it includes both old and new denominations of Peruvian banknotes, offering comprehensive coverage of both the obverse and reverse sides of the bills.
- The dataset serves as a highly valuable resource for the development of applications centered around the classification and detection of Peruvian banknotes, making it particularly advantageous for researchers specializing in banknote classification and identification.
- Its versatility is a key asset, as it can be effectively employed for various applications, encompassing training, validation, and testing phases for both banknote classification and identification models.
- Moreover, the dataset plays a pivotal role in the precise valuation of Peruvian banknotes, significantly aiding financial transactions and monetary activities.

- It's important to note that this dataset is not intended for the classification of whether banknotes are genuine or counterfeit. Instead, its primary focus lies in enabling advanced machine and deep learning models for the accurate recognition and categorization of Peruvian banknotes.
- Beyond its financial applications, this dataset extends its impact to the development of tools for banknote recognition and classification, thereby benefiting not only visually impaired individuals but also bank customers and governmental agencies alike. The dataset's potential to enhance accessibility and facilitate monetary interactions underscores its broader societal significance.

## 2. Data Description

The creation of a banknote dataset is of paramount importance for several compelling reasons. Firstly, the accurate recognition of banknotes is an essential task for automated teller machines (ATMs) and currency identification machines [7] cited by [4]. Moreover, the development of systems capable of authenticating banknotes is essential [8]. Additionally, banknote recognition remains a significant challenge for individuals with visual impairments, underscoring its vital role in promoting financial accessibility [9].

Our banknote dataset represents a comprehensive collection of Peruvian currency. It encompasses 16 distinct classes, encompassing denominations of 10, 20, 50, and 100 Peruvian soles. Notably, each denomination is treated as a separate class, accounting for both the obverse and reverse sides of the banknotes. The dataset includes banknote images captured under various environmental conditions, ranging from well-illuminated settings to low-light environments and cluttered backgrounds. Furthermore, it contains images of partially folded or partially obscured banknotes. Sample images from the dataset, captured in diverse environmental scenarios, are depicted in Figs. 1 and 2.

The Peruvian banknote images are stored within the "Peruvian\_Banknotes" folder, serving as the central repository. This primary folder is further subdivided into three subfolders: "training\_set," "validation\_set," and "testing\_set". Both the training and validation sets comprise 16 subfolders, each housing classes representative of the dataset. These classes encompass banknotes from the 2011 and 2019 currency families [5], encompassing denominations of 10, 20, 50, and 100 Peruvian soles, and encompassing both the obverse and reverse sides of the banknotes. Finally, the testing folder includes 261 images sourced from everyday environments, summing up to a total of 9,315 images of Peruvian banknotes.

According with Table 1, the distribution of denominations in the dataset is illustrated in Fig. 3. Fig. 3 displays the representation of different banknote denominations. The 10 Soles note from the new family comprises 8% of the dataset, while the 10 Soles note from the old family accounts for 11%. The new family 20 Soles note is 10% of the dataset, while the old family 20 Soles note represents 12%. The new family 50 Soles note constitutes 13% of the dataset, while the old family 50 Soles note holds the highest share at 20%. The new family 100 Soles note is 10% of the dataset, while the old family 100 Soles note is 14%.

**Table 1**

Total denominations.

	Denomination				Total
	10	20	50	100	
Banknote family new	728	894	1190	939	3751
Banknote family old	1025	1162	1831	1285	5303
Total	1753	2056	3021	2224	9054

The directory structure of the Peruvian Banknotes dataset is shown in Fig. 4. Fig. 4 describes the folder structure of the Peruvian Banknotes Dataset. The dataset's root folder, labeled as



**Fig. 1.** Banknote images taken in various environments (Scenario 1-4/From 7).

“PERUVIAN\_BANKNOTES”, encompasses three primary subfolders: “training\_set”, “validation\_set” and “testing\_set”. Within these subfolders, the dataset is further categorized into 16 distinct classes, each representing a different denomination of Peruvian currency, encompassing both the obverse and reverse sides of the banknotes.

To further clarify, this dataset exclusively focuses on genuine Peruvian banknotes for recognition and classification purposes. It does not include any counterfeit banknotes.

### 3. Experimental Design, Materials and Methods

#### 3.1. Experimental design

The process of acquiring image data is illustrated in Fig. 5. Banknote images were obtained using the high-resolution rear camera of a Redmi Note 12 smartphone (see Table 2 for device

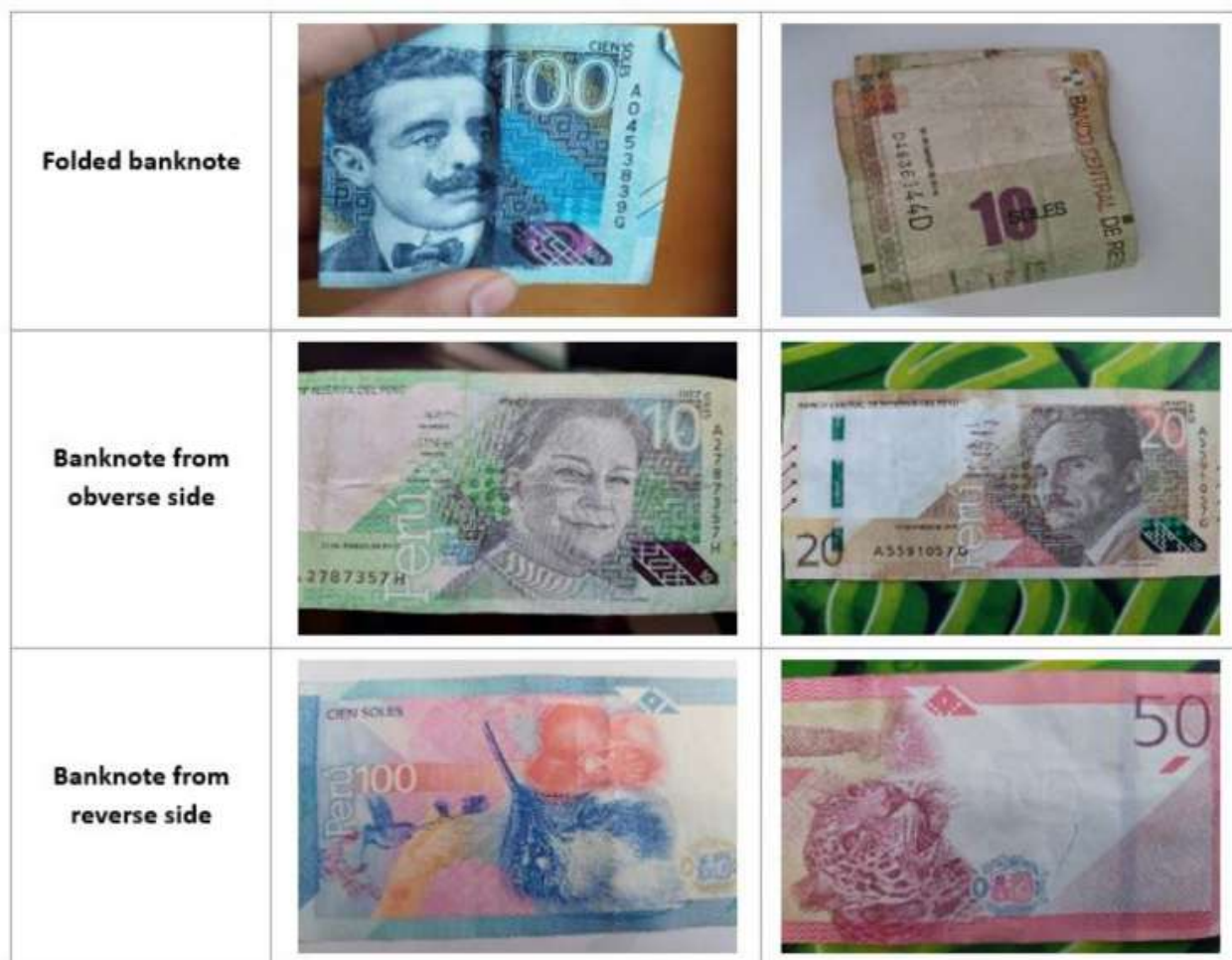


Fig. 2. Banknote images taken in various environments (Scenario 5-7/From 7).

Table 2

Device specifications for banknote capture.

	Description
Processor	Snapdragon 685
RAM	8 GB LPDDR4X
Storage	128 GB
Front Camera	13 megapixels f/2.45
Main Rear Camera	50 MP f/1.8
Wide-angle	8 MP f/2.2
Macro	2 MP f/2.4
Operating System	Android 12 + MIUI 14

Table 3

Data acquisition steps.

Step	Duration	Activity
Data Gathering	April to July	Daily during daytime captured the banknote images
Image manual processing	July to September	Separate the images into their respective folders. Resize the images and change their names accordingly.

specifications). A total of 9,315 images were captured and subsequently categorized and stored in respective folders based on their denominations and family values.

Table 3 outlines the steps involved in the data acquisition process, while Table 4 provides specifics regarding image acquisition and preprocessing tools. As indicated in Table 5, the banknote images were captured from various angles and backgrounds. Subsequently, following the

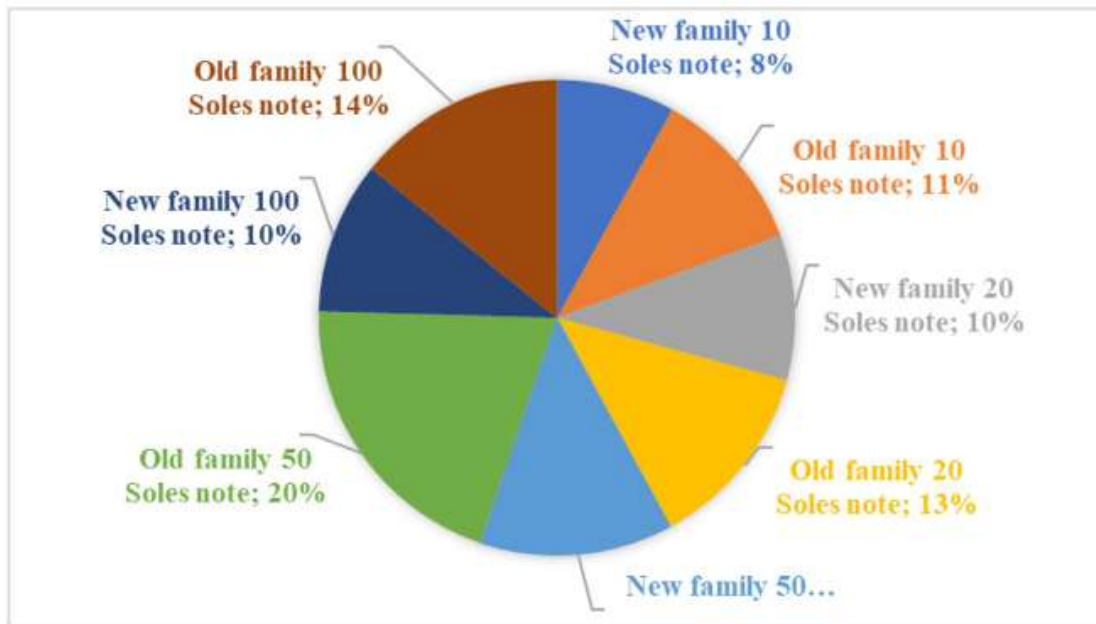


Fig. 3. Percentage of each currency denomination in the Peruvian Banknotes dataset.

```

PERUVIAN_BANKNOTES
├── testing_set
├── training_set
│   ├── billete100_anverso_antiguo
│   ├── billete100_anverso_nuevo
│   ├── billete100_reverso_antiguo
│   ├── billete100_reverso_nuevo
│   ├── billete10_anverso_antiguo
│   ├── billete10_anverso_nuevo
│   ├── billete10_reverso_antiguo
│   ├── billete10_reverso_nuevo
│   ├── billete20_anverso_antiguo
│   ├── billete20_anverso_nuevo
│   ├── billete20_reverso_antiguo
│   ├── billete20_reverso_nuevo
│   ├── billete50_anverso_antiguo
│   ├── billete50_anverso_nuevo
│   ├── billete50_reverso_antiguo
│   └── billete50_reverso_nuevo
└── validation_set
    ├── billete100_anverso_antiguo
    ├── billete100_anverso_nuevo
    ├── billete100_reverso_antiguo
    ├── billete100_reverso_nuevo
    ├── billete10_anverso_antiguo
    ├── billete10_anverso_nuevo
    ├── billete10_reverso_antiguo
    ├── billete10_reverso_nuevo
    ├── billete20_anverso_antiguo
    ├── billete20_anverso_nuevo
    ├── billete20_reverso_antiguo
    ├── billete20_reverso_nuevo
    ├── billete50_anverso_antiguo
    ├── billete50_anverso_nuevo
    ├── billete50_reverso_antiguo
    └── billete50_reverso_nuevo
  
```

Fig. 4. Peruvian Currency dataset directory structure.

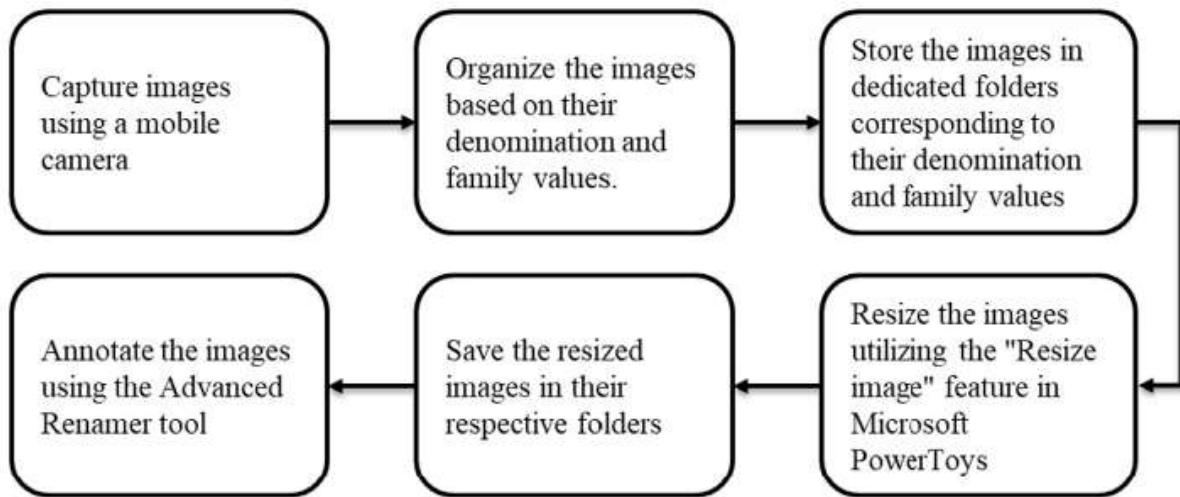


Fig. 5. Peruvian Banknote dataset acquisition process.

Table 4

Specification of image acquisition and preprocessing tools.

No.	Particular	Details
1	Camera	(i) Make and Model: Redmi Note 12 (ii) Rear Camera 50-megapixel (f/1.8) (iii) Rear autofocus
2	Battery	5.000 mAh
3	Labelling Software	Advanced Renamer
4	Resizing Software	Microsoft PowerToys
5	Image Resolution	640 × 480 pixels
6	Image Format	JPG

Table 5

Banknote image capture conditions.

Denominations	Direction of image Capturing	Different backgrounds considered for image capturing	No. of images of each denomination	Total No. of images
10 New and Old, 20 New and Old, 50 New and Old, 100 New and Old, Peruvian Soles.	Front Direction, Front Direction, Rotated 180°, Backward Direction, Backward Direction	illuminated, Dark, cluttered, occluded.	10 Soles – 1753 20 Soles – 2056 50 Soles – 3021 100 Soles – 2224	9315
Both obverse and reverse side.	Rotated 180°, Half folded.		Test Soles – 261	

capture of the Peruvian banknote images, they were meticulously organized into dedicated folders. For a comprehensive breakdown of the folder structure and file names, please refer to Table 6.

The dataset creation process lasted for 5 months, starting in April 2023, and concluding in September 2023. It is important to mention that it was a rigorous process, beginning with the capture of banknote images and involving manual image processing. The manual processing helped in reducing the image sizes and sorting each image into its appropriate folder. A total of 16 folders were used, which represent the classes for training machine and deep learning models.

**Table 6**

Structure and naming of image directories.

Type	Description	Nomenclature used
Directories	The captured images were organized into 16 distinct folders. These folders represent different combinations of banknote denominations, sides (obverse, reverse), and banknote families (new/old).	For example: <ul style="list-style-type: none"> <li>• banknote10_obverse_old</li> <li>• banknote10_reverse_old</li> <li>• banknote10_obverse_new</li> <li>• banknote10_reverse_new</li> </ul>
Files	Files Each image was labeled following specific nomenclature to enable identification and categorization during the training process. The nomenclature includes: <ul style="list-style-type: none"> <li>• The dataset type to which the image belongs (training, validation, or testing).</li> <li>• The denomination of the banknote (10, 20, 50, or 100).</li> <li>• The banknote family (new or old).</li> <li>• A sequential enumeration for unique identification of each image.</li> </ul>	Ejemplo de etiqueta: "training_10_old_001"

### 3.2. Materials or specification of image acquisition system

The banknote images were captured using a Redmi Note 12 smartphone equipped with a 50 MP rear camera. To ensure uniformity, all original images, originally sized at  $3072 \times 3072$  pixels, were resized to  $640 \times 480$  dimensions using the "Resize image" utility in Microsoft PowerToys. These images were saved in .jpg format. The image capture process encompassed various environmental conditions, including different lighting conditions, backgrounds, angles, as well as scenarios involving folded and occluded banknotes, mirroring real-world situations.

Subsequently, the captured images were meticulously organized into training, validation, and testing sets. The dataset comprises a total of 16 distinct classes, representing Peruvian banknotes with denominations of 10, 20, 50, and 100 Peruvian soles. Notably, each denomination is treated as a separate class, accounting for both the obverse and reverse sides of the banknotes. For a detailed overview of the directory structure of the image dataset, please refer to [Table 6](#). To enhance their utility, the images were annotated using the Advanced Renamer tool, with annotations and banknote images stored in their respective folders.

### 3.3. Method

All banknote images were captured using the rear camera of a Redmi Note 12 mobile phone, encompassing various angles and backgrounds. These images, originally sized at  $3072 \times 3072$  pixels, were subsequently resized to  $640 \times 480$  pixels using the "Resize image" utility within PowerToys. The resizing and renaming process is detailed in [Table 4](#) for reference. [Table 5](#) provides a comprehensive overview of the classes, the quantity of images taken, and the environmental conditions under which these images were taken.

### Limitations

None.

## Ethics statement

There is no funding present for present effort. There is no conflict of interest. The data is available in public domain.

## Ethical Statement

We appreciate the opportunity to submit our dataset, the “Annotated Dataset of Peruvian Banknotes,” for your review and possible publication in Data in Brief. We acknowledge the importance of ethics in scientific publishing and wish to assure you that we have adhered to the highest ethical standards at every stage of our work.

In accordance with the ethical publishing guidelines provided by Elsevier and Data in Brief, we have considered and complied with the following key ethical aspects:

1. **Authorship of the paper:** We have limited authorship of our article to those who have made significant contributions to the conception, design, execution, and/or interpretation of our study. Furthermore, we have ensured transparency of author contributions by including a CRediT statement in our article.
2. **Originality and plagiarism:** All materials presented in our dataset are entirely original. We have ensured that, in the event of using the work and/or words of others, it has been appropriately cited and referenced.
3. **Data access and retention:** We are prepared to provide public access to raw data related to our dataset for editorial review. We believe data availability is essential for transparency and the reliability of scientific research.
4. **Multiple, redundant, or simultaneous publication:** We are committed to not publishing manuscripts describing essentially identical research in more than one journal or primary publication. We acknowledge and comply with Elsevier’s policies on prior publication.
5. **Acknowledgment of sources:** Proper acknowledgment has been given to the work of other researchers and authors whose materials have influenced our dataset.
6. **Disclosure and conflicts of interest:** Full disclosures of all relationships that could be viewed as presenting a potential conflict of interest in relation to our dataset have been provided.
7. **Fundamental errors in published works:** If we were to discover a significant error in our dataset in the future, we undertake to promptly notify the journal editor and cooperate with the retraction or correction process as necessary.
8. **Reporting standards:** An accurate and objective description of the work performed in creating our dataset, as well as a discussion of its significance, has been presented.
9. **Hazards and human or animal subjects:** No materials, chemicals, procedures, or equipment presenting unusual hazards inherent in their use have been used. Additionally, our dataset does not involve the participation of human or animal subjects.
10. **Use of patient images or case details:** Given that our dataset consists of images of Peruvian banknotes, it does not require the approval of ethics committees or informed consent from patients or volunteers.

We hope this ethical statement reinforces our commitment to integrity and ethics in research and ensures the quality and reliability of our dataset and its potential publication in Data in Brief.

## Data Availability

[Peruvian Banknotes \(Original data\)](#) (Mendeley Data)

## CRedit Author Statement

**Nicolás Esleyder Caytairo-Silva:** Conceptualization, Methodology, Software, Writing – original draft, Project administration; **Jackeline Melady Peña-Alejandro:** Investigation, Resources, Data curation, Conceptualization; **Eveling Gloria Castro-Gutierrez:** Conceptualization, Supervision, Project administration, Validation; **Jose Sulla-Torres:** Supervision, Validation, Resources; **Benjamin Maraza-Quispe:** Supervision, Validation, Resources.

## Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

The authors would like to thank to the editor and the reviewers for their helpful suggestions and valuable comments.

## Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] S. Kulkarni, y I. Shende, Sasturkar<sup>3</sup>, «Banknote classification using transfer learning», *EPRA Int. J. Multidiscipl. Res. (IJMR)* 9 (5) (2023) 382–388 n.
- [2] V. Meshram, K. Patil, y V. Meshram, «Evaluation of top pretrained models using transfer learning on banknote dataset with quality parameter», *Ingénierie des Systèmes d'Information* 28 (3) (2023) 693–701 n.
- [3] «An efficient deep learning model using network pruning for fake banknote recognition», César Pachón; Dora Ballesteros; Diego Renza 233 (2023) 1–10.
- [4] K.P.P.C. Vidula Meshrama, «Dataset of Indian and Thai banknotes with annotations», *Data Brief* 41 (2022) 1–7.
- [5] BCRP, «Banco Central de Reserva del Perú,» 2023. [En línea]. Available: <https://www.bcrp.gob.pe/billetes-y-monedas/familia-de-billetes.html>. [Último acceso: 13 September 2023].
- [6] N. Caytairo-Silva, «Peruvian Banknotes», Mendeley Data, Arequipa, 2023.
- [7] y T.D. Pham, Y.H. Park, S.Y. Kwon, K.R. Park, D.S. Jeong, S. Yoon, «Efficient banknote recognition based on selection of discriminative regions with one-dimensional visible-light line sensor», *Sensors* 16 (3) (2016). n.
- [8] «Banknote validation through an embedded RFID chip and an NFC-enabled smartphone» Mohamed Hamdy Eldefrawy, Muhammad Khurram Khan, 2015.
- [9] T.D. Pham, Y.W. Lee, C. Park, y K.R. Park, «Deep learning-based detection of fake multinational banknotes in a cross-dataset environment utilizing smartphone cameras for assisting visually impaired individuals», *Mathematics* 10 (9) (2023) n.