



EDUCACIÓN

SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Tecnológico Nacional de México

**Centro Nacional de Investigación
y Desarrollo Tecnológico**

Tesis de Doctorado

**Modelo de diferenciación de orejas humanas naturales,
mediante imágenes**

presentada por

MSC. Eydi López Hernández

como requisito para la obtención del grado de
Doctor en Ciencias de la Computación

Directora de tesis

Dra. Andrea Magadán Salazar

Codirector de tesis

Dr. Osslán Osiris Vergara Villegas

Cuernavaca, Morelos, México. Junio de 2026.

 Centro Nacional de Investigación y Desarrollo Tecnológico	ACEPTACIÓN DE IMPRESIÓN DEL DOCUMENTO DE TESIS DOCTORAL		Código: CENIDET-AC-006-D20
	Referencia a la Norma ISO 9001:2008 7.1, 7.2.1, 7.5.1, 7.6, 8.1, 8.2.4		Revisión: 0
			Página 1 de 1

Cuernavaca, Mor., a 23 de enero de 2026

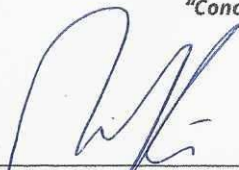
DR. CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO
P R E S E N T E

AT'n: DRA. ANDREA MAGADÁN SALAZAR
PRESIDENTA DEL CLAUSTRO DOCTORAL DEL
DEPARTAMENTO DE CIENCIAS COMPUTACIONALES

Los abajo firmantes, miembros del Comité Tutorial de la estudiante **M.C. EYDI LÓPEZ HERNÁNDEZ** manifiestan que después de haber revisado el documento de tesis titulado **"Modelo de diferenciación de orejas humanas naturales, mediante imágenes"**, realizado bajo la dirección de la **Dra. Andrea Mgadán Salazar** y codirección del **Dr. Osslan Osiris Vergara Villegas**, el trabajo se **ACEPTA** para proceder a su impresión.

ATENTAMENTE

Excelencia en Educación Tecnológica®
"Conocimiento y Tecnología al Servicio de México"


DRA. ANDREA MAGADÁN SALAZAR
TECNM/CENIDET

Osslan Osiris Vergara Villegas
DR. OSSLAN OSIRIS VERGARA VILLEGAS
UNIVERSIDAD AUTÓNOMA DE CIUDAD JUÁREZ


DR. RAÚL PINTO ELÍAS
TECNM/CENIDET


DR. NIMROD GONZÁLEZ FRANCO
TECNM/CENIDET


DR. NOÉ ALEJANDRO CASTRO SÁNCHEZ
TECNM/CENIDET


DR. MIGUEL ÁNGEL ZÚÑIGA GARCÍA
ITESM

c.c.p: C Verónica Sotelo Boyas/ Jefa del Departamento de Servicios Escolares
c.c.p: C Noé Alejandro Castro Sánchez / Jefe del Departamento de Ciencias Computacionales
c.c.p: Expediente



Cuernavaca Mor, 12/febrero/2026

Oficio No. SAC/062/2026

Asunto: Autorización de impresión de tesis

EYDI LÓPEZ HERNÁNDEZ
CANDIDATA AL GRADO DE DOCTORA EN CIENCIAS
DE LA COMPUTACIÓN
P R E S E N T E

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"Modelo de diferenciación de orejas humanas naturales, mediante imágenes"**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

ATENTAMENTE

Excelencia en Educación Tecnológica®

"Conocimiento y Tecnología al servicio de México"

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO



c.c.p. Departamento de Ciencias Computacionales
Departamento de Servicios Escolares

CMAZ/lmz



2026
año de
Margarita
Maza

Dedicatoria

A la memoria de mis padres, la Sra. Bertha Hernández y el Sr. Lorenzo López, quienes desde el cielo me siguen guiando y cuidando. Por sus enseñanzas, por los valores que sembraron en mí y por mostrarme el verdadero significado del esfuerzo, la dignidad y la calidad humana. Hasta el cielo, con todo mi amor.

A mis amados hermanitos, Saby, Galdy, Cacun, Marce y Made, por estar a mi lado en todo momento y en cualquier circunstancia. Son parte esencial de mi historia, de mi vida y de lo que soy.

A mis queridos sobrinos, mis primeros maestros de vida, por recordarme la importancia de disfrutar cada etapa y cada instante. Para ustedes, con todo mi cariño: para todo hay tiempo; vivan, aprendan, sueñen y disfruten cada momento.

A mi esposo, Félix, mi amigo, mi cómplice de locuras y mi apoyo constante durante este proyecto. Tu presencia, paciencia y amor fueron fundamentales en este camino.

A mi amada Luna, por enseñarme el amor más profundo, por ser mi inspiración diaria y por recordarme, con tu existencia, la razón más grande para seguir adelante.

A mi familia, mi orgullo, mi paz y mi dignidad, porque en ustedes encuentro mis raíces, mi fuerza y el sentido de pertenencia que me sostiene.

Agradecimientos

Expreso mi más sincero reconocimiento y gratitud a las instituciones y personas que hicieron posible la culminación del proyecto doctoral:

Al Tecnológico Nacional de México / CENIDET, por facilitarme el uso de sus instalaciones y recursos, fundamentales para el desarrollo de esta investigación.

Al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT), por el apoyo económico brindado, el cual fue el motor para la culminación del trabajo.

A mi directora de tesis, la Dra. Andrea Magadán Salazar, con profunda admiración, cariño y respeto. Su labor trascendió la dirección académica; agradezco su interés genuino por mi bienestar integral, su tiempo compartido y sus invaluable consejos.

A mi codirector, el Dr. Osslan Osiris Vergara Villegas, por su tiempo, dedicación y las recomendaciones precisas que fortalecieron el proyecto.

A los integrantes de mi comité revisor: Dr. Noé Alejandro Castro Sánchez, Dr. Nimrod González Franco, Dr. Raúl Pinto Elías, Dr. Gerardo Reyes Salgado y Dr. Miguel Ángel Zuñiga García. Sus revisiones críticas y puntos clave fueron esenciales para llevar la tesis a buen término.

Al Dr. Jorge Alberto Fuentes Pacheco, por su apoyo técnico y los conocimientos aplicados que permitieron resolver retos cruciales de la problemática abordada.

A mis compañeros y amigos del departamento, quienes se convirtieron en mi familia extendida. Gracias por hacer de este proceso una experiencia llena de momentos felices e inolvidables.

A Dios, desde lo más profundo de mi ser, por guiar cada uno de mis pasos y recordarme en todo momento su amor y cuidado.

Gracias, gracias, gracias.

RESUMEN

La oreja posee características distintivas que han contribuido a su creciente relevancia en las últimas décadas. Entre sus ventajas se encuentran la estabilidad morfológica en el tiempo y la posibilidad de captura sin contacto físico. Por ello, se utiliza como rasgo biométrico para los sistemas de reconocimiento de personas. No obstante, al igual que otras modalidades biométricas, los sistemas basados en oreja pueden ser vulnerables a ataques de presentación (PAs), especialmente ante el acceso creciente a técnicas de impresión 3D y a modelos generativos. En el contexto anterior, la presente tesis propone un modelo computacional para diferenciar entre orejas humanas naturales y artificiales mediante imágenes, en escenarios controlados y no controlados.

Como parte de la investigación, se conformó el Conjunto de Imágenes de Orejas Humanas Naturales y Artificiales (CIOHNA), integrado por muestras naturales provenientes de cuatro conjuntos públicos y capturas propias, así como por muestras artificiales generadas a partir de dibujos, impresiones 3D, imágenes sintéticas e imágenes producidas mediante herramientas de inteligencia artificial.

La solución propuesta se organiza en dos etapas principales. La primera corresponde a la localización de la oreja. Para ello, se integra un enfoque híbrido que combina MediaPipe, utilizado para identificar puntos faciales de referencia, con el modelo YOLOv5s entrenado para detectar orejas humanas. A partir de los puntos faciales se estiman las dimensiones de la oreja con base en las proporciones antropométricas propuestas por Leonardo da Vinci, mientras que YOLOv5s actúa como mecanismo complementario cuando los puntos faciales no son suficientes o no pueden ser localizados adecuadamente. Esta etapa fue evaluada con imágenes provenientes de conjuntos públicos y capturas propias, considerando condiciones controladas y no controladas, alcanzando Exactitud de 97.00 %, Precisión de 99.00 %, *Recall* de 96.00 % y *F1-score* de 97.00 %.

La segunda etapa aborda la clasificación entre orejas humanas naturales y artificiales como estrategia de detección de ataques de presentación. Para ello, se propone el Modelo 4C, basado en una representación de cuatro canales compuesta por la imagen RGB y un canal adicional de textura generado mediante Patrones Binarios Locales (LBP). La representación se integra a la red neuronal convolucional VGG19 modificada para procesar la entrada RGB + LBP. En la evaluación experimental, el Modelo 4C alcanzó una exactitud global de 98.33 %, superando al modelo de referencia basado únicamente en RGB. Asimismo, desde la perspectiva PAD, obtuvo APCER de 2.36 %, BPCER de 0.97 % y ACER de 1.67 %, lo que evidencia una reducción de errores tanto en seguridad como en usabilidad.

En conjunto, la tesis aporta tres elementos principales: la construcción del conjunto CIOHNA, un método híbrido para la localización automática de la oreja y el Modelo 4C para la diferenciación entre orejas naturales y artificiales.

ABSTRACT

The ear possesses distinctive characteristics that have contributed to its growing relevance over recent decades. Among its principal advantages are morphological stability over time and the possibility to capture without physical contact. For this reason, it is used as a biometric feature in person recognition systems. However, like other biometric modalities, ear-based systems can be vulnerable to presentation attacks (PAs), especially given the growing availability of 3D printing techniques and generative models. In this context, the present thesis proposes a computational model to differentiate between natural and artificial human ears through images in controlled and uncontrolled scenarios.

As part of the research, the Natural and Artificial Human Ear Image Dataset (CIOHNA) was created, comprising natural samples obtained from four public datasets and self-acquired captures, as well as artificial samples generated from drawings, 3D prints, synthetic images, and images produced using artificial intelligence tools.

The proposed solution is organized into two main stages. The first involves ear localization. To achieve this, a hybrid approach is integrated that combines MediaPipe, used to identify facial landmarks, with the YOLOv5s model trained to detect human ears. The dimensions of the ear are estimated from facial landmarks based on the anthropometric proportions proposed by Leonardo da Vinci, while YOLOv5s serves as a complementary mechanism when facial landmarks are insufficient or cannot be properly located. This stage was evaluated using images from public datasets and our own captures, considering both controlled and uncontrolled conditions, achieving an Accuracy 97.00 %, Precision of 99.00 %, Recall of 96.00 %, and an F1-score of 97.00 %.

The second stage addresses the classification between natural and artificial human ears as a strategy for detecting presentation attacks. To this end, the 4C Model is proposed, based on a four-channel representation composed of the RGB image and an additional texture channel generated using Local Binary Patterns (LBP). The representation is integrated into the VGG19 convolutional neural network, modified to process the RGB + LBP input. In the experimental evaluation, the 4C Model achieved an overall accuracy of 98.33%, outperforming the baseline model based solely on RGB. Furthermore, from the PAD perspective, it achieved an APCER of 2.36%, a BPCER of 0.97%, and an ACER of 1.67%, demonstrating a reduction in errors in both safety and usability.

Overall, the thesis contributes three main elements: the construction of the CIOHNA dataset, a hybrid method for automatic ear localization, and the 4C Model for differentiating between natural and artificial ears.

CONTENIDO

RESUMEN	I
ABSTRACT	II
LISTA DE TABLAS	VII
LISTA DE FIGURAS	VIII
LISTA DE ECUACIONES	X
GLOSARIO	XI
CAPÍTULO 1 INTRODUCCIÓN	1
1.1. PLANTEAMIENTO DEL PROBLEMA	1
1.1.1. Análisis	1
1.1.2. Complejidad del problema	2
1.2. DESCRIPCIÓN DE SOLUCIÓN	2
1.3. OBJETIVOS.....	3
1.3.1. General.....	3
1.3.2. Específicos.....	3
1.4. ALCANCES Y LIMITACIONES.....	3
1.4.1. Alcances	3
1.4.2. Limitaciones	4
1.5. JUSTIFICACIÓN.....	4
1.6. ESTRUCTURA DE LA TESIS	4
CAPÍTULO 2 MARCO TEÓRICO	6
2.1. BIOMETRÍA	6
2.1.1. Modalidades de los sistemas biométricos	6
2.1.2. Anatomía de la oreja	8
2.2. DETECCIÓN DE ATAQUES DE PRESENTACIÓN (PAD).....	9
2.2.1. Definiciones Clave	9
2.3. FUNDAMENTOS DE LOCALIZACIÓN Y DETECCIÓN DE OBJETOS	10
2.3.1. Proporciones del cuerpo humano.....	10
2.3.2. MediaPipe	12
2.3.3. YOLOv5s	12
2.4. PATRONES BINARIOS LOCALES (LBP)	13
2.5. ARQUITECTURA VGG19	14
2.6. MÉTRICAS DE EVALUACIÓN	15
2.6.1. Matriz de confusión	15

2.6.2.	Métricas de clasificación	16
2.6.2.1.	Exactitud (Accuracy)	16
2.6.2.2.	Precisión (Precision).....	16
2.6.2.3.	Sensibilidad (Recall o Sensitividad)	16
2.6.2.4.	F1-Score	16
2.6.3.	Métricas PAD (estándar en anti-spoofing)	17
2.6.3.1.	APCER (Attack Presentation Classification Error Rate)	17
2.6.3.2.	BPCER (Bona Fide Presentation Classification Error Rate).....	17
2.6.3.3.	ACER (Average Classification Error Rate)	17
2.6.3.4.	Métricas complementarias reportadas en la literatura: FGR, FFR y HTER.....	18
CAPÍTULO 3	ESTADO DEL ARTE	19
3.1.	TÉCNICAS DE LOCALIZACIÓN	19
3.2.	RECONOCIMIENTO DE PERSONAS MEDIANTE LA OREJA	22
3.2.1.	Técnicas tradicionales	23
3.2.2.	Arquitecturas de aprendizaje profundo (Deep Learning)	25
3.2.3.	Modelos híbridos	28
3.3.	SISTEMAS DE DETECCIÓN DE ATAQUES DE PRESENTACIÓN (PAD/ANTI-SPOOFING)	30
3.3.1.	PAD en biometría	31
3.3.2.	PAD en la oreja humana	31
CAPÍTULO 4	PROPUESTA DE SOLUCIÓN	36
4.1.	DESCRIPCIÓN GENERAL	36
4.1.1.	Localización y Segmentación (Enfoque Híbrido)	37
4.1.1.1.	Localización de la oreja para entornos no controlados	39
4.1.1.2.	Localización de la oreja para entornos controlados	40
4.1.2.	Extracción de características (Fusión RGB + LBP).....	41
4.1.2.1.	Transformación 3C a 4C	42
4.1.3.	Estructura del Modelo 4C.....	43
4.1.4.	Cambios en la red VGG19.....	44
4.2.	CONFIGURACIÓN EXPERIMENTAL	45
4.2.1.	Datos y partición	45
4.2.2.	Preprocesamiento y generación del cuarto canal (LBP).....	45
4.2.3.	Configuración de entrenamiento del clasificador	45
4.2.4.	Entorno de ejecución	45
4.3.	HERRAMIENTAS DE DESARROLLO	45
CAPÍTULO 5	CONJUNTO DE IMÁGENES.....	47
5.1.	MUESTRAS ARTIFICIALES - INSTRUMENTOS DE ATAQUE	47

5.1.1.	Dibujo	47
5.1.1.1.	Internet	49
5.1.1.2.	Generador de Imágenes (Image Generator)	49
5.1.2.	Impresión 3D: Filamento Rígido y Silicón	49
5.1.2.1.	Filamento Rígido	51
5.1.2.2.	Silicón	51
5.1.3.	Imágenes sintéticas con apariencia realista	51
5.1.3.1.	MetaHuman Creator	52
5.1.3.2.	DALL-E	53
5.1.4.	Imágenes hiperrealistas generadas con Meta IA Generativa	54
5.2.	OREJAS HUMANAS NATURALES	55
5.2.1.	Análisis Matemático de Imágenes (AMI)	56
5.2.2.	EarVN1.0.....	56
5.2.3.	IIT Delhi	57
5.2.4.	ITWE (In-the-Wild-Ear)	58
5.2.5.	Cámara web	59
CAPÍTULO 6	PRUEBAS Y RESULTADOS	61
6.1.	AMBIENTE DE PRUEBAS	61
6.2.	EVALUACIÓN DEL MÉTODO DE LOCALIZACIÓN.....	61
6.2.1.	Entrenamiento de YOLOv5s	62
6.2.2.	Comparativa de localización de orejas humanas	64
6.2.3.	Método de localización propuesto.....	66
6.3.	EVALUACIÓN EXPERIMENTAL DEL MODELO 4C.....	69
6.3.1.	Diseño experimental	69
6.3.1.1.	Conjunto de imágenes de entrenamiento	69
6.3.1.2.	Conjunto de imágenes de validación	70
6.3.1.3.	Entrada evaluada: comparación entre 3C y 4C.....	71
6.3.1.4.	Hiperparámetros y configuración experimental.....	71
6.3.2.	Evaluación de desempeño con entrada 3C y 4C	72
6.3.2.1.	Integración final del Modelo 4C.....	72
6.3.3.	Análisis de errores mediante matriz de confusión	73
6.3.4.	Evaluación del Modelo 4C mediante métricas PAD actuales	73
6.3.5.	Análisis visual del aprendizaje del modelo	75
6.3.5.1.	Mapas de características	75
6.3.5.2.	Mapas de activación mediante Grad-CAM	76
6.3.6.	Evaluaciones parciales del Modelo 4C	76
6.4.	COMPARATIVA MEDIANTE MÉTRICAS PAD HISTÓRICAS	79
6.5.	DISCUSIÓN DE LOS RESULTADOS	80

CAPÍTULO 7	CONCLUSIONES	81
7.1.	CONCLUSIONES GENERALES.....	81
7.2.	OBJETIVOS ALCANZADOS	82
7.3.	PRODUCTOS ACADÉMICOS OBTENIDOS	83
7.4.	ACTIVIDADES ACADÉMICAS REALIZADAS	83
7.5.	TRABAJO FUTURO	86
REFERENCIAS		87

LISTA DE TABLAS

TABLA 2.1 COMPARATIVA DE SISTEMAS BIOMÉTRICOS PRESENTADA POR KAMBOJ ET AL. (2022A)	7
TABLA 2.2 MATRIZ DE CONFUSIÓN PARA EVALUAR LA CARACTERIZACIÓN DE LA OREJA EN LA IMAGEN SIGUIENDO EL ESQUEMA DE KOHAVI Y PROVOST (1998).	15
TABLA 3.1 GRUPOS Y TÉCNICAS DE LOCALIZACIÓN DE OBJETOS (CHAVES ET AL., 2018)	20
TABLA 3.2 ESTADO DEL ARTE DE TÉCNICAS DE LOCALIZACIÓN DEL ÁREA DE LA OREJA (2020–2025).	21
TABLA 3.3. ESTADO DEL ARTE DE TÉCNICAS TRADICIONALES PARA EL RECONOCIMIENTO DE PERSONAS MEDIANTE LA OREJA.	24
TABLA 3.4. ESTADO DEL ARTE DE ARQUITECTURAS DE APRENDIZAJE PROFUNDO (DEEP LEARNING) PARA EL RECONOCIMIENTO DE PERSONAS MEDIANTE LA OREJA.	26
TABLA 3.5. ESTADO DEL ARTE DE ARQUITECTURAS DE MODELOS HÍBRIDOS PARA EL RECONOCIMIENTO DE PERSONAS MEDIANTE LA OREJA.	29
TABLA 3.6. TÉCNICAS PARA LA DETECCIÓN DE ATAQUES DE SUPLANTACIÓN (ANTI-SPOOFING)	34
TABLA 5.1 COMPENDIO DEL CONJUNTO IMÁGENES DE OREJAS HUMANAS ARTIFICIALES	55
TABLA 5.2 COMPENDIO DEL CONJUNTO IMÁGENES DE OREJAS HUMANAS NATURALES	60
TABLA 6.1 ESPECIFICACIONES DEL EQUIPO DE CÓMPUTO UTILIZADO EN LAS PRUEBAS	61
TABLA 6.2 DISTRIBUCIÓN DE LOS CONJUNTOS DE IMÁGENES PARA EL ENTRENAMIENTO DE YOLOV5S.....	62
TABLA 6.3 RESULTADOS DEL ENTRENAMIENTO DE YOLOV5S.....	63
TABLA 6.4 DISTRIBUCIÓN DE LOS CONJUNTOS DE IMÁGENES UTILIZADOS EN LA VALIDACIÓN	64
TABLA 6.5 RESULTADOS DE EVALUAR POR SEPARADO A MEDIAPIPE Y YOLOV5S PARA IMÁGENES QUE CONTIENEN UNA OREJA HUMANA EN AMBIENTES CONTROLADOS Y NO CONTROLADOS.	64
TABLA 6.6 DISTRIBUCIÓN DEL CONJUNTO DE IMÁGENES UTILIZADO PARA LA VALIDACIÓN DEL MÉTODO PROPUESTO	66
TABLA 6.7 RESULTADOS DE APLICAR EL MÉTODO DE LOCALIZACIÓN PROPUESTO.....	67
TABLA 6.8 DISTRIBUCIÓN DE LOS CONJUNTOS DE IMÁGENES PARA EL ENTRENAMIENTO	70
TABLA 6.9 DISTRIBUCIÓN DE LOS CONJUNTOS DE IMÁGENES PARA LA VALIDACIÓN.....	70
TABLA 6.10 HIPERPARÁMETROS Y CONFIGURACIÓN ADICIONAL UTILIZADOS EN LA EXPERIMENTACIÓN	71
TABLA 6.11 COMPARATIVA DE LOS RESULTADOS DE DESEMPEÑO ENTRE EL MODELO 3C Y 4C	72
TABLA 6.12 COMPARACIÓN DEL DESEMPEÑO ENTRE EL MODELO DE REFERENCIA (RGB) Y EL MODELO PROPUESTO (4C) SEGÚN MÉTRICAS PAD ACTUALES	74
TABLA 6.13 DISTRIBUCIÓN DE LOS CONJUNTOS EVALUADOS EN LAS PRUEBAS PARCIALES DEL MODELO 4C...	77
TABLA 6.14 RESULTADO DE LAS MÉTRICAS APLICADAS A LAS PRUEBAS PARCIALES.....	78
TABLA 6.15 COMPARATIVA DEL MODELO PROPUESTO CON EL ESTADO DEL ARTE MEDIANTE MÉTRICAS PAD HISTÓRICAS	79
TABLA 7.1 TABLA DE EVIDENCIA DE LOS LOGROS SOBRE OBJETIVOS	82

LISTA DE FIGURAS

FIGURA 1.1 ESQUEMA GENERAL PARA EL DESARROLLO Y EVALUACIÓN DEL MODELO PROPUESTO.....	2
FIGURA 2.1 IMÁGENES DE LOS SISTEMAS BIOMÉTRICOS TOMADAS DE (KAMBOJ ET AL., 2022A).....	8
FIGURA 2.2 ESTRUCTURA DE LA OREJA HUMANA (HASSABALLAH ET AL., 2019).....	8
FIGURA 2.3 REPRESENTACIÓN DE UNA OREJA HUMANA A) NATURAL B) ARTIFICIAL.....	10
FIGURA 2.4 PROPORCIONES FACIALES UTILIZADAS PARA LA ESTIMACIÓN DE LA REGIÓN DE LA OREJA BASADAS EN LOS ESTUDIOS ANATÓMICOS DE LEONARDO DA VINCI (1490): A) BOCETO ANATÓMICO, B) ROSTRO FRONTAL Y C) ROSTRO LATERAL.....	11
FIGURA 2.5 UBICACIÓN DE SEIS PUNTOS DEL ROSTRO (MEDIAPIPE, 2022).....	12
FIGURA 2.6 MALLA FACIAL (MEDIAPIPE, 2022).....	12
FIGURA 2.7 ESQUEMA DE SISTEMA DE DETECCIÓN DE YOLO (REDMON ET AL., 2016).....	12
FIGURA 2.8. PROCESO DEL DESCRIPTOR LBP A) VISTA DE LA ECUACIÓN Y B) CONVERSIÓN DE BINARIO A DECIMAL.	14
FIGURA 2.9 ESQUEMA DE LA ARQUITECTURA VGG19 (CONFIGURACIÓN E) (SIMONYAN & ZISSERMAN, 2015).	14
FIGURA 3.1 REPRESENTACIÓN DE LAS ETAPAS DE RECONOCIMIENTO MEDIANTE LA OREJA.....	23
FIGURA 3.2. TAXONOMÍA DE LA BIOMETRÍA DE LA OREJA (KAMBOJ ET AL., 2022A).....	28
FIGURA 3.3. TIPOS DE SUPLANTACIÓN PARA ATAQUE O ENGAÑO A) IMAGEN EN VIVO, B) IMAGEN IMPRESA, C) REPRODUCCIÓN DE VIDEO (TOPRAK & TOYGAR, 2020).....	32
FIGURA 3.4. ESTRUCTURA DEL MODELO ANTI-SPOOFING (TOPRAK & TOYGAR, 2020).....	32
FIGURA 3.5. DIAGRAMA GENERAL DEL MODELO PROPUESTO (TOPRAK & TOYGAR, 2021).....	33
FIGURA 4.1 FLUJO DEL MODELO PROPUESTO PARA LA IDENTIFICACIÓN DE OREJAS HUMANAS NATURALES Y ARTIFICIALES MEDIANTE EXTRACCIÓN DE TEXTURA LBP Y LA RED VGG19.....	36
FIGURA 4.2 DIAGRAMA PARA LA LOCALIZACIÓN DEL ÁREA DE INTERÉS.....	38
FIGURA 4.3. ILUSTRACIÓN DEL PROCESO DE LOCALIZACIÓN DE LA OREJA PARA ENTORNOS NO CONTROLADOS MEDIANTE MEDIAPIPE.....	39
FIGURA 4.4 LOCALIZACIÓN DE PUNTOS DE REFERENCIA FACIALES PARA LA ESTIMACIÓN DE LA ORIENTACIÓN DEL ROSTRO A) REPRESENTACIÓN DEL TRAGO, B) ROSTRO EN VISTA FRONTAL Y C) PERFIL DEL ROSTRO A ANALIZAR.	40
FIGURA 4.5 LOCALIZACIÓN DE LA REGIÓN DE INTERÉS PARA LA OREJA UTILIZANDO PUNTOS DE REFERENCIA FACIALES C) MALLA FACIAL, B) PUNTOS DE REFERENCIA PARA LOCALIZAR LA OREJA, C) PUNTOS LOCALIZADOS EN EL ROSTRO.....	40
FIGURA 4.6 ESQUEMA DE ENTRENAMIENTO Y LOCALIZACIÓN DE OREJAS HUMANAS MEDIANTE YOLOv5s.....	41
FIGURA 4.7 REPRESENTACIÓN DEL RESULTADO DE LA TRANSFORMACIÓN DE LA IMAGEN 3C A 4C.....	43
FIGURA 4.8 FLUJO DEL MODELO 4C PROPUESTO PARA LA CARACTERIZACIÓN Y CLASIFICACIÓN DE OREJAS HUMANAS NATURALES Y ARTIFICIALES MEDIANTE UNA REPRESENTACIÓN DE CUATRO CANALES (RGB + LBP) Y LA RED VGG19.....	44
FIGURA 5.1 EJEMPLOS DE LA DISTRIBUCIÓN DE LAS IMÁGENES DE OREJAS HUMANAS EN DIBUJO.....	48
FIGURA 5.2 EJEMPLOS DE IMÁGENES DE OREJAS HUMANAS DESCARGADAS DESDE INTERNET.....	49
FIGURA 5.3 EJEMPLOS DE IMÁGENES DE OREJAS HUMANAS ARTIFICIALES GENERADAS POR IA (OPENAI, 2024).	49
FIGURA 5.4 EJEMPLOS DE IMÁGENES DE OREJAS HUMANAS EN IMPRESIÓN 3D: FILAMENTO RÍGIDO Y SILICÓN.....	50

FIGURA 5.5 EJEMPLOS DE IMÁGENES EN IMPRESIÓN 3D CON FILAMENTO RÍGIDO	51
FIGURA 5.6 EJEMPLOS DE IMÁGENES EN IMPRESIÓN 3D CON SILICÓN	51
FIGURA 5.7 EJEMPLOS DE IMÁGENES DE OREJAS HUMANAS REALISTAS GENERADAS POR IA (EPIC GAMES, 2024).	52
FIGURA 5.8 EJEMPLOS DE IMÁGENES DE OREJAS HUMANAS REALISTAS GENERADAS CON METAHUMAN CREATOR.(EPIC GAMES, 2024).....	53
FIGURA 5.9 EJEMPLOS DE IMÁGENES SINTÉTICAS DE OREJAS HUMANAS CON APARIENCIA REALISTA GENERADAS MEDIANTE DALL-E EN CHATGPT (OPENAI, 2024).	53
FIGURA 5.10 EJEMPLOS DE IMÁGENES DE OREJAS HUMANAS HIPERREALISTAS CREADAS CON META IA GENERATIVA DE WHATSAPP (META, 2024).....	54
FIGURA 5.11 EJEMPLOS DE IMÁGENES DEL CONJUNTO AMI.	56
FIGURA 5.12 EJEMPLOS DE IMÁGENES DE OREJAS HUMANAS NATURALES DEL CONJUNTO EARVN1.0 (TRUONG HOANG, 2019).	57
FIGURA 5.13 EJEMPLOS DE IMÁGENES DEL CONJUNTO IITD.	58
FIGURA 5.14 EJEMPLOS DE IMÁGENES DE OREJAS HUMANAS NATURALES DEL CONJUNTO ITWE.	58
FIGURA 5.15 EJEMPLOS DE IMÁGENES TOMADAS CON LA CÁMARA WEB.....	59
FIGURA 6.1 MUESTRAS DEL CONJUNTO DE IMÁGENES DE ENTRENAMIENTO YOLOV5S A) OREJA HUMANA Y B) NO OREJA HUMANA.....	62
FIGURA 6.2 MUESTRAS DE LA VALIDACIÓN DENTRO DE ENTRENAMIENTO DE YOLOV5S.	63
FIGURA 6.3 MUESTRAS DEL CONJUNTO DE IMÁGENES PARA LA VALIDACIÓN A) OREJA HUMANA AMBIENTE CONTROLADO B) OREJA HUMANA AMBIENTE NO CONTROLADO	64
FIGURA 6.4 MUESTRAS DE IMÁGENES EN LA EVALUACIÓN EN AMBIENTE NO CONTROLADO A) MEDIAPIPE LOCALIZA LA OREJA B) MEDIAPIPE NO LOCALIZA LA OREJA, C) YOLOV5S LOCALIZA LA OREJA Y D) YOLOV5S NO LOCALIZA OREJA.	65
FIGURA 6.5 MUESTRAS DE IMÁGENES EN LA EVALUACIÓN EN AMBIENTE CONTROLADO A) MEDIAPIPE NO LOCALIZA LA OREJA, B) YOLOV5S NO LOCALIZA LA OREJA Y C) YOLOV5S NO LOCALIZA LA OREJA.....	65
FIGURA 6.6 MUESTRAS DEL RESULTADO DE LA LOCALIZACIÓN DE LA OREJA HUMANA A) MEDIAPIPE Y B) YOLOV5S.....	66
FIGURA 6.7 MUESTRAS DEL CONJUNTO DE IMÁGENES PARA LA VALIDACIÓN A) OREJA HUMANA AMBIENTE CONTROLADO B) OREJA HUMANA AMBIENTE NO CONTROLADO Y C) NO OREJA HUMANA.....	67
FIGURA 6.8 MUESTRAS DEL RESULTADO DE LA EVALUACIÓN DEL MÉTODO PROPUESTO PARA LA LOCALIZACIÓN DEL ÁREA DE INTERÉS A) OBJETO LOCALIZADO CORRECTAMENTE B) IMÁGENES ESTABLECIDAS COMO NO OREJA.	68
FIGURA 6.9 MATRIZ DE CONFUSIÓN DEL MODELO PROPUESTO PARA LA CLASIFICACIÓN DE OREJAS NATURALES Y ARTIFICIALES.	73
FIGURA 6.10 VISUALIZACIÓN DE MAPAS DE CARACTERÍSTICAS DE LA PRIMERA CAPA CONVOLUCIONAL. (A) MODELO 3C Y (B) MODELO 4C.....	75
FIGURA 6.11 VISUALIZACIÓN DE MAPAS DE ACTIVACIÓN MEDIANTE GRAD-CAM A) MODELO 3C Y B) MODELO 4C.	76

LISTA DE ECUACIONES

(ECUACIÓN 2.1)..... 13

(ECUACIÓN 2.2)..... 13

(ECUACIÓN 2.3)..... 16

(ECUACIÓN 2.4)..... 16

(ECUACIÓN 2.5)..... 16

(ECUACIÓN 2.6)..... 17

(ECUACIÓN 2.7)..... 17

(ECUACIÓN 2.8)..... 17

(ECUACIÓN 2.9)..... 17

(ECUACIÓN 2.10)..... 18

(ECUACIÓN 2.11)..... 18

(ECUACIÓN 2.12)..... 18

(ECUACIÓN 4.1)..... 42

(ECUACIÓN 4.2)..... 42

(ECUACIÓN 4.3)..... 43

GLOSARIO

3C	Tres Canales (RGB)
4C	Cuatro Canales (RGB + imagen resultante de aplicar LBP a la imagen original)
2D	Dos Dimensiones
3D	Tres Dimensiones
ACER	Tasa promedio de error de clasificación (<i>Average Classification Error Rate</i>)
APCER	Tasa de error de clasificación de presentación de ataques (<i>Attack Presentation Classification Error Rate</i>)
BPCER	Tasa de error de clasificación de presentación de buena fe (<i>Bona Fide Presentation Classification Error Rate</i>)
CNN	Red neuronal convolucional (<i>Convolutional Neural Network</i>)
CRR	Tasa de reconocimiento correcta (<i>Correct Recognition Rate</i>)
CENIDET	Centro Nacional de Investigación y Desarrollo Tecnológico
DL	Aprendizaje profundo (<i>Deep learning</i>)
EER	Tasa de Error Igual (<i>Equal Error Rate</i>)
FAR	Tasa de falsa aceptación o falsos positivos (<i>False Acceptance Rate</i>)
FN	Falso Negativo
FP	Falso Positivo
FR	Referencia completa (<i>Full Reference</i>)
FRR	Tasa de falso rechazo o falsos negativos (<i>False Rejection Rate</i>)
HTER	Tasa de error total medio (<i>Half Total Error Rate</i>)
IA	Inteligencia Artificial (<i>Artificial Intelligence</i>)
IQA	Medidas de la Calidad de la Imagen (<i>Image Quality Assessment</i>)
IQM	Evaluación de la Calidad de la Imagen (<i>Image Quality Measures</i>)
JCR	Informes de citas de revistas (<i>Journal Citation Reports</i>)
KNN	K Vecino más cercano (<i>K Nearest Neighbor</i>)
LBP	Patrones binarios locales (<i>Local Binary Patterns</i>)
MI	Información Mutua (<i>Mutual Information</i>)
ML	Aprendizaje máquina (<i>Machine Learning</i>)
N	Negativo
NIST	Instituto Nacional de Estándares y Tecnología (<i>National Institute of Standards and Technology</i>)
NR	No referenciada (<i>No Reference</i>).
P	Positivo
PA	Ataque de Presentación (<i>Presentation Attack</i>)
PAD	Detección de ataques de presentación (<i>Presentation Attack Detection</i>)
PAI	Instrumento de ataque de presentación (<i>Presentation Attack Instrument</i>)
PCA	Análisis de Componentes Principales (<i>Principal Component Analysis</i>)
RGB	Rojo, Verde, Azul (<i>Red, Green, Blue</i>)
ReLU	Unidad Lineal Rectificada (<i>Rectified Linear Unit</i>)
ROI	Región de interés (<i>Region of interest</i>)
SVM	Máquinas de vector soporte (<i>Support Vector Machines</i>)
VN	Verdadero Negativo
VP	Verdadero Positivo

CAPÍTULO 1

INTRODUCCIÓN

La gestión de identidad en los ecosistemas digitales contemporáneos enfrenta el desafío de equilibrar la usabilidad y la robustez de la seguridad. Los mecanismos tradicionales de autenticación, basados en conocimiento (contraseñas) o posesión (tokens), han demostrado ser insuficientes ante la sofisticación de los ataques actuales (Temoshok et al., 2025). En dicho escenario, los sistemas biométricos se consolidan como una alternativa viable para la verificación de identidad, fundamentándose en rasgos fisiológicos o conductuales inherentes al individuo.

Entre las diversas modalidades biométricas, la oreja humana ha adquirido relevancia científica como un rasgo distintivo debido a su estabilidad morfológica, captura sin contacto y baja variabilidad (Kamboj et al., 2022a), (Xu et al., 2023). A diferencia del rostro, cuya morfología puede alterarse por expresiones o envejecimiento a corto plazo, la estructura del oído externo exhibe una estabilidad temporal significativa y una invarianza notable a los gestos, preservando su unicidad incluso en gemelos idénticos (Hassaballah et al., 2019). Las propiedades anteriormente descritas, han motivado el desarrollo de sistemas automáticos de reconocimiento basado en imágenes de la oreja, especialmente en aplicaciones forenses y de control de acceso.

Históricamente fundamentada en los trabajos antropométricos de Bertillon (1890), la biometría de la oreja ha transitado de mediciones manuales a sistemas de visión por computadora avanzados (Hurley et al., 2008), (Anagnostopoulos, 2013), (Kamboj et al., 2021). Sin embargo, la transición de los sistemas desde entornos de laboratorio hacia escenarios operativos reales (*in-the-wild*) ha revelado dos limitaciones sustanciales: la degradación del rendimiento bajo condiciones no controladas (iluminación, oclusión, postura) y, de manera crítica la vulnerabilidad ante Ataques de Presentación (*Presentation Attacks*) en los cuales se emplean fotografías impresas, imágenes digitales, pantallas o prótesis sintéticas para suplantar la identidad de un individuo legítimo.

El auge de tecnologías de impresión 3D y de modelos generativos basados en inteligencia artificial ha reducido significativamente el costo y la complejidad de los ataques, incrementando el riesgo de falsos positivos en sistemas biométricos que no incorporan mecanismos explícitos de detección de vida (*liveness detection*). En consecuencia, la detección de ataques de presentación (PAD) se ha convertido en un requisito indispensable para la viabilidad y despliegue de sistemas biométricos seguros, tal como lo establecen los estándares internacionales ISO/IEC 30107-1 (International Organization for Standardization, 2023a).

1.1. Planteamiento del problema

El problema del presente trabajo de investigación es identificar que la oreja contenida en una imagen corresponde a la de una persona humana y no se trate de un modelo falso realizado en dibujo o impresión 3D (silicón u otro material plástico).

Si bien el estado del arte ha avanzado en la *identificación* de sujetos (quién es) a partir del análisis de imágenes de la oreja, existe información limitada en la validación de la *naturaleza* de la muestra (qué es). Los sistemas existentes, optimizados para extraer características geométricas o de apariencia global, fallan al detectar las micro-texturas y propiedades de reflectancia que distinguen la piel humana viva de materiales sintéticos (papel, pantallas, polímeros), especialmente cuando las imágenes presentan ruido ambiental, desenfoque o variaciones de iluminación propias de capturas con dispositivos móviles. El no contemplar desarrollos de detección de vida (*liveness detection*) compromete la integridad del sistema biométrico, permitiendo accesos no autorizados (falsos positivos) en aplicaciones de seguridad.

1.1.1. Análisis

En la sociedad actual, la autenticación basada en características biométricos se ha convertido en un enfoque confiable y ampliamente utilizado para la identificación en el ámbito forense y civil. La investigación criminal, la identificación de personas desaparecidas, el comercio electrónico, y el control de acceso a dispositivos son algunos ejemplos de sus aplicaciones (Sarangi et al., 2019). La oreja como característica biométrica, en sus inicios no presentó mayor relevancia, debido a las dificultades que se presentaron para adquirir imágenes adecuadas (Anagnostopoulos, 2013).

En la actualidad, lo anterior ha cambiado considerablemente debido a qué se identificaron características distintivas como que es invariante a los estados de ánimo u expresiones faciales, es bajamente afectada por la edad ya que su forma se mantiene igual entre los ocho y setenta años de edad (Wang et al., 2021), (Kamboj et al., 2022a), (Fukuda et al., 2025); además, está cerca de la simetría bilateral (Emeršič et al., 2020). Por lo anterior, se considera un sistema biométrico confiable para el reconocimiento de personas y su uso empieza a extenderse a aplicaciones de seguridad y acceso a lugares y dispositivos.

Sin embargo, para que un sistema biométrico basado en la oreja sea confiable en aplicaciones reales, no basta con reconocer la identidad del individuo. Es imprescindible verificar previamente que la imagen procesada corresponde efectivamente a una oreja humana natural y no un modelo falso realizado en dibujo o impresión 3D (silicón u otro material plástico). Investigaciones referentes a la detección de suplantación de identidad mediante el análisis de imágenes del rostro han crecido en los últimos años (Abdullakutty et al., 2021), (Sharma & Selwal, 2023), (Singh et al., 2023), (Muhammad et al., 2024), (Pooshideh et al., 2024), (Shaheed et al., 2024); no obstante, el detectar la suplantación mediante el análisis de imágenes de la oreja permanece poco explorado en el contexto de la biometría de la oreja, como se evidencia en el análisis del estado del arte. El poco interés es principalmente debido a la complejidad para caracterizar una oreja humana en imágenes que presentan variaciones significativas de iluminación, oclusiones parciales por cabello o accesorios, cambios de escala y rotación, así como ruido inherente a dispositivos móviles de captura.

1.1.2. Complejidad del problema

- La localización del área de interés dentro de la imagen es una etapa difícil ya que, algunas de sus partes pueden estar ocluidas por cabello, accesorios como aretes, audífonos o sombras, lo que dificulta su detección precisa en imágenes capturadas en condiciones reales.
- Considerando que parte de las imágenes fueron adquiridas en un ambiente real y no controlado, presentan variaciones en la escala, iluminación, de perspectiva, rotación y oclusión. Provocando que la extracción de las características consideradas para su descripción no sea consistente en todos los casos.
- La caracterización de una imagen de una persona viva con respecto a una imagen generada es cada vez es compleja.

1.2. Descripción de solución

Para abordar el problema planteado, la tesis siguió una solución compuesta por cinco etapas, las cuales se ilustran en la Figura 1.1.



Figura 1.1 Esquema general para el desarrollo y evaluación del modelo propuesto.

Cada una de las etapas, se describen a continuación de manera general:

- 1. Fundamentos:** análisis teórico y estudio del estado del arte de investigaciones relacionadas con la identificación de una oreja humana natural, que permitió identificar las ventajas y limitaciones de los enfoques existentes.
- 2. Especificación:** selección de las herramientas y técnicas para el desarrollo e implementación del modelo híbrido en la localización y reconocimiento de la oreja.
- 3. Desarrollo:** diseño del modelo, el cual consiste en la combinación de una técnica tradicional de extracción de textura con una red neuronal convolucional. Después de la construcción del modelo híbrido denominado Modelo 4C se realizó la codificación del sistema inteligente. Las aportaciones esperadas con el proyecto son principalmente la

propuesta del modelo que capta la diferencia entre una oreja humana natural sobre una sintética, apoyando con ello a la detección de posibles ataques de suplantación de personas mediante la oreja. Así como, la creación de un banco de imágenes que contiene orejas naturales y artificiales.

4. **Pruebas:** etapa de experimentación del módulo de localización y del modelo de reconocimiento *anti-spoofing* bajo distintos escenarios tales como piel alrededor de la oreja, rostro completo, hombro y cuello; y cambios ligeros de iluminación, escala y rotación a partir de los conjuntos de datos públicos y banco de datos propio.
5. **Evaluación:** verificación de los resultados del modelo propuesto mediante las métricas estándar utilizadas para el reconocimiento reportadas en el estado del arte.

1.3. Objetivos

Establecer el rumbo del proyecto fue punto importante para obtener los resultados planteados al inicio del proyecto de tesis, en los siguientes dos puntos se describen el objetivo general y los específicos que representan el trabajo de tesis.

1.3.1. General

Proponer un modelo computacional para la diferenciación automática entre orejas humanas naturales y artificiales (*anti-spoofing*) en imágenes adquiridas en entornos no controlados, mediante una arquitectura híbrida que fusione características de textura local con representaciones profundas.

1.3.2. Específicos

- Analizar las limitaciones de las técnicas actuales referentes a la caracterización e identificación de la oreja, para seleccionar las adecuadas al problema según su desempeño.
- Implementar redes profundas con los mejores resultados documentados, para validar su desempeño ante escenarios complejos.
- Construir un banco de imágenes de orejas humanas naturales y artificiales adquiridas en ambientes no controlados para seleccionar las imágenes de orejas a procesar.
- Formular el modelo de caracterización mediante la combinación de una técnica tradicional de descripción de textura y una red neuronal de aprendizaje profundo.
- Evaluar el modelo propuesto con las métricas estándar empleadas en el reconocimiento y en la detección de suplantación de orejas.

1.4. Alcances y limitaciones

En el apartado se especifican los alcances y limitaciones establecidos en el proyecto de tesis.

1.4.1. Alcances

- El modelo detecta y reconoce una oreja humana real en un ambiente controlado.
- Los factores considerados que pueden presentarse en un ambiente no controlado son: cambios en la iluminación, rotación, escala, forma y oclusión.
- El modelo fue desarrollado y evaluado utilizando conjuntos de imágenes públicos, así como con el banco de imágenes propio.

- Para evaluar el desempeño del modelo propuesto se utilizaron cuatro métricas clásicas en el área de reconocimiento (Exactitud, Precisión, *Recall* y *F1-score*), tres de suplantación (*APCER*, *BPCER*, *ACER*) detalladas en el Capítulo 2 (Marco Teórico – Métricas de evaluación), para evaluar el desempeño del sistema.

1.4.2. Limitaciones

- No se garantiza la certeza del reconocimiento si la imagen es de baja resolución, con un alto grado de oclusión (por arriba del 50 %) o con demasiados accesorios.
- El tamaño del banco de imágenes propio fue condicionado por la disposición de los usuarios para realizar las sesiones, la disponibilidad del equipo y el lugar para realizarlo.

1.5. Justificación

La integridad y seguridad de las personas es una necesidad que está presente en toda población actualmente, convirtiéndose en un tema de interés para la mayoría de los seres humanos. Como individuos, saber que nuestra integridad o la de un ser querido está en riesgo, pone a las personas en una constante preocupación. Por tal motivo, se busca satisfacer la necesidad mediante diversos mecanismos o herramientas tecnológicas, como la biometría, que es la ciencia enfocada en reconocer a los seres humanos. Los sistemas biométricos constituyen una herramienta importante en áreas como la vigilancia, la ciencia forense y muchos sistemas inteligentes.

La identificación de personas mediante el uso de la oreja es uno de los sistemas biométricos que ha cobrado importancia en los últimos años. Considerando que la detección de la oreja es la etapa inicial para la identificación de personas, es primordial verificar que la imagen de entrada sea una oreja humana real y no un ataque de suplantación mediante un modelo artificial, aun cuando dicha fotografía puede ser adquirida en ambientes poco controlados y con ello tener problemas de ruido, cambios en la intensidad luminosa, tamaño, pose, entre otros. El beneficio que se busca con el proyecto de tesis es proponer una caracterización computacional eficiente de la oreja humana en diferentes condiciones, para poder diferenciarla de figuras semejantes que pueden ser categorizadas como falsas y, con ello apoyar en la detección de suplantación de identidad.

1.6. Estructura de la tesis

El resto del trabajo está dividido en seis capítulos descritos como:

Capítulo 2. Marco teórico. Presenta las principales definiciones para el entendimiento del trabajo desarrollado. Se describe en términos generales la biométrica y las modalidades existentes, se conceptualiza la anatomía de la oreja humana como sistema biométrico y se establecen sus dimensiones mediante las proporciones del cuerpo humano. Además, se fundamentan las bases teóricas sobre descriptores de textura (como LBP y sus variantes), arquitecturas de aprendizaje profundo (redes neuronales convolucionales) y las herramientas de visión por computadora utilizadas para la implementación del sistema. Se introduce el término ataque de presentación (PAD). Asimismo, se definen las métricas de evaluación seleccionadas para medir el desempeño del modelo.

Capítulo 3. Estado del arte. Análisis de la literatura consultada y relacionada con el tema de investigación dentro de los últimos cinco años. El capítulo se estructura en tres núcleos temáticos: (1) Técnicas de localización de la región de interés, (2) Métodos de reconocimiento de personas mediante la oreja donde se revisan descriptores de textura para la extracción de características y (3) Sistemas de detección de ataques de suplantación (*anti-spoofing*) facial y de oreja.

Capítulo 4. Metodología propuesta. Presenta la contribución principal de la tesis: el diseño y desarrollo del 'Modelo 4C'. Se detalla la arquitectura de la propuesta metodológica desglosando sus dos etapas fundamentales: (1) Localización Robusta de la Región de Interés (ROI) y (2) Clasificación mediante Fusión de Características (Modelo 4C). Asimismo, se describen las herramientas de desarrollo implementadas.

Capítulo 5. Conjunto de imágenes. Describe la conformación y limpieza/normalización de los bancos de imágenes empleados. Se detallan las características del conjunto propio (muestras artificiales de impresión 3D en filamento y silicón) y de los conjuntos naturales (bases de datos públicas y recopilaciones web).

Capítulo 6. Pruebas y resultados. Establece el diseño experimental para la validación del modelo. Se definen cuatro casos de estudio para evaluar el método de localización y se detalla la validación del Modelo 4C. Se presentan los resultados cuantitativos obtenidos, incluyendo una comparativa con el estado del arte frente a ataques de suplantación, finalizando con una discusión sobre los resultados del sistema propuesto.

Capítulo 7. Conclusiones. Sintetiza los logros alcanzados en relación con los objetivos planteados. Se enumeran las aportaciones, los productos académicos derivados de la investigación y se proponen líneas de trabajo futuro para la continuidad del proyecto.

CAPÍTULO 2

MARCO TEÓRICO

El presente capítulo establece los fundamentos teóricos y conceptuales necesarios para abordar el problema de la detección de ataques de presentación en biometría de la oreja. Se describen las bases anatómicas de la oreja humana, los principios de la detección de objetos y análisis de textura, así como las arquitecturas de aprendizaje profundo que sustentan la propuesta metodológica de la investigación.

2.1. Biometría

Es una ciencia interdisciplinaria que estudia los métodos para identificar o verificar la identidad de una persona en función de sus rasgos físicos o de comportamiento (Bansal & Madasu, 2021). Es una herramienta utilizada en áreas como la seguridad, el control de acceso, la autenticación en dispositivos móviles, la banca electrónica y la ciencia forense, ya que proporcionan una alternativa más robusta frente a contraseñas o tarjetas que pueden perderse, olvidarse o falsificarse (Hassaballah et al., 2019), (Shaheed et al., 2024). La biometría aprovecha características anatómicas o conductuales que en general son únicas, permanentes y difíciles de replicar lo que permiten establecer correspondencias fiables entre un sujeto y una identidad registrada (Bertillon, 1890), (Kamboj et al., 2022a), (Kamboj et al., 2021), (Fukuda et al., 2025).

2.1.1. Modalidades de los sistemas biométricos

Las modalidades biométricas pueden clasificarse, de manera general, en fisiológicas y conductuales. Las primeras se basan en rasgos físicos relativamente estables del cuerpo humano, como la huella dactilar, el rostro, el iris, la retina, la geometría de la mano y la oreja. Las segundas se fundamentan en patrones de comportamiento, tales como la voz, la firma manuscrita o la dinámica de tecleo.

Entre las modalidades más empleadas se encuentran el rostro, la huella dactilar, el iris, la palma de la mano, los nudillos (Kamboj et al., 2022a) (Kamboj et al., 2021) (Pooshideh et al., 2024) (Shaheed et al., 2024) (Fukuda et al., 2025) y también la oreja Kamboj et al. (2022a), (Youbi et al., 2019). En la Tabla 2.1, Kamboj et al. (2022a) presentan una comparativa de las modalidades biométricas relevantes, tomando en cuenta si hay contacto con el sensor que captura la imagen a procesar, el conjunto de información que se busca, si es necesaria la cooperación del usuario, las ventajas y los retos a que se enfrenta cada uno de ellos.

Tabla 2.1 Comparativa de sistemas biométricos presentada por Kamboj et al. (2022a).

Biométrico	Sensor	Características	Cooperación	Ventajas	Desafíos
Rostro	Sin contacto	Espaciado de ojos, nariz, contorno de labios, orejas, cabello, textura de la piel.	Baja	Capacidad para operar de forma encubierta.	Máscara, expresión facial, anteojos, cosméticos, gemelos idénticos, cirugía, envejecimiento.
Huella dactilar	Contacto	Puntos singulares, minucias, crestas, pliegues y cicatrices.	Alta	Bajo costo, fácil de usar, más comercial.	Condición de la piel, intrusivo, dedo artificial, no apto para personas que trabajan en la industria laboral.
Iris	Sin contacto	Criptas, corona, cuello en zigzag, ligamentos arqueados.	Alta	Resistencia al impostor.	Sensor caro, desenfoque de movimiento, iluminación, intrusivo, condiciones de salud.
Palma mano	Contacto	Líneas, pliegues y crestas en la palma de la mano.	Alta	Sensor de bajo costo, no intrusivo.	Variación de condiciones de iluminación escala y traslación.
Nudillos	Sin contacto	Líneas de nudillos, pliegues.	Baja	Sensor barato, único, bien protegido.	Intrusivo
Oreja	Sin contacto	Hélix, lóbulo, trago, concha.	Baja	No intrusivo, útil en situaciones como COVID-19, poco afectado por la edad	Accesorios en la oreja, cabello, condiciones de iluminación.

La Tabla 2.1 sintetiza, desde una perspectiva comparativa, que ningún rasgo biométrico puede considerarse universal, ya que cada modalidad implica compromisos entre tipo de sensor (contacto/sin contacto), nivel de cooperación del usuario, ventajas operativas y limitaciones en condiciones reales. En particular, el rostro permite operar de manera encubierta, pero es vulnerable a factores como mascarilla, expresiones, gafas/cosméticos, envejecimiento e incluso escenarios como gemelos idénticos; el iris destaca por su alta singularidad y resistencia al impostor, aunque requiere alta cooperación, sensores potencialmente costosos y es sensible a desenfoque por movimiento, iluminación e incluso condiciones de salud; la huella dactilar es una de las opciones más comerciales por su bajo costo, pero su desempeño y uso se ven condicionados por el estado de la piel, es intrusivo y ataques con dedos artificiales. Asimismo, nudillos y palma ofrecen alternativas de costo relativamente bajo, pero presentan restricciones de ser intrusivo (nudillos) o variaciones de adquisición (palma: iluminación, escala y traslación). La oreja se posiciona como un biométrico sin contacto y de baja cooperación, relevante por su naturaleza no intrusiva, su utilidad en contextos donde se busca reducir contacto (p. ej., COVID-19) y, de forma destacada, porque puede adquirirse aun cuando el rostro está cubierto por mascarilla, mientras que sus retos se concentran en oclusiones por accesorios/cabello y condiciones de iluminación. En la Figura 2.1 se muestran imágenes de los sistemas biométricos que se describen en el trabajo de Kamboj et al. (2022a).

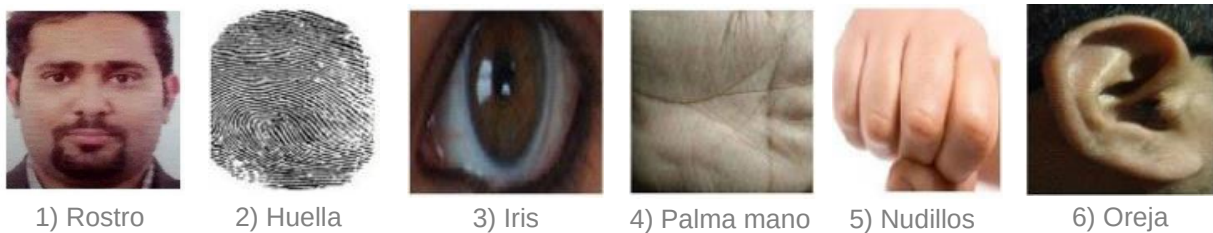


Figura 2.1 Imágenes de los sistemas biométricos tomadas de (Kamboj et al., 2022a).

Los sistemas que han tenido mayor uso para el reconocimiento de personas en los últimos años son: el rostro, la huella dactilar y la oreja. Siendo los que mayores desafíos han tenido que resolver. En ese panorama, la oreja destaca por su morfología estable en el tiempo, captura sin contacto y menor afectación por expresiones o maquillaje; los retos en la práctica son la oclusión por cabello/accesorios, la pose y la iluminación, especialmente en ambientes no controlados (Bertillon, 1890), (Raveane et al., 2019), (Kamboj et al., 2022b), (Kamboj et al., 2022a), (Kamboj et al., 2021), (Fukuda et al., 2025). Las conclusiones han sido reiteradas en encuestas y revisiones de 2022–2024, así como en trabajos aplicados que muestran desempeños competitivos con CNN y técnicas de mejora de imagen.

2.1.2. Anatomía de la oreja

La oreja humana como modalidad biométrica es competente y distintiva, la cual contiene información estable y confiable, así como una estructura que no muestra cambios drásticos con el paso del tiempo, captura sin contacto y menor afectación por expresiones o maquillaje. Si bien la estructura externa es relativamente simple, la variabilidad entre dos orejas es lo suficientemente distinguible incluso para gemelos idénticos (Hassaballah et al., 2019). Estudios anatómicos y biométricos han demostrado que la configuración geométrica de las regiones es distintiva entre individuos y su estructura externa se mantiene similar con el paso de los años. La parte externa está formada por el borde de la hélice y el lóbulo. La parte interna tiene características prominentes como el antehélix, la incisura intertrágica, la concha, la fosa triangular, el pilar del hélix y el trago (Kamboj et al., 2022a), (Kamboj et al., 2021). La Figura 2.2 muestra la estructura de la oreja y sus distintos componentes.

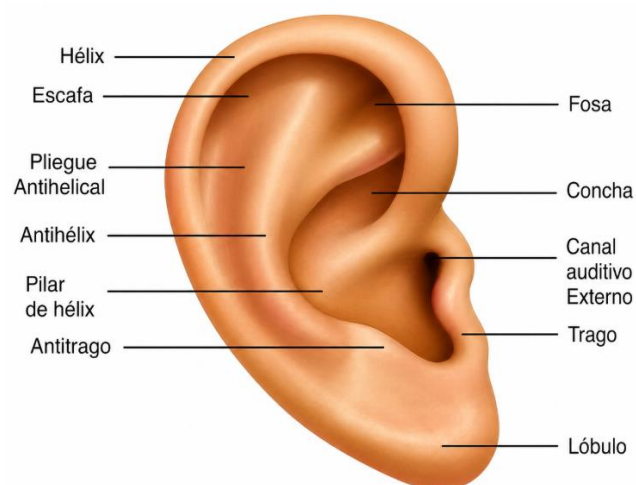


Figura 2.2 Estructura de la oreja humana (Hassaballah et al., 2019).

La oreja humana es una estructura cartilaginosa recubierta de piel, cuya morfología se define durante el desarrollo embrionario y permanece prácticamente invariable desde la infancia hasta la senectud, salvo por un crecimiento escalar predecible debido a la gravedad y la pérdida de elasticidad del colágeno (Abaza et al., 2013). Las características discriminativas principales, ilustradas en la literatura clásica (Abaza et al., 2013), incluyen:

- **Hélix y Antehélix:** Los bordes exteriores e interiores curvos.
- **Trago y Antitrigo:** Prolongaciones cartilaginosas que protegen el canal auditivo.
- **Concha:** La depresión central profunda.
- **Lóbulo:** La parte inferior carnosa.

Desde la perspectiva de la visión por computadora, la profundidad y las sombras proyectadas por las estructuras (especialmente la concha y el hélix) son críticas para distinguir una oreja volumétrica real de una representación plana (fotografía 2D), hecho que constituye la base física para la detección de ataques de impresión.

2.2. Detección de Ataques de Presentación (PAD)

La norma ISO/IEC 30107-1 (2023a) define el marco estandarizado para la evaluación de vulnerabilidades en sistemas biométricos.

2.2.1. Definiciones Clave

- **Ataque de Presentación (*Presentation Attack, PA*):** Presentación al subsistema de captura biométrica de un instrumento con la intención de interferir en su operación normal (suplantación o evasión).
- **Instrumento de Ataque de Presentación (*Presentation Attack Instrument, PAI*):** Objeto o mecanismo utilizado para el ataque. En el contexto auricular, los PAI más comunes son fotografías impresas (papel mate/brillante), pantallas electrónicas (ataques de repetición) y máscaras sintéticas.
- **Muestra *Bona Fide*:** Interacción biométrica originada por un sujeto vivo y legítimo, sin intención de ataque.

Aun cuando los sistemas de reconocimiento de personas mediante los rasgos biométricos brindan ventajas en la vida cotidiana, la falta de seguridad de los sistemas puede desencadenar resultados desastrosos. Las personas (intrusos) pueden intentar acceder a los sistemas de seguridad aplicando técnicas ilegales. Es decir, si el impostor intenta tener acceso a la cuenta de otra persona, a lo que se le conoce como ataque de suplantación de identidad (Toprak & Toygar, 2021). Toprak & Toygar (2021) afirman que los ataques más comunes son la fotografía impresa, la fotografía digital, la reproducción de video, la máscara y la cirugía plástica. El proyecto se enfoca principalmente en reconocer una oreja humana natural sobre una oreja en dibujo o impresión 3D (silicón u otro material plástico). En la Figura 2.3, se muestra un ejemplo de una oreja humana natural y una hecha de silicona.

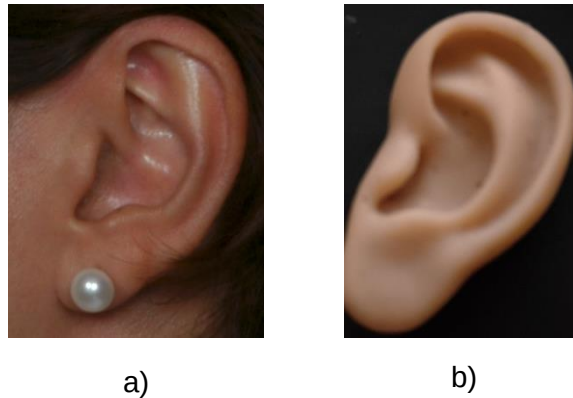


Figura 2.3 Representación de una oreja humana a) natural b) artificial.

La adopción masiva de biometría obliga a considerar la suplantación (*Presentation Attack Detection, PAD*). Las revisiones de 2024–2025 insisten en integrar PAD desde el diseño (detección de fotos/impresiones/moldes, verificación de vivacidad y calidad, pérdidas y arquitecturas específicas), tanto para rasgos faciales como perioculares; las pautas son aplicables a oreja y a sistemas multimodales que la incluyan. La seguridad frente a ataques de presentación (PAD) —fotografías, vídeo, máscaras o artefactos impresos/silicona— es un foco central. Revisiones recientes analizan *datasets* de imágenes de orejas, protocolos y técnicas (p. ej., análisis de textura, fusión multimodal, aprendizaje profundo), y subrayan retos de generalización entre dominios (Pooshideh et al., 2024). La modalidad de la oreja adquiere relevancia por tres razones:

1. Por su morfología estable y su captura sin contacto, ofrece un rasgo difícil de alterar con expresiones o gestos intencionados.
2. Permite una vía adicional de autenticación complementaria a otras modalidades (rostro, huella) en sistemas biométricos multimodales, lo que mejora la resistencia a ataques.
3. Pero precisamente su auge implica que los sistemas que usan oreja deben integrar mecanismos PAD específicos: detección de moldes, detección de dispositivos impresos, verificación de que la captura es real y no una imagen bidimensional, control de la iluminación, etc.

En conjunto, dichos elementos conforman la base conceptual y técnica del sistema propuesto orientado a reconocimiento ante intentos de suplantación.

2.3. Fundamentos de Localización y Detección de Objetos

Para el procesamiento de imágenes en entornos no controlados (*in-the-wild*), es imperativo aislar la Región de Interés (ROI) antes de la clasificación.

2.3.1. Proporciones del cuerpo humano

Para especificar las dimensiones de la oreja humana, se consideró lo propuesto por Leonardo da Vinci, en su obra en el *Hombre de Vitruvio* (*Uomo vitruviano*) (da Vinci, 1490), estudio que menciona que la oreja es una tercera parte del rostro y es proporcional a la distancia que va de la parte inferior de la barbilla hasta la nariz.

De igual forma, la distancia entre el nacimiento del cabello y las cejas mantiene dicha relación (da Vinci, 1883) (Losardo et al., 2015) (Nicholson, 2019) como se muestra en la Figura 2.4.

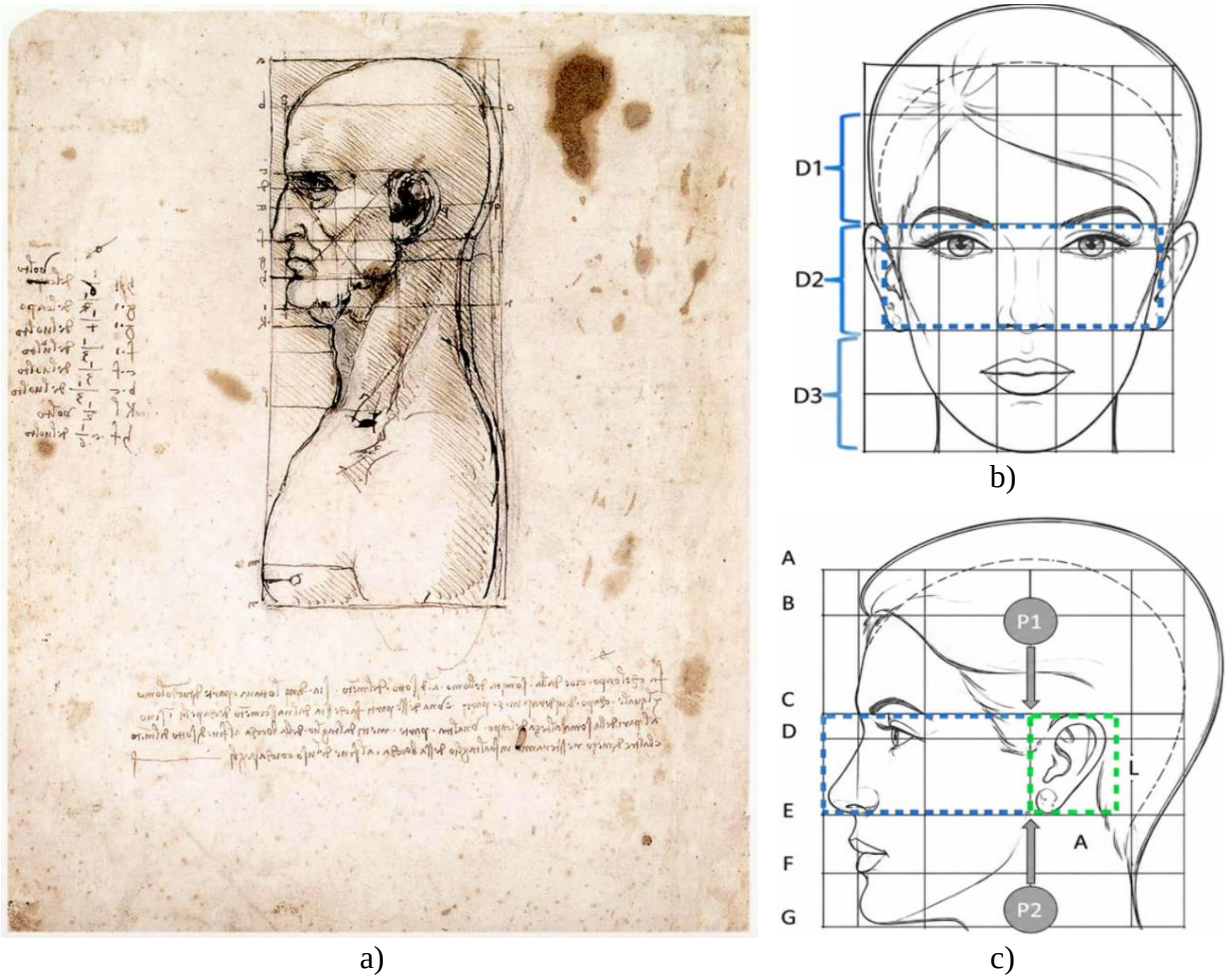


Figura 2.4 Proporciones faciales utilizadas para la estimación de la región de la oreja basadas en los estudios anatómicos de Leonardo da Vinci (1490): a) boceto anatómico, b) rostro frontal y c) rostro lateral.

La imagen muestra las proporciones del rostro humano basadas en principios clásicos atribuidos a Leonardo da Vinci (1490). En el inciso a) se presenta un boceto anatómico realizado por Leonardo da Vinci que ilustra la cabeza humana en vista lateral, donde se analizan las proporciones y relaciones geométricas entre diferentes partes del rostro.

En los incisos b) y c) muestran una adaptación de estos principios para el análisis facial. En b) se observa la vista frontal del rostro dividida en tres regiones verticales principales: desde la línea del cabello hasta las cejas (D1), de las cejas a la base de la nariz (D2) y de la base de la nariz al mentón (D3). El rectángulo azul punteado indica la zona central del rostro, que se extiende aproximadamente desde las cejas hasta la base de la nariz.

Por su parte, en c) se presenta la vista lateral del rostro, donde se indican las referencias geométricas utilizadas para estimar la posición de la oreja. El recuadro verde señala la posición y altura aproximada de la oreja, la cual se encuentra generalmente alineada entre la línea de las cejas y la base de la nariz. Los puntos P1 y P2 indican referencias verticales

utilizadas para estimar la altura de la oreja (L) dentro de las proporciones faciales. Estas referencias permiten definir una región de interés para la localización de la oreja dentro del rostro.

2.3.2. MediaPipe

Es una herramienta que apoya el reconocimiento de rostros mediante la localización de diferentes puntos en el rostro detallando ojos, nariz, boca, entre otros. MediaPipe ofrece soluciones de aprendizaje automático personalizables y multiplataforma de código abierto (MediaPipe, 2022). MediaPipe contiene dos formas de localizar el rostro: mediante la ubicación de seis puntos estratégicos y mediante una malla facial. La primera sólo ubica seis partes del rostro, como son: ojos (dos puntos), tragos (dos puntos) nariz y boca (Véase Figura 2.5), mientras que la segunda forma coloca 477 puntos alrededor y/o sobre cada una de las partes que componen el rostro generando así una malla facial (Véase Figura 2.6).

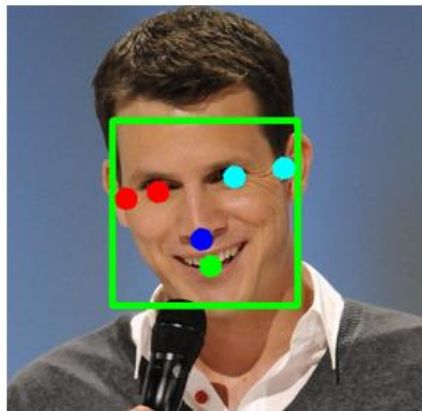


Figura 2.5 Ubicación de seis puntos del rostro (MediaPipe, 2022).

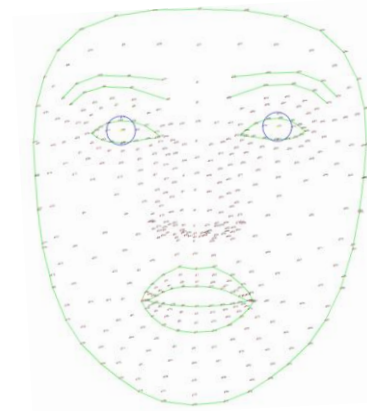


Figura 2.6 Malla facial (MediaPipe, 2022).

2.3.3. YOLOv5s

YOLO (You Only Look Once) introdujo un enfoque unificado para la detección de objetos al plantear la detección como un problema de regresión hacia cuadros delimitadores (*bounding boxes*) y probabilidades de clase, de modo que una sola red neuronal predice directamente dichas salidas a partir de la imagen completa, permitiendo un flujo extremo a extremo (Redmon et al., 2016). Fue creada por Joseph Redmond como principal autor (Redmon et al., 2016), consiste en una red neuronal convolucional que predice simultáneamente múltiples cuadros delimitadores y las probabilidades de la clase de objeto que delimitan los cuadros (ver Figura 2.7).

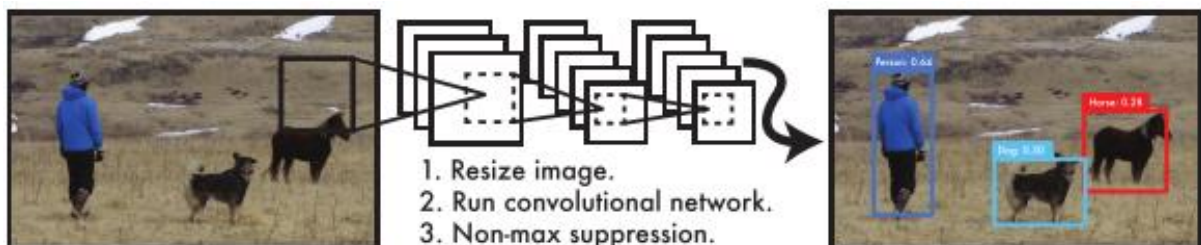


Figura 2.7 Esquema de sistema de detección de YOLO (Redmon et al., 2016).

YOLOv5s (Small) es una de las variantes más ligeras de la familia YOLOv5, desarrollada por *Ultralytics* (2023). Su arquitectura se organiza en tres bloques principales: (i) *backbone*, basado en CSPDarknet53 para la extracción de características; (ii) *neck*, que utiliza SPPF (*Spatial Pyramid Pooling–Fast*) y PANet (*Path Aggregation Network*) para fusionar información multi-escala; y (iii) *head*, responsable de generar las predicciones finales.

En este proceso, los datos fluyen en forma de tensores, definidos como estructuras de datos que generalizan el concepto de matrices a n dimensiones (Goodfellow et al., 2016). En YOLOv5s, estos tensores finales contienen la información necesaria (coordenadas, confianza y clase) para identificar los objetos en la imagen de forma simultánea en una sola pasada por la red. La arquitectura ha demostrado ser eficiente para la detección de objetos por lo que es utilizada en diferentes proyectos de investigación (Li et al., 2021), (Dai et al., 2022), (Qin et al., 2022), (Lei et al., 2023).

2.4. Patrones Binarios Locales (LBP)

El descriptor LBP (*Local Binary Patterns*) es un descriptor de textura local que extrae características de una imagen a partir de la relación de intensidad entre un píxel central y sus vecinos. Su formulación general fue consolidada por Ojala et al. (2002), quienes definieron un vecindario compuesto por P puntos de muestreo distribuidos de manera uniforme sobre un círculo de radio R alrededor del píxel central g_c . Para cada píxel vecino cuya intensidad es g_p ($p=0, 1, \dots, P-1$), se evalúa la diferencia $g_p - g_c$ mediante una función umbral, generando así una secuencia binaria que posteriormente se transforma a un valor decimal ponderado por potencias de dos. De tal forma que el descriptor LBP se define como lo indica la Ecuación 2.1.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (\text{Ecuación 2.1})$$

Donde:

- $LBP_{P,R}$ representa el valor decimal del patrón en el píxel central.
- g_c es la intensidad de gris del píxel central.
- g_p representa la intensidad de los P píxeles vecinos.
- p indexa la posición del vecino sobre el círculo
- 2^p pondera la contribución binaria para formar el código decimal.
- P es el número de puntos de muestreo del vecindario.
- R es el radio del vecindario (en píxeles).
- $s(g_p - g_c)$ es la función umbral definida por la Ecuación 2.2.

$$s(g_p - g_c) = \begin{cases} 1 & \text{si } g_p - g_c \geq 0 \\ 0 & \text{si } g_p - g_c < 0 \end{cases} \quad (\text{Ecuación 2.2})$$

Donde $s(x)$ toma el valor de 1 cuando $g_p - g_c \geq 0$ y 0 en caso contrario.

En la formulación propuesta por Ojala et al. (2002), los vecinos se describen como una cadena circular indexada alrededor del píxel central. Esto significa que cada vecino recibe un índice p dentro de una secuencia ordenada sobre la circunferencia, y dicho índice determina el valor binario 2^p utilizado en la conversión decimal. Por ello, el valor final del código LBP depende de la de indexación adoptada, la condición es que dicha convención se mantenga constante

durante todo el proceso de extracción. La Figura 2.8 La figura ilustra el proceso general de obtención del descriptor LBP.

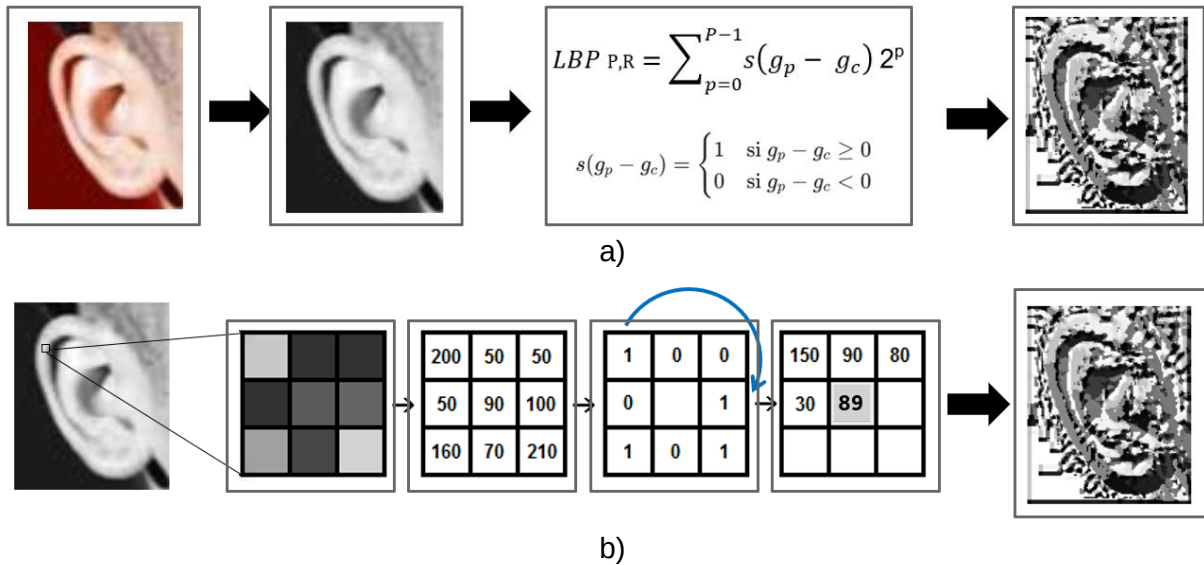


Figura 2.8. Proceso del descriptor LBP a) vista de la ecuación y b) conversión de binario a decimal.

En la imagen inciso a) se observa el flujo general: la imagen original se convierte a escala de grises, se aplica la ecuación del operador LBP y se genera una nueva imagen codificada por patrones locales. En el inciso b) se presenta un ejemplo a partir de un vecindario de 3x3, se comparan los valores de los píxeles vecinos con el valor central, se obtiene un patrón binario y luego se transforma en un valor decimal.

2.5. Arquitectura VGG19

En 2015 en la Universidad de Oxford Simonyan y Zisserman presentaron su trabajo titulado “Redes convolucionales muy profundas para el reconocimiento de imágenes a gran escala”, dando paso a la red VGG19 (Simonyan & Zisserman, 2015). La arquitectura VGG19 recibe imágenes de $224 \times 224 \times 3$ y está compuesta por 16 capas convolucionales organizadas en cinco bloques, seguidas de tres capas totalmente conectadas para la clasificación final (ver Figura 2.9).

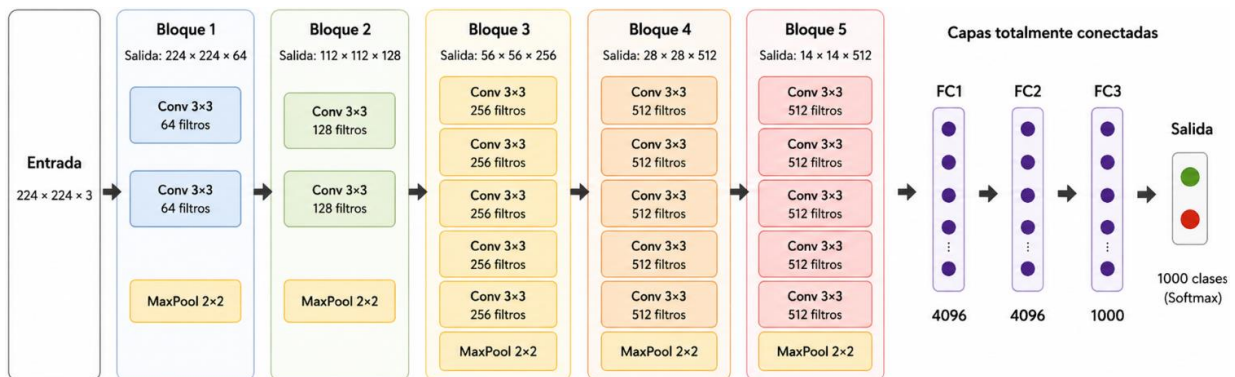


Figura 2.9 Esquema de la arquitectura VGG19 (Configuración E) (Simonyan & Zisserman, 2015).

VGG enfatiza el desarrollo en términos de mayor profundidad de las redes, fue diseñada con varias configuraciones con tamaños entre 11 y 19 capas, aunque las versiones de mejor rendimiento tenían 16 o más capas (Aggarwal, 2018). Una innovación importante de VGG es que redujo el tamaño de los filtros, pero aumentó la profundidad.

VGG19 se caracteriza por su profundidad (19 capas con pesos) y el uso exclusivo de filtros de convolución pequeños de 3x3. La elección de filtros pequeños es fundamental teóricamente: una pila de dos capas de 3x3 tiene el mismo campo receptivo efectivo que una de 5x5, pero con menos parámetros y más no-linealidades (funciones de activación ReLU). Dicha propiedad permite a VGG19 modelar texturas complejas con gran precisión espacial, lo que la hace idónea para tareas donde el detalle fino (textura de la piel) es más relevante que la semántica global (Simonyan & Zisserman, 2015).

2.6. Métricas de evaluación

Las métricas de evaluación son aquellas que permiten identificar qué tan bueno y fiable es el desempeño del modelo evaluado. Mediante ellas se puede observar qué tan hábil es el modelo para hacer la clasificación. Para la evaluación de la experimentación se utilizaron cuatro métricas clásicas de clasificación (descritas en el punto 2.6.2) y tres métricas de detección de ataques de presentación (PAD, descritas en el punto 2.6.3).

2.6.1. Matriz de confusión

La matriz de confusión muestra las clasificaciones predichas y reales, tiene un tamaño de $\ell \times \ell$, donde ℓ es el número de valores de etiquetas diferentes (Kohavi & Provost, 1998). La Tabla 2.2 presenta la matriz de confusión utilizada para evaluar la caracterización de la oreja en la imagen. La matriz permite contrastar la clase real con la clase predicha por el modelo, identificando los casos de verdadero positivo, falso positivo, falso negativo y verdadero negativo

Tabla 2.2 Matriz de confusión para evaluar la caracterización de la oreja en la imagen siguiendo el esquema de Kohavi y Provost (1998).

Predicción Real / Actual		N/F	P/V
		Artificial	Natural
N/F	Artificial	Verdadero negativo (VN)	Falso positivo (FP)
P/V	Natural	Falso negativo (FN)	Verdadero positivo (VP)

Donde:

- Verdadero Negativo (**VN**): el modelo predice correctamente las muestras negativas, para el caso del trabajo de tesis son las orejas artificiales.
- Verdadero Positivo (**VP**): el modelo predice correctamente las muestras positivas, para el caso del trabajo de tesis son las orejas naturales.
- Falso Positivo (**FP**): el modelo predice incorrectamente las muestras negativas, para el caso del trabajo de tesis son las orejas naturales.
- Falso Negativo (**FN**): el modelo predice incorrectamente las muestras positivas, para el caso del trabajo de tesis son las orejas artificiales.

- Positivas (**P**): Total de muestras positivas, para el caso del trabajo de tesis son las orejas naturales.
- Negativas (**N**): Total de muestras negativas, para el caso del trabajo de tesis son las orejas artificiales.

La terminología actual de verdadero positivo, falso positivo, falso negativo y verdadero negativo puede interpretarse como una adaptación del lenguaje original de la teoría de detección de señales al análisis de clasificadores automáticos (Kohavi & Provost, 1998). A partir de la estructura es posible analizar el comportamiento del clasificador y derivar métricas de desempeño que complementan la exactitud global del sistema

2.6.2. Métricas de clasificación

Para evaluar el desempeño de los modelos de clasificación, se utilizan métricas clásicas derivadas de la matriz de confusión. Las métricas consideradas en el proyecto de tesis son Exactitud, Precisión, Sensibilidad y *F1-score*, las cuáles permiten cuantificar el desempeño del modelo para diferenciar entre una oreja humana natural y una artificial.

2.6.2.1. Exactitud (Accuracy)

Representa el porcentaje de casos identificados de manera correcta sobre el total de casos considerados. En el contexto del estudio, indica la tasa de coincidencias correctamente aceptadas y rechazadas (tanto oreja natural como artificial) en todo el modelo (Qin et al., 2022), su valor se obtiene mediante la Ecuación 2.3.

$$Accuracy = \frac{VP + VN}{P + N} \quad (\text{Ecuación 2.3})$$

2.6.2.2. Precisión (Precision)

Indica la fiabilidad del modelo cuando predice una clase positiva. Es el porcentaje de casos positivos que fueron identificados correctamente respecto al total de casos que el modelo etiquetó como positivos. El valor se obtiene mediante la Ecuación 2.4.

$$Precision = \frac{VP}{VP + FP} \quad (\text{Ecuación 2.4})$$

2.6.2.3. Sensibilidad (Recall o Sensitividad)

Mide la capacidad del modelo para encontrar todos los casos positivos reales. Es el porcentaje de casos positivos correctamente identificados respecto a *todas* las muestras que realmente son positivas en el conjunto de datos. Es dado por la Ecuación 2.5.

$$Recall = \frac{VP}{VP + FN} \quad (\text{Ecuación 2.5})$$

2.6.2.4. F1-Score

Mide la capacidad del modelo para encontrar todos los casos positivos reales. Es el porcentaje de casos positivos correctamente identificados respecto a *todas* las muestras que realmente son positivas en el conjunto de datos. se obtiene mediante la Ecuación 2.6.

$$F1 - Score = \frac{2 * Precision * Recall}{Presicion + Recall} \quad (Ecuación 2.6)$$

2.6.3. Métricas PAD (estándar en *anti-spoofing*)

Para cuantificar el desempeño de un sistema de Detección de Ataques de Presentación (PAD), se utilizan las métricas estandarizadas en la norma ISO/IEC 30107-3 (2023b). A diferencia de las métricas binarias clásicas, las métricas PAD se enfocan específicamente en la vulnerabilidad del sistema ante ataques de suplantación (*spoofing*) y en la usabilidad para usuarios legítimos. A continuación, se detallan dichas métricas con sus respectivas formulaciones:

2.6.3.1. APCER (Attack Presentation Classification Error Rate)

Se define como la proporción de ataques de presentación que son incorrectamente clasificados como muestras *bona fide* (auténticas). En términos de seguridad, representa la tasa de falsos positivos de vida (el sistema clasifica una oreja artificial o ataque como muestra natural o *bona fide*). Su cálculo se establece mediante la Ecuación 2.7:

$$APCER = \frac{N_{PA_err}}{N_{PA}} \quad (Ecuación 2.7)$$

Donde:

- N_{PA_err} : Número de ataques de presentación clasificados erróneamente como *bona fide*.
- N_{PA} : Número total de ataques de presentación evaluados.

2.6.3.2. BPCER (Bona Fide Presentation Classification Error Rate)

Se define como la proporción de muestras *bona fide* (genuinas) que son incorrectamente clasificadas como ataques. En términos de usabilidad, representa la tasa de falsos negativos de vida (el sistema clasifica una oreja natural o *bona fide* como una muestra artificial o de ataque). Su valor se obtiene mediante la Ecuación 2.8:

$$BPCER = \frac{N_{BF_err}}{N_{BF}} \quad (Ecuación 2.8)$$

Donde:

- N_{BF_err} : Número de presentaciones *bona fide* clasificadas erróneamente como ataques.
- N_{BF} : Número total de presentaciones *bona fide* evaluadas.

2.6.3.3. ACER (Average Classification Error Rate)

Es el promedio aritmético de las tasas de error *APCER* y *BPCER*. La métrica *ACER* se utiliza para resumir el rendimiento general del sistema en un único valor, equilibrando la seguridad y la usabilidad. Se calcula mediante la Ecuación 2.9 (ISO, 2023):

$$ACER = \frac{APCER + BPCER}{2} \quad (Ecuación 2.9)$$

Donde:

- *ACER* es su promedio aritmético y resume el desempeño del sistema.

- *APCER* y *BPCER* son las tasas definidas en las Ecuaciones 2.7 y 2.8.

2.6.3.4. Métricas complementarias reportadas en la literatura: *FGR*, *FFR* y *HTER*

Además de las métricas PAD estandarizadas en ISO/IEC 30107-3 (2023b), algunos trabajos de anti-spoofing biométrico reportan métricas complementarias basadas en la clasificación binaria entre muestras reales y falsas. Entre ellas se encuentran *FGR* (*False Genuine Rate*), *FFR* (*False Fake Rate*) y *HTER* (*Half Total Error Rate*). Estas métricas han sido utilizadas en evaluaciones previas de anti-spoofing para expresar los errores cuando una muestra falsa se clasifica como real o cuando una muestra real se clasifica como falsa (Toprak & Toygar, 2020), (Toprak & Toygar, 2021).

La métrica *FGR* indica la proporción de muestras falsas clasificadas incorrectamente como genuinas (Ecuación 2.10):

$$FGR = \frac{N_{falsa \rightarrow genuina}}{N_{falsa}} \quad (\text{Ecuación 2.10})$$

Donde:

- $N_{falsa \rightarrow genuina}$: número de muestras artificiales clasificadas como naturales.
- N_{falsa} : número total de muestras artificiales evaluadas.

La métrica *FFR* indica la proporción de muestras genuinas clasificadas incorrectamente como falsas (Ecuación 2.11):

$$FFR = \frac{N_{genuina \rightarrow falsa}}{N_{genuina}} \quad (\text{Ecuación 2.11})$$

Donde:

- $N_{genuina \rightarrow falsa}$: número de muestras naturales clasificadas como artificiales.
- N_{falsa} : número total de muestras naturales evaluadas.

Finalmente, *HTER* resume el error promedio del sistema mediante el promedio aritmético de *FGR* y *FFR* (Ecuación 2.12):

$$HTER = \frac{FGR + FFR}{2} \quad (\text{Ecuación 2.12})$$

En este capítulo se presentaron los fundamentos teóricos de la investigación: biometría de la oreja, ataques de presentación, localización de la región de interés, extracción de características mediante LBP, arquitectura VGG19. También se describieron las métricas clásicas (*Exactitud*, *Precisión*, *Recall* y *F1-score*), PAD y complementarias utilizadas en el análisis, como *APCER*, *BPCER*, *ACER*, *FGR*, *FFR* y *HTER*.

Con ello, el capítulo establece la base conceptual necesaria para comprender la construcción del conjunto de imágenes, la extracción de características mediante la representación 4C y la evaluación experimental del modelo propuesto en los capítulos posteriores.

CAPÍTULO 3

ESTADO DEL ARTE

El reconocimiento biométrico de personas se ha investigado intensamente debido a la necesidad de sistemas de autenticación más seguros, eficientes y confiables a través de la combinación de la computación, óptica, acústica, biosensores y bioestadística u otros medios de alta tecnología (Wang et al., 2021). El reconocimiento mediante imágenes de la oreja es una autenticación biométrica que tuvo sus inicios en 1890 con el criminólogo francés A. Bertillon (1890). El sistema biométrico de la oreja ha tenido su mayor importancia en la ciencia forense, al ser la oreja una parte del cuerpo humano que mejor se preserva. Sin embargo, la transición de los sistemas desde entornos controlados hacia escenarios reales (*in-the-wild*) ha puesto de manifiesto limitaciones críticas relacionadas con la localización automática del área de interés, la robustez del reconocimiento bajo condiciones no controladas y, de forma más reciente, la vulnerabilidad frente a ataques de suplantación (*Presentation Attacks*).

El capítulo presenta una revisión del estado del arte, organizada en tres núcleos temáticos fundamentales:

1. Técnicas de localización de la oreja,
2. Reconocimiento de personas mediante la oreja, y
3. Detección de ataques de presentación (PAD/*anti-spoofing*).

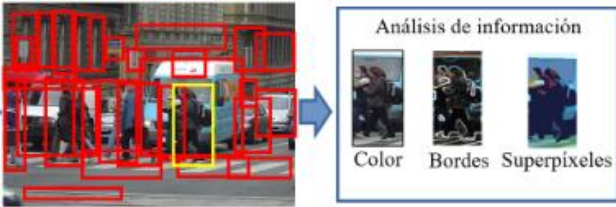
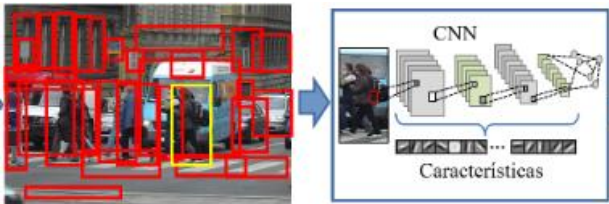
El análisis permite identificar las metodologías actuales, establecer las técnicas utilizadas, los alcances logrados y las limitaciones presentes.

3.1. Técnicas de localización

La localización precisa de un objeto que aparece en una imagen es requerida en muchas aplicaciones, como en la medicina para el diagnóstico clínico asistido por computadora, en la industria para inspección visual, para la detección de obstáculos en vehículos o en robots, entre otros (Chaves et al., 2018).

En el año 2018 se presentó un trabajo donde se realizó una revisión sistemática de los métodos para la localización de objetos, los cuales dividieron en tres grupos: métodos basados en ventanas deslizantes, conjuntos de regiones candidatas y aprendizaje profundo (Chaves et al., 2018). En la Tabla 3.1 se muestra información de los grupos y técnicas de localización que utilizan dichos grupos.

Tabla 3.1 Grupos y técnicas de localización de objetos (Chaves et al., 2018)

Grupo	Descripción	Técnicas de localización
Ventanas Deslizantes	Se recorre la imagen en búsqueda del objeto de interés utilizando una ventana de tamaño variable.  Extracción de características Clasificador	V.J.: Viola and Jones
		HOG+SVM
		DPM: Modelo de parte deformable (<i>Deformable Part Model</i>)
Conjunto de regiones candidatas	Genera un conjunto de regiones de interés, bajo el criterio que los objetos comparten características semejantes y que los diferencian del fondo.  Análisis de información Color Bordes Superpíxeles	Objetividad (<i>Objectness</i>)
		Búsqueda selectiva (<i>Selective Search</i>)
		Cajas de borde (<i>Edge Boxes</i>)
Aprendizaje profundo (<i>Deep Learning</i>)	Aprenden directamente de los datos considerando diferentes niveles de abstracción, en especial las Redes Neuronales Convolucionales (CNN, del inglés <i>Convolutional Neural Network</i>). Se enfoca en el análisis de imágenes, mostrando un buen rendimiento en problemas de reconocimiento, detección y segmentación de objetos.  CNN Características	RCNN (<i>Region-based Convolutional Neural Networks</i>)
		Fast-RCNN (<i>Fast Region-based Convolutional Neural Networks</i>)
		Faster-RCNN (<i>Faster Region-based Convolutional Neural Networks</i>)
		Mask-RCNN (<i>Mask Region-based Convolutional Neural Networks</i>)
		YOLO (<i>You Only Look Once</i>)
		SSD (<i>Single Shot Detector</i>)

Como se observa en la Tabla 3.1, las técnicas de localización de objetos han evolucionado desde métodos tradicionales basados en el análisis de la imagen mediante ventanas deslizantes, hacia enfoques que emplean regiones candidatas y, posteriormente, modelos basados en aprendizaje profundo.

Con el objetivo de analizar el estado del arte en el reconocimiento biométrico basado en la oreja, se realizó una revisión de trabajos publicados en el periodo 2020–2025, enfocándose particularmente en la forma en que abordan la localización del área de interés. En la Tabla 3.2 se presenta una síntesis de los artículos revisados, indicando el tipo de localización empleada (cuando existe) o, en su defecto, muestras de los conjuntos utilizados cuando las imágenes ya contienen la región de la oreja previamente segmentada. Asimismo, se incluye información de cada trabajo en términos de metodología, modelo utilizado, bases de datos de evaluación y métricas reportadas.

Tabla 3.2 Estado del arte de técnicas de localización del área de la oreja (2020–2025).

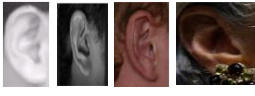
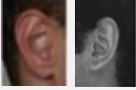
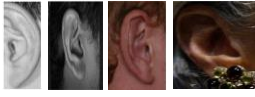
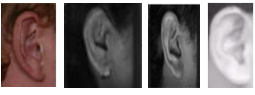
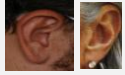
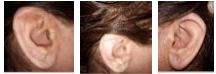
#	Referencia	Localización	Observaciones
1	(Ganapathi et al., 2020)		Utiliza estadísticas geométricas en nubes de puntos 3D basándose en variaciones de superficie. Evaluado en UND (Colección J2) alcanzando un desempeño de Rank-1 de 98.84 %.
2	(Sivanarain & Viriri, 2020)		Propone la fusión de descriptores locales (FLD) combinando DTP y LBP. Se aplican directamente sobre recortes de las bases de datos. Evaluado en IITD usando Rank-1, alcanzando una tasa máxima de 97.44 %.
3	(Alagarsamy & Murugan, 2020)	Adaptive Runge-Kutta (AARK)	Utiliza umbralización adaptativa AARK (<i>Adaptive Approach Runge-Kutta</i>) para determinar límites de la oreja, seguido de un clasificador CART (<i>Classification and Regression Tree</i>). Evaluado en IITD (98.42 %) y AMI (92.54 %).
4	(Sajadi & Fathi, 2020)		Combina características locales LPQ (<i>Local Phase Quantization</i>) y globales (<i>Gabor-Zernike</i>). Evaluado en USTB, IITD, AMI y AWE usando Exactitud, alcanzando 100 % ambiente controlado y 57.20 % no controlado.
5	(Toprak & Toygar, 2020)		Sistema de <i>Anti-spoofing</i> basado en medidas de calidad de imagen IQA (<i>Image Quality Assessment</i>) para detectar ataques de impresión. Evaluado en AMI y UBEAR, logrando exactitudes de 98.17 % y 96.48 % respectivamente.
6	(Hassaballah et al., 2020)		Propone el descriptor RLOP para extraer estructuras locales binarizadas con invarianza a la rotación. Utilizan imágenes previamente recortadas de los datasets. Evaluado en IITD, AMI y AWE usando Rank-1, alcanzando 99.70 % (controlado) y 51.01 % (no controlado).
7	(Alkababji & Mohammed, 2021)	Faster R-CNN	Sistema de reconocimiento basado en una red de extracción de funciones seguida de dos subredes de detección. Evaluado en la base AMI utilizando Exactitud, alcanzando un 98.05 % en condiciones controladas.
8	(Kamboj et al., 2021)	EME (<i>Ear Mask Extraction</i>)	Utiliza una red EME para segmentación a nivel de píxel y filtros Curvature Gabor. Evaluado en IITD y AMI usando Exactitud, alcanzando 99.38 % y 98.24 % respectivamente.
9	(Bansal & Madasu, 2021a)		No realizan localización; el algoritmo aplica normalización y filtrado de Hanman directamente sobre la imagen. Evaluado en IITD y USTB usando Exactitud, alcanzando 99.27 % y 98.00 %.
10	(Toprak & Toygar, 2021)		Combina CNN con medidas de calidad de imagen (IQM) para detectar ataques de suplantación. La localización está implícita en el uso de <i>datasets</i> pre-recortados de orejas. Evaluado en AMI, UBEAR, IITD y USTB usando Exactitud, alcanzando 99.50 % (controlado) y 98.11 % (no controlado).
11	(Khaldi & Benzaoui, 2021)		Emplea redes generativas adversarias (DCGAN) para colorear imágenes en escala de grises antes del reconocimiento con CNN. Evaluado en AMI y AWE usando Rank-1, alcanzando 99.16 % (controlado) y 53.33 % (no controlado).
12	(Khiarak, 2021)		Propone un método de detección de ataques de presentación (PAD) mediante aprendizaje por transferencia con redes profundas. Utiliza el <i>dataset WUT-Ear V1.0</i> donde la región de interés ya está localizada. Evaluado bajo métricas de error, alcanzando una Exactitud del 99.83 %.

Tabla 3.2 Estado del arte de técnicas de localización del área de la oreja (2020–2025).
(continuación)

#	Referencia	Localización	Observaciones
13	(Bokade & Kanphade, 2022)		Sistema multimodal que integra rostro, oreja y mano para autenticación. Evaluado en EarVN1.0 y ORL (<i>Olivetti Research Laboratory</i>) y <i>PalmSet</i> , usando Exactitud, alcanzando un 99.72 % en condiciones multimodales.
14	(Kamboj et al., 2022b)	EME (<i>Ear Mask Extraction</i>)	El método propuesto se enfoca en EIQA (<i>Ear Image Quality Assessment</i>) alcanzando 97.46 % de exactitud en la clasificación de calidad de imágenes de oreja.
15	(Ramos-Cooper et al., 2022)	Mask R-CNN	Utiliza la red <i>Mask R-CNN</i> para la segmentación de la oreja en imágenes no restringidas. Implementan la detección a nivel de píxel para aislar la oreja de fondos complejos. Evaluado en <i>VGGFace-Ear</i> y AWE usando Rank-1, alcanzando un 91.1 % y 62.5 % respectivamente.
16	(Ganesan et al., 2023)		Combina matriz de covarianza con aprendizaje profundo. La fase de localización se omite al utilizar nubes de puntos 3D y proyecciones 2D ya normalizadas. Evaluado en UND (J2) usando Rank-1, alcanzando una Exactitud del 99.63 %.
17	(Lei et al., 2023)	YOLOv5s-MG (<i>single-stage detector</i> , <i>MobileNetV3</i> + <i>GhostConv/C3Ghost</i>)	Realizan la localización y clasificación en una sola etapa para mejorar la velocidad de inferencia. Evaluado en EarVN1.0 y USTB usando Exactitud, alcanzando 98.24 % y 99.41 %.
18	(Bouaouina et al., 2024)	Mean-CAM (ROI basada en mapas de activación)	Combina DCGAN y Mean-CAM-CNN para reconocimiento. Utilizan mapas de activación para localizar y extraer la región de interés mediante máscaras. Evaluado en MAI y AWE usando Rank-1, alcanzando 100 % y 76.25 %.
19	(Ganapathi et al., 2024)		Reconocimiento mediante descriptores binarios 3D. Evaluado en UND-J2 alcanzando 98.62 % de exactitud.
20	(Fukuda et al., 2025)	CNN + Grad-CAM	Identificación multi-ángulo con CNN + Grad-CAM. Evaluado en un dataset propio (Multi-angle) usando Rank-1, alcanzando 96.50 %.

La revisión del estado del arte entre 2020 y 2025 revela que solo el 40 % de los estudios (ocho de 20) implementan una etapa explícita de localización, mientras que la mayoría opta por omitir este proceso al emplear conjuntos de datos públicos con imágenes pre-alineadas y recortadas. Esta tendencia subraya una dependencia crítica hacia el pre-procesamiento manual de los *datasets*, los cuales serán descritos a detalle en el capítulo cuatro, y resalta la necesidad de desarrollar algoritmos de segmentación más robustos para aplicaciones en entornos no controlados.

3.2. Reconocimiento de personas mediante la oreja

Generalmente, las etapas que integran el proceso de reconocimiento de personas mediante el sistema biométrico de la oreja (Wang et al., 2021), son cinco:

1. **Detección:** es la base para el reconocimiento mediante la oreja y corresponde a la localización de la misma.
2. **Preprocesamiento:** procedimiento para el mejoramiento de las condiciones originales de la imagen.

3. **Extracción de características:** es la parte más crítica del reconocimiento mediante la oreja, tradicionalmente consiste en localizar puntos clave dentro de la imagen para extraer características, tradicionalmente geométricas.
4. **Módulo de Clasificación:** se subdivide en dos vertientes interdependientes.
 - **El Clasificador:** Representa la arquitectura lógica o el modelo matemático (ej. Redes Neuronales, SVM, o KNN) previamente entrenado con una galería de referencia.
 - **La Clasificación:** Es la ejecución algorítmica donde las características extraídas se comparan contra los patrones almacenados para determinar una medida de similitud.
5. **Decisión:** es la última etapa, donde se interpreta el resultado de la clasificación y se toma una decisión final.

En la Figura 3.1, se muestra una representación del flujo de las etapas que se llevan a cabo para realizar el reconocimiento de personas mediante la imagen de la oreja.



Figura 3.1 Representación de las etapas de reconocimiento mediante la oreja.

Dentro del núcleo temático de reconocimiento de personas mediante la oreja se delimitaron tres tipos de enfoques: Técnicas tradicionales, Arquitecturas de aprendizaje profundo (*Deep learning*) y Modelos híbridos.

3.2.1. Técnicas tradicionales

En el reconocimiento de personas mediante imágenes de la oreja suelen distinguirse tres enfoques: basados en 3D, basados en apariencia (color y textura) y basados en características geométricas (Omara et al., 2016). La taxonomía reciente consolida la clasificación y subraya que, dentro de basados en apariencia, los descriptores de textura (p. ej., LBP y variantes, LPQ, LDP/LTP, Gabor, GLCM/Haralick) siguen siendo relevantes como base o como complemento de métodos profundos, sobre todo en escenarios con datos limitados o variabilidad fuerte de pose/iluminación (Benzaoui et al., 2023).

Los autores Benouis & Regouid (2020) proponen un sistema biométrico basado en una técnica llamada *Shifted 1D-LBP*. En el sistema la imagen es recortada manualmente y se transforma a un tamaño de 50x50 píxeles, después se realiza la extracción de características mediante *Shifted 1D-LBP*, el cual es un algoritmo derivado de la técnica LBP (*Local Binary Pattern*). Para la parte del reconocimiento, se usa el algoritmo KNN (*K Nearest Neighbor*). El sistema utiliza el conjunto USTB, reporta *CRR* (*Correcte Recognition Rate*), *FAR* (*False Accept Rate*)

y *FRR* (*False Reject Rate*) como métricas de evaluación.

Sivanarain & Viriri (2020) presentaron en la conferencia IMITEC “2nd International Multidisciplinary Information Technology and Engineering Conference”, una investigación para el reconocimiento de personas utilizando descriptores de textura locales. El objetivo fue evaluar descriptores de textura locales: patrones binarios locales (LBP), patrones ternarios locales (LTP), patrones direccionales locales (LDP) y patrones ternarios direccionales (DTP) para una identificación humana robusta utilizando la oreja humana. En el estudio, se propuso un enfoque para la fusión de descriptores de textura locales (LBP, LTP, LDP y DTP). Donde la fusión de la técnica DTP con LBP, obtuvo los mejores puntajes.

La Tabla 3.3 sintetiza el estado del arte en torno al reconocimiento de personas mediante la oreja a partir de técnicas tradicionales. En ella se reúnen trabajos representativos, considerando sus objetivos, contribuciones principales, técnicas empleadas, conjuntos de imágenes y métricas reportadas. En este sentido, su función es descriptiva y documental, más que comparativa, ya que los estudios considerados difieren en sus protocolos experimentales, conjuntos de imágenes y criterios de evaluación.

Tabla 3.3. Estado del arte de técnicas tradicionales para el reconocimiento de personas mediante la oreja.

Referencia	Objetivo	Contribución principal	Técnicas	Dataset	Métrica	Resultado (%)
(Alqaralleh & Toygar, 2018)	Superar las limitaciones de los sistemas unimodales.	Modelo que combina el trago y la oreja para el reconocimiento.	LBP, k-NN	USTB	Exactitud	100.00
(Youbi et al., 2019)	Mostrar la robustez y la estabilidad del método propuesto.	Modelo variante del descriptor de LBP llamado MLBP.	Multi-scale LBP	IITD-I	Rank-1	98.40
				IITD-II		98.64
				USTB		98.33
(Ganapathi et al., 2020)	Uso de características locales para reconocimiento 3D.	Detección y descripción de puntos clave mediante estadísticas geométricas.	Matriz de covarianza	UND-J2	Rank-1	98.60
					Rank-2	100.00
					EER	1.50
(Sivanarain & Viriri, 2020)	Identificación mediante textura.	Fusión de descriptores	LBP, LTP, LDP o DTP	IITD-I	Exactitud	95.88
				IITD-II		97.44
(Alagarsamy & Murugan, 2020)	Enfoque automatizado en imágenes 2D.	Segmentación adaptativa (AARK) y clasificador CART.	AARK, DWT, CART	IITD	Exactitud	6.46
					Precisión	97.69
					Recall	95.45
(Sajadi & Fathi, 2020)	Extraer características en el dominio de la frecuencia.	Fusión de características locales y globales.	LPQ, Gabor-Zernike, Algoritmo genético, K-NN.	USTB	Exactitud	100.00
				IITD-I		99.20
				IITD-II		97.13
(Hassaballah et al., 2020)	Proporcionar patrones invariantes de rotación, ruido e iluminación.	Descriptor RLOP (<i>Robust local oriented patterns</i>)	LBP, LDP, LOOP	IITD-I	Exactitud	95.13
				IITD-II		97.98
				AMI		72.29
				AWE		54.10

Tabla 3.3 Estado del arte de técnicas tradicionales para el reconocimiento de personas mediante la oreja (continuación).

Referencia	Objetivo	Contribución principal	Técnicas	Dataset	Métricas	Resultado (%)
(Bansal & Madasu, 2021a)	Modelar certeza/incertidumbre en entornos no controlados.	HAE (<i>Hanman-Anirban Entropy</i>), HT (<i>Hanman Transform</i>), HF (<i>Hanman Filter</i>), WHC (<i>Weighted Hanman Classifier</i>).	HAE, HT, HF, WHC.	IITD-I	Exactitud	97.10
				IITD-II		97.40
(Bokade & Kanphade, 2022)	Sistema multimodal	Integración de rasgos biométricos.	Métodos clásicos.	EarVN1.0, ORL, PalmSet	Exactitud	97.53
(Benzaoui et al., 2023)	Análisis del estado del arte.	Síntesis de descriptores de textura.	LBP/derivados, Gabor, BSIF, LPQ, Haralick.	-----	-----	Revisión teórica
(Oyebiyi et al., 2023)	Revisión sistemática (2013–2022).	Clasificación de métodos locales.	LBP/LPQ/LDP/Gabor/GLCM como “ <i>local methods</i> ”.	10 bases	-----	73 artículos analizadas.
(Ganapathi et al., 2024)	Uso de descriptores binarios 3D para reconocimiento humano mediante la oreja.	Descriptores binarios 3D mediante correntopía.	Descriptor binario 3D.	UND-J2	Exactitud	98.62
					EER	1.54

A partir de la información presentada en la Tabla 3.3 se observa que la métrica más utilizada para evaluar el desempeño de estos sistemas es Exactitud. En condiciones controladas, se reportan valores cercanos o superiores al 97 % en conjuntos de imágenes como IITD-II y IITD-II. De manera similar, algunos trabajos alcanzan valores de hasta 100 % en *datasets* como UND-J2 y USTB-1, lo que muestra la eficacia de estos métodos cuando las imágenes presentan condiciones estables de iluminación, pose y fondo. Sin embargo, cuando los métodos se evalúan en entornos no controlados, el desempeño puede disminuir de forma considerable. Esto se observa en el conjunto de imágenes AWE, donde se reportan Exactitud de 54.10 % (Hassaballah et al., 2020), debido a la presencia de variaciones en iluminación, pose, oclusiones y fondo.

En conjunto, los resultados presentados en la tabla muestran que, aunque los métodos tradicionales siguen siendo efectivos en escenarios controlados, su desempeño puede verse limitado en ambientes no controlados.

3.2.2. Arquitecturas de aprendizaje profundo (*Deep Learning*)

Existen trabajos que emplean redes neuronales profundas (CNN, GAN, Attention, etc.) como núcleo del modelo, sin combinar directamente descriptores clásicos (LBP, Gabor, Haralick, etc.). Se propone, por ejemplo, un método de reconocimiento de oreja humana por fusión de rasgos basado en características de canal y convolución dinámica (CFDCNet); aun cuando es profundo, fusiona canales y realza patrones locales; el trabajo ubica los descriptores de textura

como línea base y motiva su combinación con DL, útil para “textura + deep”. Primero extraen las características de forma adaptativa mediante convolución dinámica, luego se aumentan los pesos mediante un mecanismo de atención de canal lo que les permite resaltar características importantes y suprimir las no validas y finalmente realizan agrupación máxima para reducir el número de parámetros y cálculos (Xu et al., 2023).

La Tabla 3.4 presenta una síntesis de trabajos representativos del estado del arte que emplean técnicas de aprendizaje profundo (*Deep Learning*) para el reconocimiento biométrico de personas mediante imágenes de la oreja. La tabla organiza la información de cada trabajo mediante las siguientes columnas. La columna Referencia indica la fuente bibliográfica del estudio. La columna Objetivo describe el propósito principal del trabajo. La columna Contribución principal resume la propuesta metodológica del estudio. Por su parte, la columna Técnicas especifica los modelos o algoritmos de aprendizaje profundo empleados en cada propuesta. La columna *Dataset* indica los conjuntos de datos utilizados para el entrenamiento o evaluación de los modelos. Finalmente, las columnas Métrica y Resultado presentan los indicadores de evaluación reportados por los autores y los valores obtenidos en los experimentos.

Tabla 3.4. Estado del arte de arquitecturas de aprendizaje profundo (*Deep Learning*) para el reconocimiento de personas mediante la oreja.

Referencia	Objetivo	Contribución principal	Técnicas	Imágenes	Métrica	Resultado (%)
(Dodge et al., 2018)	Predecir la clase a la que pertenece la imagen de entrada.	Mostración que las redes preentrenadas logran una línea de base sólida para el reconocimiento de la oreja sin restricciones.	AlexNet	AWE, CVLE	Rank-5	97.82
			VGG16			99.38
			VGG19			99.69
			ResNet18			99.38
			Resnet50			99.69
(Kamboj et al., 2021)	Segmentar automáticamente la región de la oreja en imágenes de perfil.	Propone EME (<i>Ear Mask Extraction</i>) para extracción de máscara de la oreja a nivel de píxel.	EME	IITD	Exactitud	99.38
				AMI		98.24
(Khaldi & Benzaoui, 2021)	Mejorar el reconocimiento en imágenes de orejas en escala de grises.	Utiliza una arquitectura basada en DCGAN para colorear imágenes antes del reconocimiento con CNN.	DCGAN + CNN	AMI	Rank-1	99.16
				AWE		53.33
(Kamboj et al., 2022b)	Evaluación de calidad de imágenes de oreja.	Clasificación de calidad de imagen con Deep CNN.	CNN + EME	IITK	Exactitud	97.46
(Ramos-Cooper et al., 2022)	Detectar y segmentar la oreja en escenarios controlados.	Segmentación profunda para aislar la oreja en imágenes complejas.	Mask R-CNN	VGGFace-Ear	Rank-1	91.10
				AWE		62.50
(Xu et al., 2023)	Extraer características relevantes descartando las menos significativas.	CNN con fusión dinámica de canales y atención (ECA).	CNN con ECA-Attention y Dynamic Convolution	AMI	Rank-1	99.70
				AWE		72.70

Tabla 3.4. Estado del arte de arquitecturas de aprendizaje profundo (Deep Learning) para el reconocimiento de personas mediante la oreja (continuación).

Referencia	Objetivo	Contribución principal	Técnicas	Imágenes	Métrica	Resultado (%)
(Freire-Obregón et al., 2023)	Mostrar que entrenar múltiples bases mejora la generalización	Muestra de que mezclar <i>datasets</i> mejora el rendimiento frente a entrenar por colección aislada.	VGGNet, InceptionV3, ResNet y ResNetXt	AMI, IITD, AWE, EarVN1.0, BIPLab	Rank-5	>70
(Lei et al., 2023)	Integrar detección y reconocimiento en una sola etapa para mejorar la velocidad.	Propone un detector ligero optimizado para reconocimiento biométrico de oreja.	YOLOv5s-MG	EarVN1.0	Exactitud	98.24
				USTB		99.41
(Bouaouina et al., 2024)	Rendimiento en imágenes oscuras/BN y guiar la atención a regiones discriminativas.	Flujo en dos pasos: DCGAN para mejorar + Mean-CAM para enmascarar ROI.	DCGAN + Mean-CAM	AMI	Rank-1	100.00
				AWE		76.25
(Freire-Obregón et al., 2025)	Explorar “ <i>ear sketches</i> ” (bocetos) con múltiples niveles de abstracción.	Representación estructural/textural tipo boceto + entrenamiento con pérdidas de métrica; evalúa <i>cross-dataset</i> .	CNN + Triplet-Loss	AMI, AWE, IITD-II, BIPLab	Reporta mejoras consistentes vs. Línea base y buena generalización entre bases.	
(Fukuda et al., 2025)	Reconocer orejas capturadas desde múltiples ángulos.	Utiliza mapas de activación para interpretar las regiones discriminativas en el reconocimiento.	CNN + Grad-CAM	<i>Dataset multi-angle</i>	Rank-1	96.50

A partir de los resultados presentados en la Tabla 3.4 se observa que las métricas más utilizadas son Exactitud descrita en el Capítulo 2 (Marco teórico) apartado 2.6 (Métricas de evaluación) y *Rank-1* que indica la probabilidad de que la identidad correcta sea la primera coincidencia en el proceso de identificación. Los resultados muestran que, en condiciones controladas, por ejemplo, en el *dataset* AMI se reportan valores de *Rank-1* de hasta 99.70 %, mientras que en otros casos se alcanza 100 %, como se observa en el trabajo de Bouaouina et al. (2024). De manera similar, varios trabajos reportan valores de Exactitud superiores al 98 % en conjuntos como USTB o EarVN1.0.

Sin embargo, cuando los métodos se evalúan en ambientes no controlados, el desempeño tiende a disminuir. Esto se observa en conjuntos como AWE, donde se reportan valores de *Rank-1* bajos (53.33 % y 62.50 %) con respecto al resto de la tabla. Esta diferencia se debe a que las imágenes en estos escenarios presentan variaciones más complejas, como cambios de iluminación, pose, oclusiones y fondo, lo que dificulta la generalización de los modelos.

En conjunto, los resultados muestran que los enfoques basados en aprendizaje profundo representan actualmente una de las principales líneas de investigación en el reconocimiento biométrico de orejas. No obstante, también muestran que el desempeño de estos sistemas depende de las condiciones de adquisición de las imágenes y de la capacidad del modelo para generalizar entre diferentes conjuntos de imágenes.

3.2.3. Modelos híbridos

Kamboj et al. (2022a)., presentaron un estudio de técnicas tradicionales y el enfoque basado en el aprendizaje profundo para el reconocimiento de personas mediante la oreja. El objetivo del estudio fue analizar el desempeño de varias arquitecturas con imágenes desafiantes en condiciones del mundo real para brindar un mayor avance en la tecnología del reconocimiento mediante la oreja. Los autores describen las ventajas que tiene el biométrico de la oreja sobre los otros, y que combinado con algún otro (cara, iris y cara lateral) se tendría un mejor rendimiento ante seguridad y los ataques de identidad. Presentan una estructura para su estudio, dividida en tres partes; bases de datos de referencia, parámetros de evaluación del rendimiento y métodos existentes para la detección y el reconocimiento de orejas, como se muestra en la Figura 3.2.

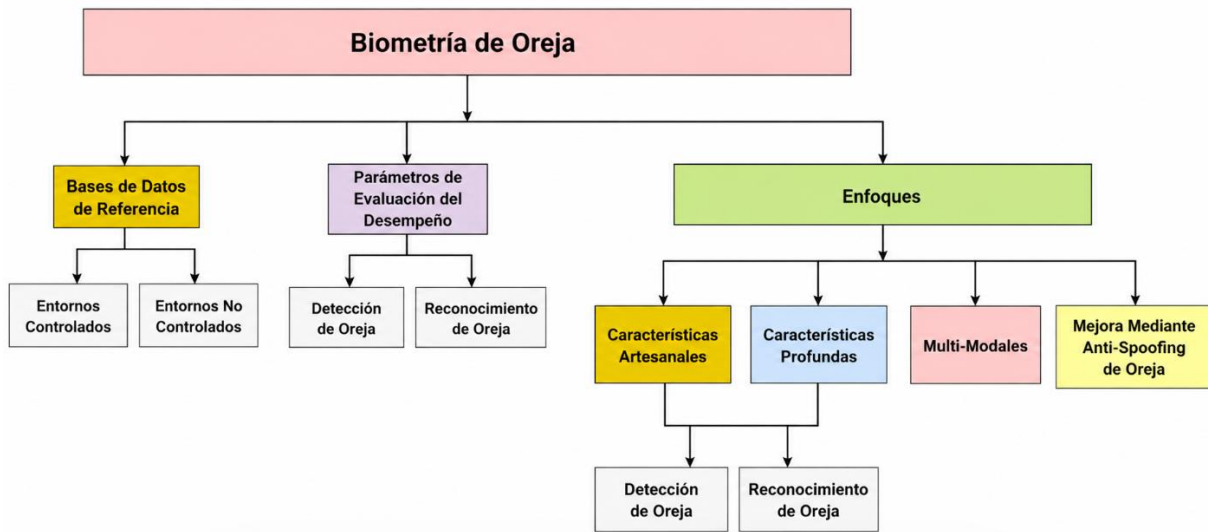


Figura 3.2. Taxonomía de la biometría de la oreja (Kamboj et al., 2022a)

Como contribución, proponen el conjunto de imágenes NITJEW por sus siglas en inglés (*National Institute of Technology Jalandhar Ear in the Wild*), para el reconocimiento de orejas sin restricciones. Llevan a cabo la evaluación comparativa de la detección y reconocimiento en seis bases de datos populares (hasta el 2020) y la propia. Las pruebas consideran variación de escenarios, oclusión, cambios de pose e iluminación. Los autores reportan que la técnica tradicional que mejores resultados presentó es el descriptor LPT con 100.00 % de reconocimiento, seguida de SIFT y LBP con 99.67 % y 98.10 % respectivamente. Con respecto al aprendizaje profundo se reporta que las Redes Neuronales Convolucionales obtuvieron entre 96.99 % y 98.27 % de reconocimiento, donde VGG-Face y ResNet obtienen 96.08 % y 93.45 % respectivamente.

Bouazouina et al. (2023) realiza una revisión del estado del arte y respalda la vigencia de LBP/LTP/LDP/BSIF/LPQ/Gabor/Haralick como línea base y componentes de fusión. Los autores proponen emplear redes generativas antagónicas convolucionales profundas (DCGAN) para mejorar las imágenes de orejas y se introduce una técnica de extracción y clasificación de características, denominada *Mean-CAM-CNN*, en el estado del arte compara y discute LBP y variantes (ALBP, RLOP) como línea base para la extracción de textura.

La Tabla 3.5 presenta trabajos relacionados con arquitecturas híbridas para el reconocimiento de personas mediante la oreja. La información se organiza por referencia, objetivo, contribución principal, técnicas empleadas, conjunto de imágenes, métrica de evaluación y resultado reportado. En general, los estudios integran técnicas tradicionales, como LBP, HOG, Gabor, PCA o descriptores de textura, con modelos de aprendizaje profundo, principalmente CNN como AlexNet, VGG, ResNet y variantes relacionadas.

Tabla 3.5. Estado del arte de arquitecturas de modelos híbridos para el reconocimiento de personas mediante la oreja.

Referencia	Objetivo	Contribución principal	Técnicas	Imágenes	Métrica	Resultado (%)
(Emeršič et al., 2020)	Analizar el desempeño de modelos de reconocimiento de oreja en escenarios no controlados.	Evaluación comparativa de técnicas tradicionales y CNN.	BSIF, DSI, Gab, HOG, LBP, LPQ, PO, RIL, RNet, SNet, VGG.	AWE	Rank-1	BSIF 55.60
						SNet 62.90
					AUC	Gab 94.70.
						SNet 93.90
(Alkababji & Mohammed , 2021)	Construir un sistema de reconocimiento de personas robusto, rápido y preciso.	Técnica para construir un sistema de reconocimiento de personas utilizando aprendizaje profundo.	AlexNet, PCA y Algoritmo genético, ANN	AMI	Exactitud	97.80
(Bansal & Madasu, 2021)	Diseñar un clasificador basado en t-normas y la función de entropía.	Clasificador Hanman Ponderado (WHC), a partir de la función de entropía Hanman-Anirban.	WHC, Función de entropía (HAE)	SBDB2	Exactitud	AlexNet 66.00
						LeNet 38.20
						VGG16 32.50
(Aiadi et al., 2022)	Reconocimiento de <i>ear prints</i> con un método ligero y no supervisado	Propone MDFNet: red de una sola capa que fusiona Magnitud y Dirección del gradiente con filtros guiados por datos (PCA). Incluye un paso de alineación de oreja mediante CNN + PCA previo a la extracción.	MDFNet, PCA, CNN usada para alineación	AWE	Rank-1	82.00
				AMI		97.67
				IITD-II		98.96
(Benzaoui et al., 2023)	Análisis del estado del arte y línea base.	Síntesis de descriptores de textura. Compara métodos clásicos (LBP, Gabor, Haralick) con arquitecturas CNN, ResNet, VGG, EfficientNet	Tradicional + DL (comparativo/fusión).	Revisión de textura clásica y su uso con DL; discute detectores/segmentación previos a la extracción.		

Tabla 3.5 Reconocimiento de personas utilizando modelos híbridos (continuación)

Referencia	Objetivo	Contribución principal	Técnicas	Imágenes	Métrica	Resultado (%)
(Ganesan et al., 2023)	Mejorar el reconocimiento o utilizando representaciones geométricas y aprendizaje profundo.	Integración de representación geométrica basada en matriz de covarianza con redes profundas para clasificación.	Covariance Matrix + Deep Learning.	UND-J2	Rank-1	99.63
(Oyebiyi et al., 2023)	Análisis del estado del arte de los métodos tradicionales de textura frente a los de DL	73 artículos analizados	LBP/LPQ /LDP/Gabor/GLCM + CNN (en revisión teórica).	Revisión sistemática que cuantifica el peso de los métodos texturales frente al DL (27 % vs. 58 % de publicaciones). Propone pipelines híbridos para entornos con poca muestra.		
(M N & Prasanna, 2024)	Reconocer personas a partir de imágenes de oreja mediante un enfoque híbrido denominado GSSO_CNN-TL.	Propone un modelo híbrido que combina preprocesamiento, aumento de datos, extracción de características clásicas y reconocimiento con CNN basada en aprendizaje por transferencia.	NLM, SIFT, SURF, PHoG y LBP. DbneAlex Net	<i>Datasets for ear detection and recognition</i>	Precisión	92.60
					Recall	95.80
					F-measure	94.20

A partir de la Tabla 3.5 se observa que los enfoques híbridos combinan descriptores tradicionales con modelos CNN para mejorar la representación de la oreja en tareas de reconocimiento. Los trabajos reportan resultados variables, con métricas como Rank-1, AUC, Exactitud, Precisión, *Recall* y *F-measure*, debido a diferencias en bases de datos, técnicas y protocolos de evaluación. Los conjuntos más utilizados son AMI, AWE, IITD, UND y otros bancos de imágenes orientados al reconocimiento de oreja, lo que confirma la relevancia de estos recursos en la evaluación de modelos híbridos.

En conjunto, la revisión permite concluir que la combinación de técnicas tradicionales con aprendizaje profundo constituye una línea de investigación pertinente para mejorar el desempeño de los sistemas biométricos basados en oreja. Esta tendencia respalda la incorporación de representaciones complementarias, como textura, geometría o características profundas, en modelos orientados al reconocimiento robusto de personas mediante este rasgo biométrico.

3.3. Sistemas de detección de ataques de presentación (PAD/anti-spoofing)

La detección de ataques de presentación (PAD) ha cobrado relevancia en biometría porque los sistemas pueden ser vulnerados mediante instrumentos que imitan rasgos genuinos. La norma ISO/IEC 30107-1 (2023a) establece el marco conceptual para especificar y evaluar métodos PAD.

3.3.1. PAD en biometría

En modalidades como rostro y huella dactilar, se han propuesto múltiples estrategias de PAD basadas en análisis de textura, calidad de imagen, señales fisiológicas y aprendizaje profundo. Las investigaciones han sido estandarizadas mediante la norma *ISO/IEC 30107-1* (2023b), que define métricas como APCER, BPCER y ACER. Uno de los problemas que aborda el proyecto de tesis es la limitada información sobre PAD para sistemas basados en la oreja, en comparación con la de biometría facial o dactilar (Sharma & Selwal, 2021), (Ametefe et al., 2022), (Sharma & Selwal, 2023) (Muhammad et al., 2024), (Pooshideh et al., 2024) .

3.3.2. PAD en la oreja humana

Como cualquier otro sistema biométrico, el basado en la oreja también es vulnerable a las amenazas o ataques de suplantación, y un ataque puede poner en peligro la seguridad del sistema. Se han propuesto algunos métodos anti-plagio para otros rasgos biométricos como el rostro, la huella dactilar y el iris. Como el trabajo de Fatemifar et al. (2019), el cual busca anomalías dentro de la imagen que permita identificar un ataque o amenaza de suplantación. La detección de anomalías es la identificación de datos inusuales. Puntos que se desvían del comportamiento de la mayoría de los datos. El trabajo desarrolló una solución para el problema de detección de suplantación de identidad mediante la fusión de múltiples expertos en anomalías.

En (Sepas-Moghaddam et al., 2018) se propone un enfoque de normalización para admitir una fusión de detectores de suplantación múltiple, utilizando solo datos normales. Sin embargo, en sistemas de autenticación mediante la oreja, las aportaciones han sido pocas, se han registrado pequeños avances para evitar la suplantación o ataques a la seguridad de los sistemas biométricos. La presentación de artefactos auditivos frente a los sensores de adquisición, utilizando Instrumentos de Ataque de Presentación (*PAI - Presentation Attack Instruments*), se conoce como ataque de suplantación o presentación, lo que genera nuevas preocupaciones de seguridad para los sistemas de reconocimiento de oreja. Se han propuesto soluciones de Detección de Ataques de Presentación (*PAD-Presentation Attack Detection*), también conocidas como soluciones *anti-spoofing*, para detectar automáticamente los ataques de suplantación.

En el 2020 Toprak y Toygar en su trabajo referente a la suplantación, afirmaron que el reconocimiento o identificación de personas mediante el uso de la oreja, es uno de los sistemas biométricos que ha ido cobrando mayor importancia. El sistema biométrico, tiene características que le dan cierta ventaja sobre los otros rasgos biométricos, por ejemplo, su estructura no es afectada con el paso del tiempo ni por las expresiones faciales, lo que permite obtener los datos con facilidad.

Por otra parte, el uso de las redes sociales en la actualidad da la posibilidad de obtener fotos o videos de una persona con facilidad, generando vulnerabilidad a los sistemas biométricos que consideran a la oreja como acceso. Los ataques pueden surgir en el nivel del sensor que captura la imagen, donde los sistemas de identificación pueden ser engañados mediante una oreja falsa tales como una imagen impresa o la reproducción de un video (Figura 3.3).

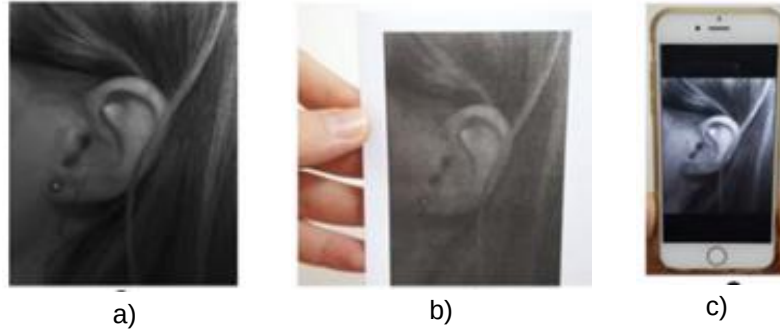


Figura 3.3. Tipos de suplantación para ataque o engaño a) imagen en vivo, b) Imagen impresa, c) Reproducción de video (Toprak & Toygar, 2020)

En el trabajo de Toprak & Toygar (2020) se propone un método para evitar los ataques de suplantación, que se fundamenta en la calidad de la imagen, la cual es un factor que muestra la diferencia entre imágenes reales e imágenes falsas. La calidad de la imagen que se adquiere de una persona viva es diferente a la imagen del impostor, que se adquiere mediante una foto impresa o una pantalla digital (Toprak & Toygar, 2020). El modelo está basado en la implementación de 25 Medidas de la Calidad de la Imagen (*IQM - Image Quality Measures*) y dividido en tres niveles de fusión: característica, puntaje y decisión (véase Figura 3.4).

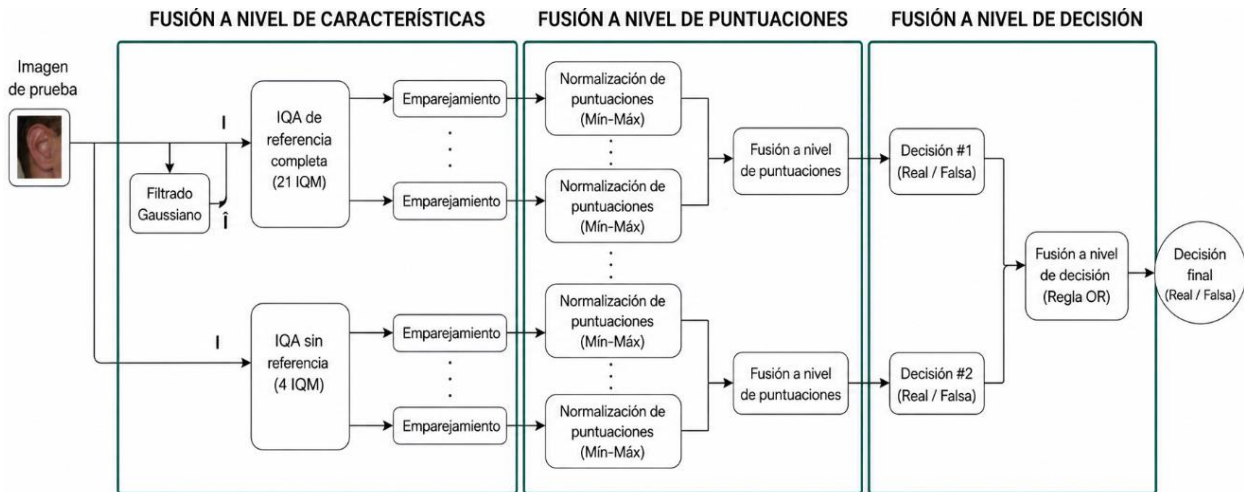


Figura 3.4. Estructura del modelo anti-spoofing (Toprak & Toygar, 2020)

Las autoras de la investigación aseveran que existen desarrollos para prevenir los ataques de suplantación en la mayoría de los sistemas biométricos (rostro, iris, huellas dactilares, voz e impresión de la palma), pero no para el sistema biométrico mediante la oreja. Toprak y Toygar presentaron en (2021) una mejora del trabajo descrito anteriormente, en la cual, se hace uso de los rasgos biométricos de la oreja, combinando Red Neuronal Convolutiva (*CNN - Convolutional Neural Network*) con cinco Medidas de Calidad de la Imagen (*IQM - Image Quality Measures*), para obtener un algoritmo robusto (Toprak & Toygar, 2021). Los resultados obtenidos se evalúan en cuatro bases de datos de oídos: AMI, UBEAR, IITD y USTB. La metodología utilizada en el estudio está dividida en Evaluación de la Calidad de la Imagen (*IQA - Image Quality Assessment*) y aprendizaje profundo basado en *CNN* (*CNN-Based Deep Learning*) (véase Figura 3.5).

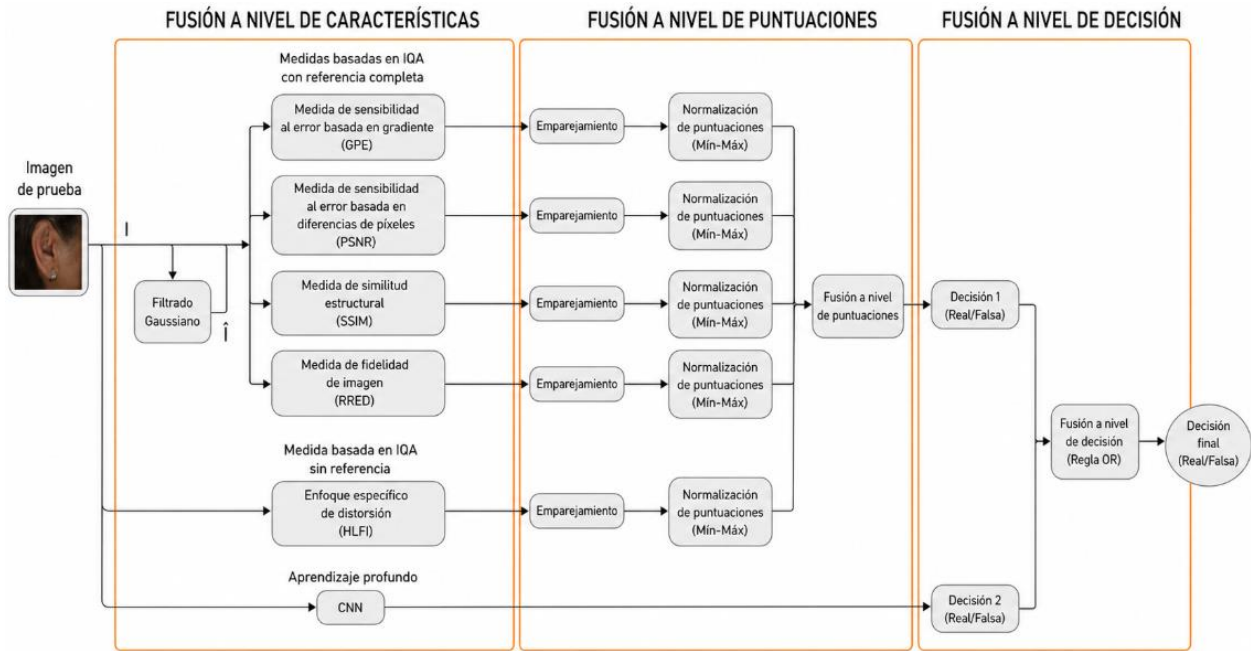


Figura 3.5. Diagrama general del modelo propuesto (Toprak & Toygar, 2021)

El estudio menciona que, utilizan dos tipos de medidas de calidad de la imagen (IQM): IQM de referencia completa (*FR - Full-Reference*) y IQM sin referencia (*NR - No Reference*).

- Dentro *FR IQM* se emplean como medidas: Relación Señal Pico a Ruido (*PSNR - Peak Signal to Noise Ratio*), Error de Fase de Gradiente (*GPE - Gradient Phase Error*), Índice de Similitud Estructural (*SSIM - Structural Similarity Index*) y Diferencia Entrópica de Referencia Reducida (*RRED - Reduced Reference Entropic Difference*).
- Para *NR IQM* se emplea: Índice de Frecuencia Alta-Baja (*HLFI - High-Low Frequency Index*).

En la Tabla 3.6, se presenta información adicional de los artículos aquí mencionados y se compara con los resultados del modelo propuesto.

Tabla 3.6. Técnicas para la detección de ataques de suplantación (anti-spoofing)

Referencia	Objetivo	Contribución	Técnicas	Imágenes	Métrica	Resultado (%)			
(Nourmoham madi-Khiarak & Pacut, 2019)	Encontrar un mejor conjunto de características discriminantes	Modelo IQA+SVM. Base de datos con orejas falsas y genuinas.	11 IQM SVM	AMI, DB propia (UoT)		AMI	UoT/Tipos de ataques		
							Foto	Pantalla	Video
					FFR	0.00	0.00	0.00	0.00
					FRR	0.00	3.06	0.00	0.00
					HTER	0.00	1.53	0.00	0.00
IQM (11): (MSE, PSNR, SNR, SC, MD, AD, NAE, UQI, PCC, MAE, RMSE). SVM.									
(Toprak & Toygar, 2020)	Reconocer imágenes de orejas reales y falsas	Modelo con 25 IQM sobre tres niveles de fusión.	25 IQM	AMI, UBEAR		AMI		UBEAR	
					FFR	2.00		20.00	
					FGR	0.00		0.00	
					HTER	1.00		10.00	
IQM (25): (MSE, AD, GME, GPE, LMSE, MD, NAE, NCC, PSNR, RMD, SNR, SME, SPE, SC, SSIM, TCD, TED, VIF, RRED, MAMS, MAS, JQI, HLFi, BIQI, NIQE)									
(Toprak & Toygar, 2021)	Identificar suplantación.	Identificación de falsificación mediante IQM y CNN.	PSNR, GPE, SSIM, RRED, HLFI, ReLU	AMI, UBEAR , IITD y USTB (set 1 y set 2)		AMI	UBEAR	IITD	USTB
					FFR	0.00	0.00	0.00	0.00
					FGR	0.00	0.00	0.00	0.00
					HTER	0.00	0.00	0.00	0.00
(Xu et al., 2024)	Identificación de ataques de software y a manipulación de plantillas.	CNN y AccDrop; protección de plantillas	CNN+ Attention + AccDrop/ DropBlock; Blockchain	AWE, EarVN1 .0	Rank-1	AWE		83.87	
						EarVN1.0		96.52	

Se observa que el conjunto de imágenes de mayor uso es AMI. Es notorio el uso de las Medidas de la Calidad de la Imagen (IQM - *Image Quality Measures*) para la detección de ataques de suplantación, ya que las autoras afirman que la calidad de la imagen que se adquiere de una persona viva es mejor a la imagen que se tiene en una foto impresa o en una pantalla digital.

A partir de la revisión realizada, se identifica limitada información de métodos de detección de ataques de presentación para biometría de oreja evaluados en entornos no controlados y frente a artefactos modernos, incluyendo imágenes generadas por inteligencia artificial. Dado lo anterior, se propone un modelo basado en fusión temprana de color y textura como etapa preliminar de seguridad en sistemas biométricos de oreja.

En contraste, la biometría de la oreja presenta un rezago significativo en investigaciones de PAD. Los pocos trabajos existentes se centran en ataques impresos y utilizan descriptores manuales o métricas de calidad de imagen. Los enfoques muestran limitaciones importantes ante ataques más sofisticados, como modelos 3D o imágenes generadas por IA.

Estudios recientes han establecido que la textura de la piel humana viva difiere de manera significativa de materiales sintéticos, y que las diferencias pueden ser capturadas mediante

descriptores como LBP. No obstante, la mayoría de los modelos profundos continúan utilizando únicamente información RGB. A partir de la revisión, se identificaron las siguientes limitaciones comunes en el estado del arte:

1. Omisión de la etapa de localización en la mayoría de los sistemas biométricos de oreja, sin analizar el impacto de errores en la etapa.
2. Dependencia de imágenes recortadas y alineadas, poco representativas de escenarios reales.
3. Escasez de estudios y *datasets* para PAD específicos para biometría de la oreja, particularmente en escenarios no controlados.
4. Uso predominante de información RGB, sensible a condiciones de captura no controladas.
5. Evaluaciones limitadas frente a ataques emergentes basados en IA generativa.

Dichas limitaciones muestran la necesidad de un sistema integral que combine localización, análisis de textura y aprendizaje profundo, evaluado bajo condiciones reales y orientado explícitamente a la detección de ataques de suplantación.

CAPÍTULO 4

PROPUESTA DE SOLUCIÓN

La propuesta de solución es la formulación de un modelo computacional para la diferenciación entre imágenes de orejas humanas naturales e imágenes generadas en condiciones de entornos controlados y no controlados, mediante una arquitectura híbrida que fusiona características de textura local con representaciones profundas. Contribuyendo a evitar un posible ataque de presentación. El capítulo cuatro aborda la conceptualización del modelo 4C y la descripción de sus etapas y las herramientas de desarrollo.

4.1. Descripción General

Para abordar la problemática de la caracterización de la oreja apoyando la Detección de Ataques de Presentación (PAD) en entornos controlados y no controlados, se propone una solución secuencial dividida en cinco fases (ver Figura 4.1), cada uno de ellos están diseñados para operar de manera modular, permitiendo su evaluación individual, donde las contribuciones principales se encuentran en el módulo dos correspondiente a la Localización de la Región de Interés (ROI) y, el módulo tres referente a la Extracción y Fusión de Características (Modelo 4C).

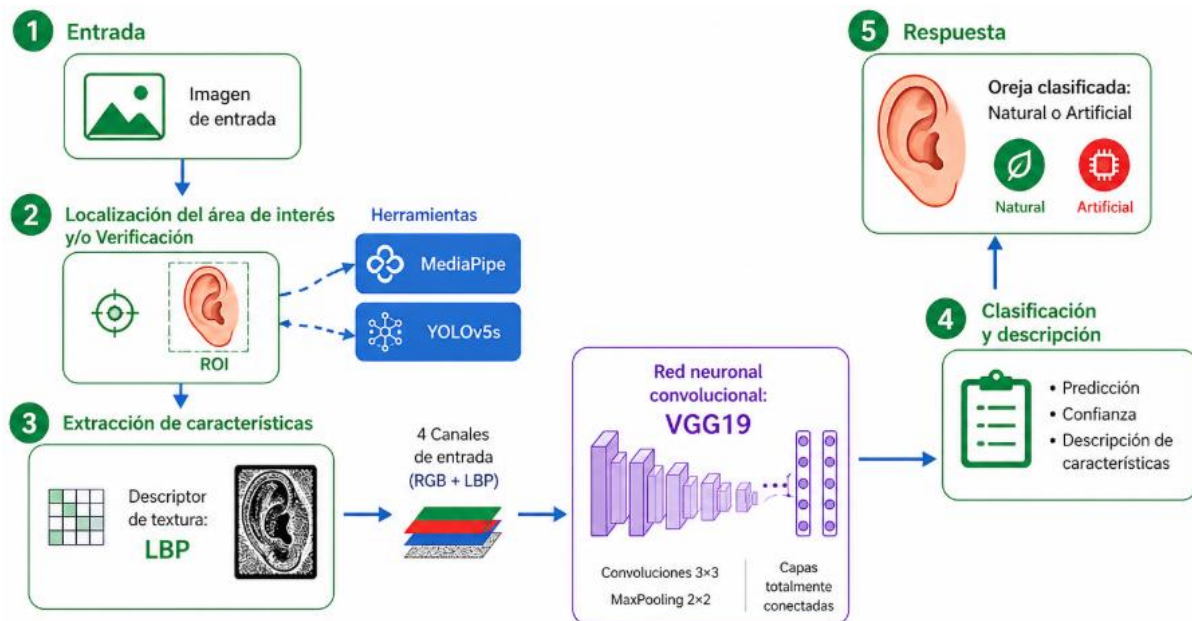


Figura 4.1 Flujo del modelo propuesto para la identificación de orejas humanas naturales y artificiales mediante extracción de textura LBP y la red VGG19.

El flujo de trabajo, ilustrado en la Figura 4.1, procesa una imagen de entrada sin restricciones de fondo o iluminación, extrae el área de la oreja y determina su naturaleza (*Bona Fide* o Ataque) mediante un análisis de textura y forma, como se describe en los siguientes puntos:

1. **Imagen de entrada.** Para el entrenamiento del sistema, recibe un conjunto de imágenes (RGB) previamente recopilado; en el capítulo cinco se detalla la información relacionada a dichos conjuntos. La región de interés extraída se somete a un proceso de preprocesamiento que incluye normalización de tamaño, conversión de formato y ajuste de intensidad. Las operaciones tienen como objetivo reducir la variabilidad inducida por las condiciones de captura y facilitar el aprendizaje del modelo de clasificación. Es conveniente mencionar que el modelo en la etapa de validación puede trabajar con imágenes de captura directa.
2. **Localización del área de interés.** En la etapa se identifica la región potencial donde se encuentra la oreja dentro de la imagen. Para ello, se emplean técnicas de detección basadas en MediaPipe y, de ser necesario, se realiza una verificación complementaria mediante el modelo YOLOv5s. La fase garantiza que el recorte obtenido contenga una oreja válida para el análisis posterior.
3. **Extracción de características.** Una vez localizada la oreja, se procede a la extracción de textura utilizando el método LBP, con el fin de capturar información local relacionada con la piel y los materiales presentes. El operador LBP genera una representación adicional que codifica la distribución de patrones de textura. La información complementa la representación visual proporcionada por los canales RGB de la imagen original. Posteriormente, la información se integra con la representación RGB de la imagen para formar una estructura de cuatro canales (imagen 4C).
4. **Clasificación mediante red neuronal convolucional.** La imagen 4C se introduce en la red VGG19, adaptada para aceptar cuatro canales de entrada (la imagen RGB más la textura). El modelo realiza la extracción profunda de características discriminativas y lleva a cabo la clasificación final mediante un proceso de aprendizaje supervisado.
5. **Respuesta del sistema.** Finalmente, el modelo emite una respuesta que indica si la imagen analizada contiene una oreja humana natural o una oreja artificial.

La fusión de un canal adicional de textura basado en LBP con imágenes RGB permite enriquecer la representación de entrada al modelo de aprendizaje profundo (VGG19), facilitando la discriminación entre superficies de piel humana y materiales artificiales. Logrando una arquitectura híbrida para detectar los ataques de presentación (PAD).

El análisis del estado del arte relacionado muestra el uso preferente de conjuntos de imágenes públicas que contienen imágenes de orejas perfectamente recortadas dentro de un rectángulo mínimo (ver Tabla 3.2). En algunos casos con presencia de cabello o aretes, sombras o cambios de iluminación.

4.1.1. Localización y Segmentación (Enfoque Híbrido)

En los sistemas biométricos basados en oreja, la localización corresponde a la etapa inicial en la que se ubica la región de la imagen donde se encuentra la oreja con el fin de aislarla para posteriores etapas de extracción de características y reconocimiento (Szeliski, 2011).

La búsqueda realizada del estado del arte para conocer cómo se estaba resolviendo la etapa de localización, se encontró poca información sobre un método o herramienta en específico que realice la actividad, debido a que la mayoría de los trabajos de reconocimiento usan conjuntos públicos que se encuentran en ambientes controlados (Alagarsamy & Murugan, 2020), (Hassaballah et al., 2020), (Sajadi & Fathi, 2020), (Sivanarain & Viriri, 2020), (Toprak & Toygar, 2020), (Bansal & Madasu, 2021a), (Toprak & Toygar, 2021), (Ramos-Cooper et al., 2022); por lo que llegan a omitir la fase, y en aquellos trabajos que si lo realizan utilizan redes neuronales convolucionales (Alkababji & Mohammed, 2021), (Kamboj et al., 2022b), (Lei et al., 2023), (Fukuda et al., 2025). Por tal motivo, se propone un sistema híbrido utilizando MediaPipe y YOLOv5s, representada en el diagrama de la Figura 4.2.

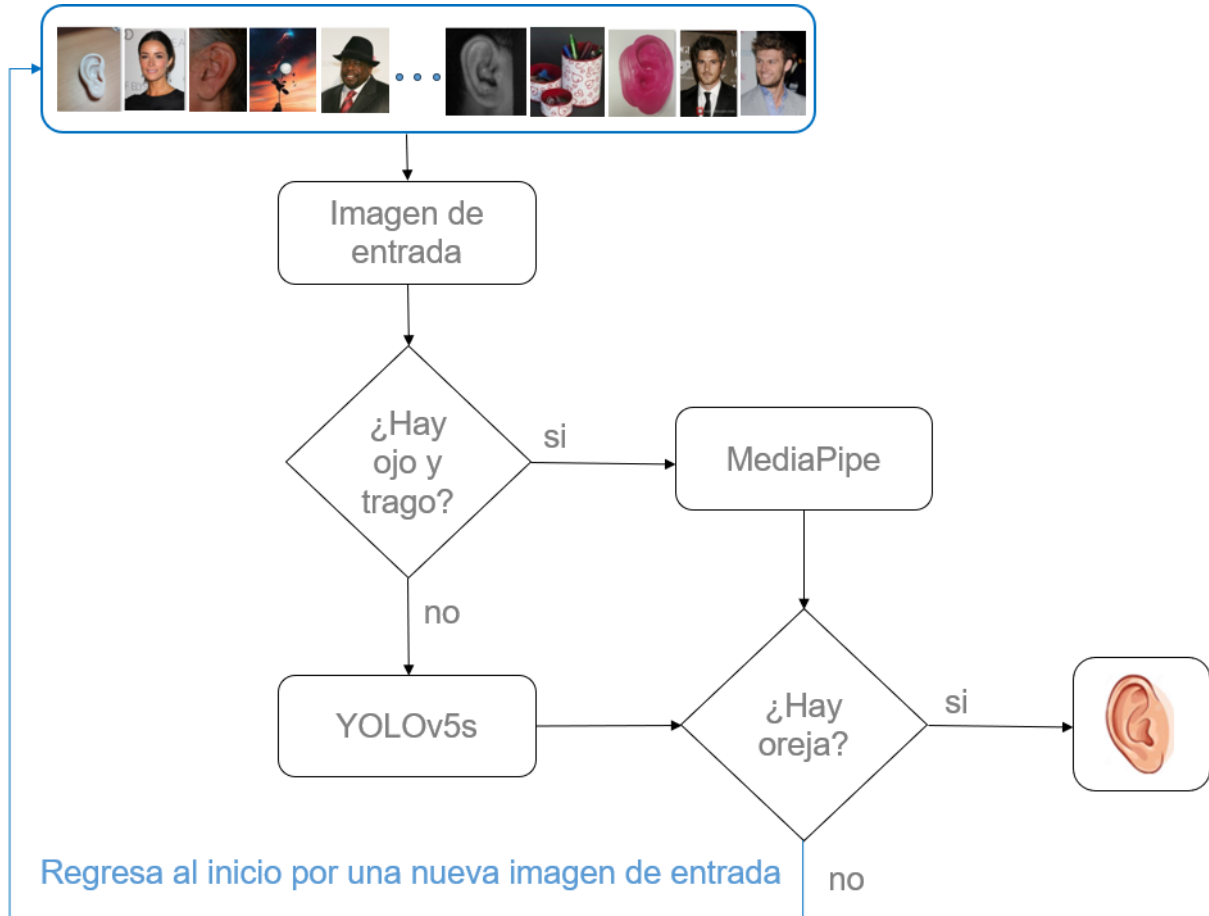


Figura 4.2 Diagrama para la localización del área de interés.

En el desarrollo de la etapa, se consideró que la imagen de entrada podría contener un rostro con diferente perspectiva; es decir, frontal e inclusive con diferentes ángulos, de tal forma que el sistema pueda localizar la oreja en escenarios reales. La segunda opción es que la imagen presente a una oreja completa o con algún grado de oclusión debido a la presencia de cabello, aretes, audífonos, entre otros.

La etapa combina dos métodos utilizados para la detección de objetos como son MediaPipe (2022) y YOLOv5s (2023). La idea principal es buscar en la imagen de entrada el área donde potencialmente pueda existir una oreja. Primero se hace la búsqueda mediante MediaPipe de

los puntos estratégicos (ojo y trago) que apoyan a detectar si en la imagen hay un rostro, permitiendo trabajar en ambientes reales; en caso de ser encontrados se sigue con el proceso de identificación de la oreja humana.

En caso contrario, se necesita el apoyo de un clasificador para localizar el ROI, en el que se ingresa al módulo de YOLOv5s para que localice la oreja en la imagen de entrada, si es encontrada se sigue con el proceso de identificación, en caso contrario se regresa al inicio del proceso general.

4.1.1.1. Localización de la oreja para entornos no controlados

El proceso de la localización propuesto se presenta en la Figura 4.3.

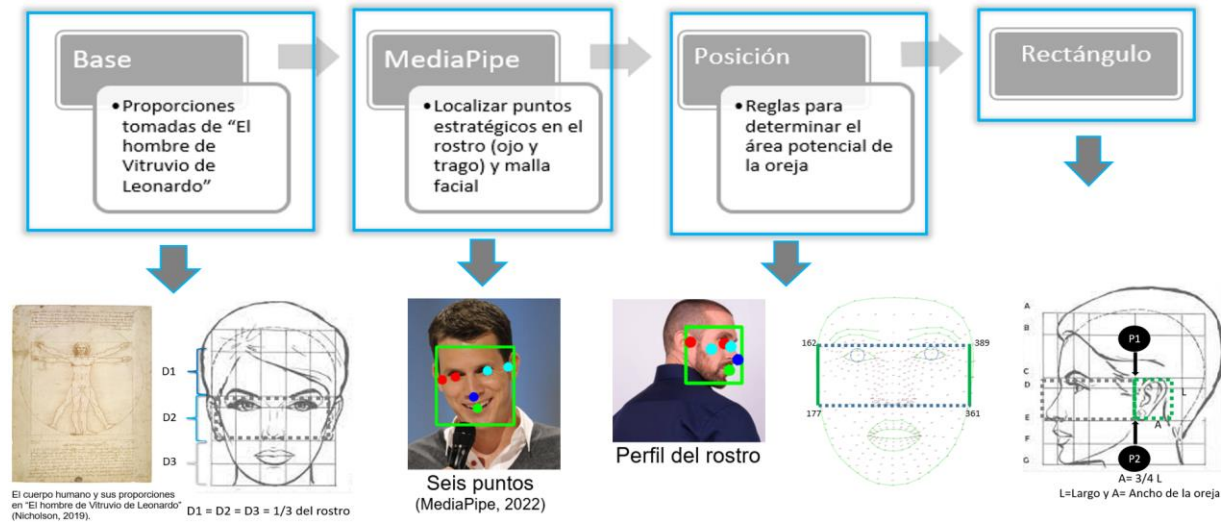


Figura 4.3. Ilustración del proceso de localización de la oreja para entornos no controlados mediante MediaPipe.

Para especificar las dimensiones de una oreja, se adoptó como referencia lo propuesto por Leonardo da Vinci en el *Hombre de Vitruvio* (*Uomo vitruviano*) (da Vinci, 1490), estudio que menciona que la oreja es una tercera parte del rostro y es proporcional a la distancia que va de la parte inferior de la barbilla hasta la nariz. De igual forma, la distancia entre el nacimiento del cabello y las cejas mantiene dicha relación (da Vinci, 1883) (Losardo et al., 2015) (Nicholson, 2019). En la presente investigación, dicha proporción se emplea para delimitar la región de interés. A partir de ello, se seleccionó MediaPipe (2022), ya que permite la localización de puntos faciales relevantes y permite trasladar este criterio proporcional a una implementación computacional dentro del proceso de detección. El proceso se lleva a cabo en dos pasos:

Paso uno: ubicación los seis puntos estratégicos del rostro que son: ojos (2 puntos), tragos (2 puntos, señalado en la Figura 4.4-a en círculo azul) nariz y boca (ver Figura 4.4-b).

La localización de los seis puntos es utilizada para determinar la orientación del rostro y con ello la información de los elementos presentes. Es decir, determinar el área de interés, señalando en color rojo los puntos que corresponden al trago y ojo del lado derecho y en azul el trago y el ojo del lado izquierdo del rostro (ver Figura 4.4-b). Se calcula la distancia euclídea

entre la posición del ojo y el trago de cada lado del rostro. El valor mayor indica el perfil (orientación) en que se encuentra el rostro (Figura 4.4-c). En caso de no localizar los seis puntos (por tener un rostro de perfil, como en la Figura 4.4-c) la plantilla predice la posición y presenta los puntos en la imagen.

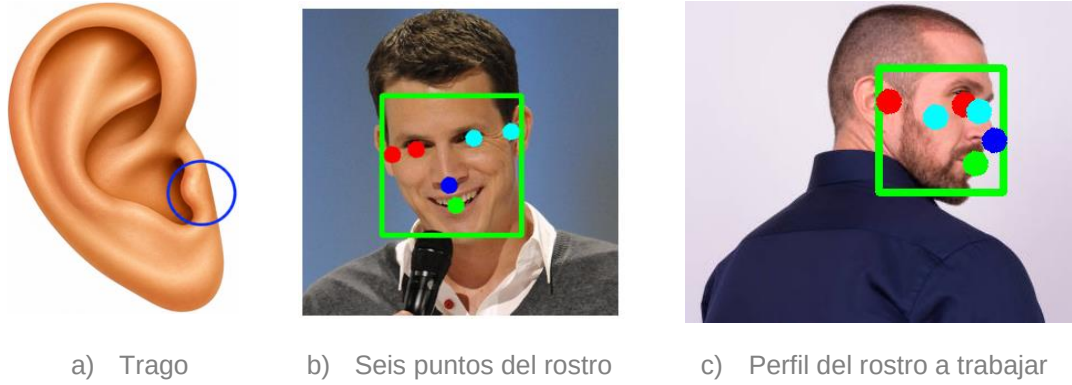


Figura 4.4 Localización de puntos de referencia faciales para la estimación de la orientación del rostro a) representación del trago, b) rostro en vista frontal y c) perfil del rostro a analizar.

Paso dos: Mediante la malla facial de MediaPipe (2022) se consideran los puntos de los costados del rostro que se encuentran a la altura de las orejas (ver Figura 4.5-a). Los valores de los puntos (de arriba hacia abajo) son: para la oreja del lado izquierdo el 389 y 361 y, para la oreja del lado derecho se toman los puntos 162 y 177; con los puntos se determina la longitud de la oreja para lo cual se usa la regla de proporción del cuerpo humano de Vitruvio (da Vinci, 1490), (véase Figura 4.5-b). Finalmente, para calcular el ancho del recuadro o enmarcado se toman $\frac{3}{4}$ partes (proporción que se obtuvo de manera experimental) de la longitud obtenida (ver Figura 4.5-c).

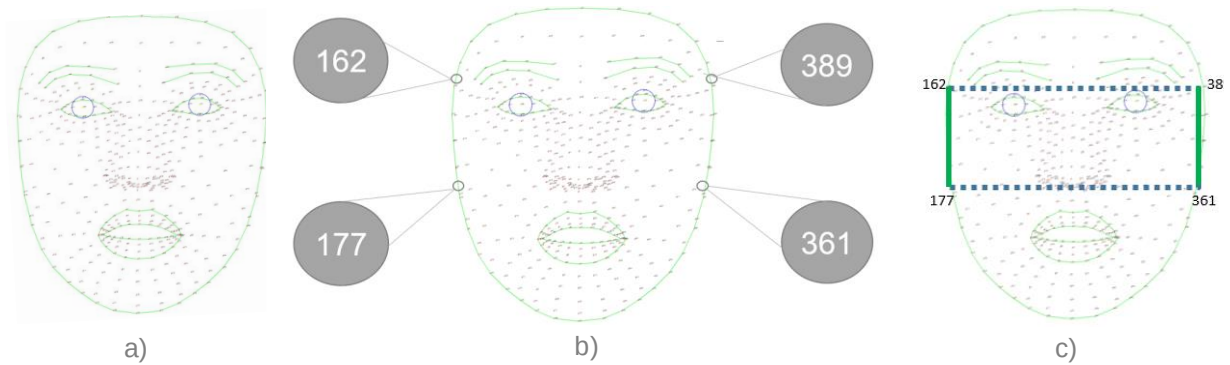


Figura 4.5 Localización de la región de interés para la oreja utilizando puntos de referencia faciales c) Malla facial, b) puntos de referencia para localizar la oreja, c) puntos localizados en el rostro.

La figura muestra el proceso de localización de la región de interés (ROI) de la oreja utilizando puntos de referencia obtenidos de la malla facial.

4.1.1.2. Localización de la oreja para entornos controlados

La integración de YOLOv5s (2023) para la localización consta de dos partes: entrenamiento y localización. Para el entrenamiento del modelo, se utilizó un conjunto integrado por imágenes

de conjuntos públicos que contienen orejas recortadas y alineadas; pero también con imágenes descargadas de la web que presentan diferentes condiciones tales como tener el rostro completo y otros aspectos presentes en un ambiente no controlado. Además, se consideró la categoría de no oreja con objetos distintos, pero con cierta relación a la entidad de interés. La información de las imágenes utilizadas se expone en el Capítulo 6 (Pruebas y Resultados) apartado 6.2 (Evaluación del método de localización - Caso uno). En la Figura 4.6 se presenta el esquema utilizado para la implementación de la localización del área de interés en ambientes controlados mediante YOLOv5s.

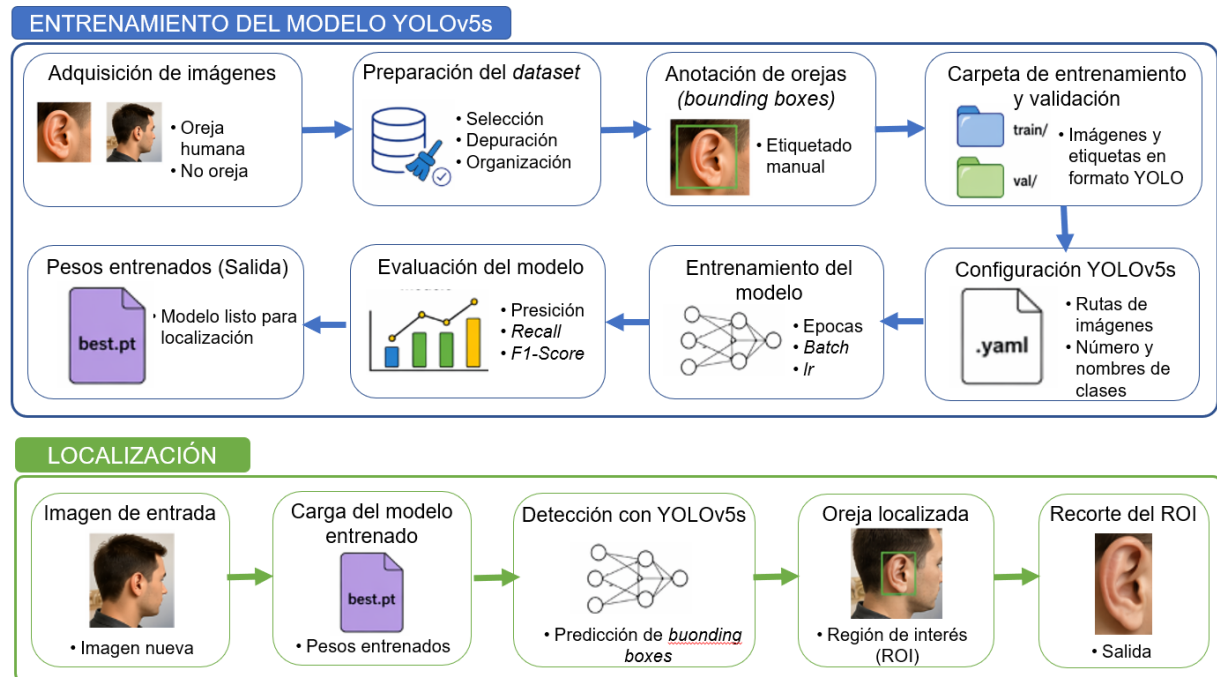


Figura 4.6 Esquema de entrenamiento y localización de orejas humanas mediante YOLOv5s.

La Figura 4.6 presenta el flujo del sistema de localización de orejas utilizando YOLOv5s, dividido en dos etapas. En entrenamiento, se realiza la adquisición de imágenes, preparación del dataset, anotación de orejas mediante *bounding boxes*, organización en carpetas de entrenamiento y validación, configuración del modelo y entrenamiento. Posteriormente, el modelo se evalúa con las métricas de precisión, *recall* y *F1-score*, generando como salida los pesos entrenados (*best.pt*).

En la etapa de localización, una imagen de entrada es procesada por el modelo entrenado para detectar la oreja mediante YOLOv5s. A partir de la *bounding box* predicha, se obtiene la región de interés (ROI) correspondiente a la oreja, que constituye la salida del proceso.

4.1.2. Extracción de características (Fusión RGB + LBP)

La extracción de características consiste en transformar los datos originales en una representación más adecuada para su procesamiento por algoritmos de aprendizaje automático (Scikit-learn, 2026). De acuerdo con Guyon y Elisseeff (2006), este problema puede entenderse mediante dos etapas principales: la construcción de características, donde los datos crudos se transforman en variables o descriptores útiles, y la selección de

características, donde se identifican aquellas características más relevantes o informativas para el modelo. En esta investigación, la extracción de características se realiza mediante la incorporación del descriptor LBP, propuesto por Ojala et al.(2002), el cual permite representar patrones locales de textura a partir de imágenes en escala de grises. El núcleo de la propuesta de solución es la fusión de la información de textura y los canales RGB en una imagen de cuatro canales. La imagen 4C se introduce en una red neuronal convolucional profunda basada en la arquitectura VGG19 (Simonyan & Zisserman, 2015).

4.1.2.1. Transformación 3C a 4C

Para convertir una imagen con tres canales de color (RGB) en una imagen de cuatro canales que incluya información de textura (T), se agrega un cuarto canal derivado de la aplicación de LBP (*Local Binary Pattern – Patrón binario local*) sobre la imagen original. En esta tesis se adopta el esquema clásico del operador, equivalente al modo “default” en implementaciones estándar, fijando los parámetros $R=1$ y $P=8$. Esta elección genera un descriptor local con ocho comparaciones alrededor del píxel central. El proceso de transformación de una imagen 3C a una 4C se describe en los siguientes cinco pasos.

1. **Entrada (Imagen Original RGB):** la imagen original cuenta con tres canales:

$$Imagen_{RGB}(x, y) = [R(x, y), G(x, y), B(x, y)]$$

Donde:

- $R(x, y)$, $G(x, y)$ y $B(x, y)$ representan la intensidad del color rojo, verde y azul, respectivamente, en el píxel ubicado en la posición (x, y) .
2. **Preprocesamiento - Conversión a Escala de Grises:** se calcula la imagen en escala de grises a partir de la imagen RGB utilizando una combinación lineal de los canales. Una conversión común es (Ecuación 4.1):

$$Imagen_{gris}(x, y) = 0.2990 \cdot R(x, y) + 0.5870 \cdot G(x, y) + 0.1140 \cdot B(x, y) \quad (\text{Ecuación 4.1})$$

Donde:

- $Imagen_{gris}(x, y)$ representa la intensidad en escala de grises en el píxel (x, y) .
 - $R(x, y)$, $G(x, y)$ y $B(x, y)$ son las intensidades de los canales rojo, verde y azul respectivamente.
 - Los valores numéricos (0.2990, 0.5870, 0.1140) provienen de estándares internacionales y estudios de percepción humana del color. En particular, son coeficientes basados en el modelo de conversión a escala de grises, definido por el estándar ITU-R BT.601-7 (2011), que a su vez está relacionado con la percepción del ojo humano hacia la luz y el color.
3. **Extracción de Textura - Calcular LBP:** se calcula el mapa de textura $L(x, y)$ aplicando el operador $LBP_{P,R}$ clásico (Sección 2.4.1) a la imagen en escala de gris $Imagen_{gris}$, utilizando una configuración de $P=8$ puntos vecinos y un radio $R=1$ (ver Ecuación 4.2).

$$L(x, y) = LBP_{P,R}(Imagen_{gris})(x, y) \quad (\text{Ecuación 4.2})$$

Donde:

- P es el número de puntos de muestreo del vecindario.
- R es el radio del vecindario.
- LBP indica la codificación clásica coherente con la Ecuación 2.1.

4. **Normalización del LBP:** el resultado del LBP se escala al rango $[0, 255]$ para que pueda ser interpretado como un canal de 8 bits, de forma análoga a los canales de color. Dividiendo por el valor máximo obtenido y multiplicando por 255 (ver Ecuación 4.3).

$$T(x, y) = \left(\frac{LBP(x, y)}{\max(LBP)} \right) \cdot 255 \quad (\text{Ecuación 4.3})$$

Donde:

- $LBP(x, y)$ es el valor del descriptor en el píxel (x, y) .
 - $\max(LBP)$ es el valor máximo posible del descriptor para el vecindario definido.
5. **Combinación de Canales (RGB + LBP):** finalmente, se incorpora el canal de textura como cuarto canal. La nueva imagen resultante queda como:

$$Imagen_{4c}(x, y) = [R(x, y), G(x, y), B(x, y), T(x, y)]$$

Donde:

- $T(x, y)$ corresponde al canal de textura obtenido a partir del LBP normalizado.
- La Imagen 4C se construye concatenando $R(x, y)$, $G(x, y)$, $B(x, y)$ y $T(x, y)$ para formar el tensor de entrada.

La conversión de la imagen RGB a un espacio de cuatro canales que otorga información adicional, apoyando a la descripción de la oreja. La representación del resultado de la transformación de la imagen 3C a 4C se muestra en la Figura 4.7.

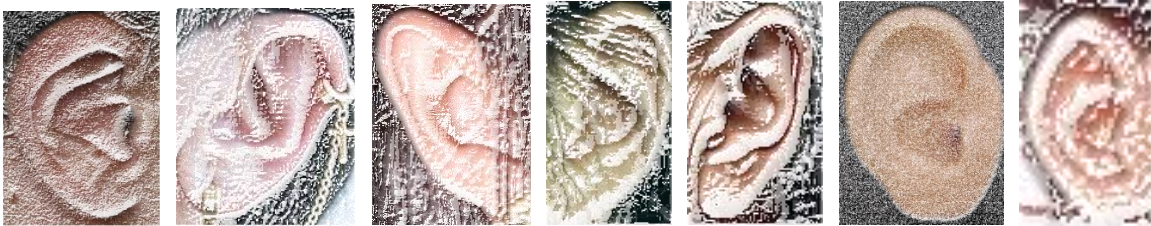


Figura 4.7 Representación del resultado de la transformación de la imagen 3C a 4C.

4.1.3. Estructura del Modelo 4C

Los modelos basados únicamente en el espacio de color RGB han reportado desempeños entre 93 % y 100 % de exactitud en tareas de reconocimiento facial y de oreja (Mo et al., 2022), (Park et al., 2022), (Xu et al., 2023).

Sin embargo, investigaciones más recientes que incorporan un canal adicional, por ejemplo, el canal alfa en imágenes RGBA han alcanzado hasta 98 % de exactitud (Espinoza-Vicuña & Nuñez-Medrano, 2023). El modelo propuesto se basa en agregar información a la imagen original desde un formato tradicional de tres canales (RGB) hacia una representación extendida de cuatro canales (RGB + LBP), con el propósito de incorporar información de

textura que refuerce la discriminación entre orejas humanas naturales y artificiales. La Figura 4.8 ilustra la secuencia del proceso propuesto, a partir de la (i) imagen original se delimita la (ii) ROI, se calcula el mapa (iii) LBP, se unen ambas representaciones para generar la imagen (iv) 4C y finalmente se procesa mediante la red (v) VGG19 para emitir la (vi) decisión final de clasificación.

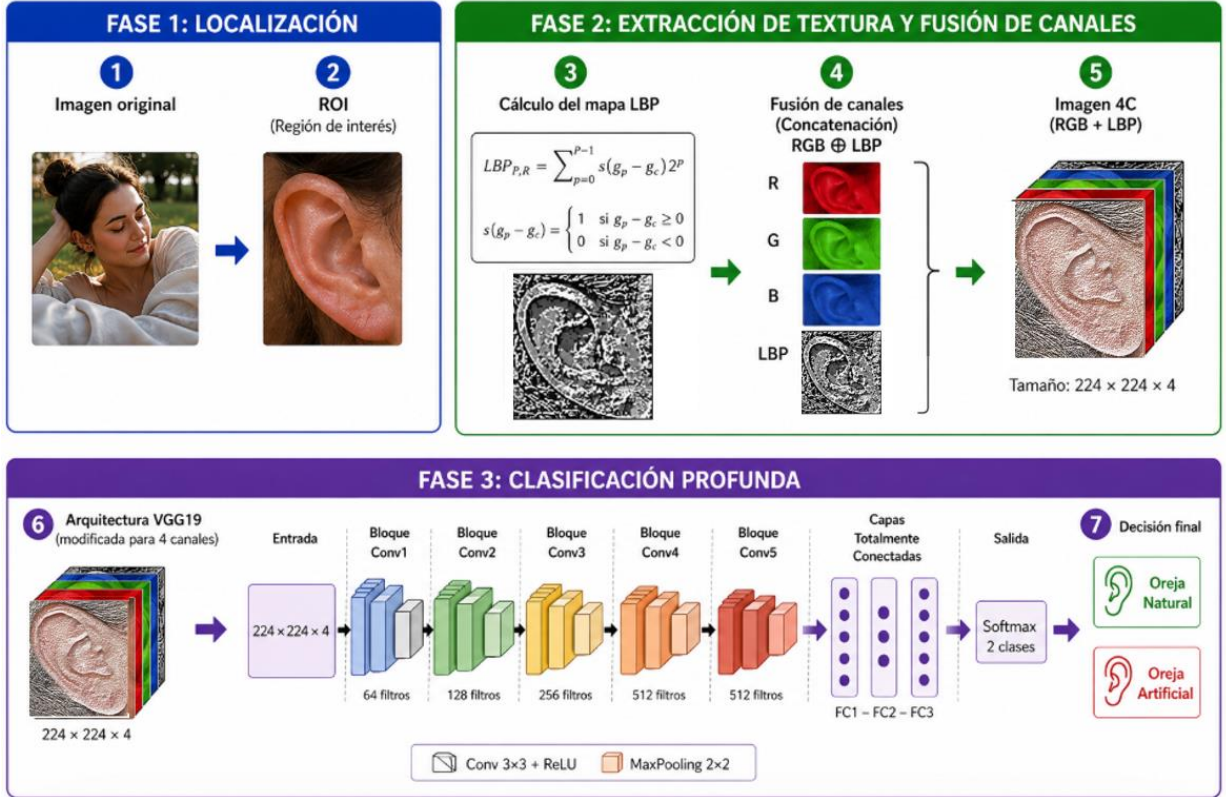


Figura 4.8 Flujo del Modelo 4C propuesto para la caracterización y clasificación de orejas humanas naturales y artificiales mediante una representación de cuatro canales (RGB + LBP) y la red VGG19.

El proceso inicia con la localización de la región de interés (ROI), correspondiente al área donde se encuentra la oreja. Una vez delimitada, se aplica el descriptor de textura *Local Binary Patterns* (LBP), el cual codifica las variaciones locales de intensidad en torno a cada píxel, generando un mapa de textura que se integra como un cuarto canal dentro de la imagen original RGB, conformando así la descripción 4C de la oreja.

Posteriormente, la imagen 4C se introduce en una arquitectura convolucional VGG19 modificada para aceptar cuatro canales de entrada. La red realiza la extracción profunda de características y la clasificación final, determinando si la oreja analizada corresponde a una muestra natural o artificial.

4.1.4. Cambios en la red VGG19

Se propuso la modificación del parámetro de entrada para la red VGG19 para que acepte cuatro canales. El entrenamiento se ejecutó con lote (*batch size*) de 32, durante 50 épocas,

con parada temprana (*patiencia*) de 5. La tasa de aprendizaje fue de $1e-05$ y el optimizador utilizado fue SGD con *momentum*=0.9. La función de pérdida empleada fue *CrossEntropyLoss*.

Las transformaciones de entrada incluyen ajuste de tamaño a (224, 224). Para VGG19 se reporta normalización con media [0.485, 0.456, 0.406] y desviación estándar [0.229, 0.224, 0.225]. La normalización específica usada para modelo 4C se presenta en la Tabla 6.12.

4.2. Configuración experimental

Con el fin de garantizar que la solución propuesta pueda ser reproducida, en esta sección se resume el protocolo experimental, incluyendo preprocesamiento, partición de datos y configuración de entrenamiento. La configuración completa de hiperparámetros se encuentra detallada en la Tabla 6.10.

4.2.1. Datos y partición

Para el entrenamiento y validación del clasificador PAD se utilizaron 2 clases: 'Artificial' y 'Natural'. En la experimentación se reporta un total de 5,400 muestras para entrenamiento y 1,440 muestras para validación. Para control de aleatoriedad se utilizó una semilla de 42.

4.2.2. Preprocesamiento y generación del cuarto canal (LBP)

Cada ROI de oreja se transforma a una representación de cuatro canales (RGB + textura) mediante el operador LBP. En particular, se aplicó *local_binary_pattern* (*img_gray*, $P=8$, $R=1$) sobre la imagen en escala de grises, y el resultado se normalizó para integrarse como cuarto canal de entrada al modelo.

4.2.3. Configuración de entrenamiento del clasificador

El modelo base utilizado fue VGG19; adicionalmente se propuso la modificación del parámetro de entrada para la red VGG19 para que acepte cuatro canales. El entrenamiento se ejecutó con lote (*batch size*) de 32, durante 50 épocas, con parada temprana (*patiencia*) de 5. La tasa de aprendizaje fue de $1e-05$ y el optimizador utilizado fue SGD con *momentum*=0.9. La función de pérdida empleada fue *CrossEntropyLoss*.

4.2.4. Entorno de ejecución

La experimentación se ejecutó con dispositivo 'cuda'. Las especificaciones del equipo de cómputo se reportan en la Tabla 6.1.

4.3. Herramientas de desarrollo

Para el desarrollo e implementación del modelo propuesto se emplearon diversas herramientas y bibliotecas del ecosistema Python, orientadas al entrenamiento de redes convolucionales basadas en la arquitectura VGG19. Las herramientas permitieron integrar procesos de entrenamiento, validación, evaluación de métricas y análisis de resultados.

A continuación, se mencionan las principales bibliotecas y entornos utilizados en la construcción del modelo:

OpenCV (OpenCV Team, 2022)

- Captura y procesamiento de video en tiempo real, lectura de la cámara, manipulación de imágenes y dibuja rectángulos o texto sobre las imágenes.
- Versión utilizada: 4.9.0

PyTorch (PyTorch Foundation, 2024)

- Permite el manejo de redes neuronales, inferencia y entrenamiento sobre CPU o GPU.
- En el proyecto de tesis se emplea para cargar el modelo final de localización, extraer características y realizar la predicción sobre la imagen de la oreja.

Ultralytics YOLOv5s (2023)

- Se emplea para detectar la oreja en caso de que MediaPipe no pueda localizarla con suficiente precisión.
- El modelo por defecto carga el archivo yolov5s.pt (un modelo pre-entrenado sobre COCO), en el caso del proyecto se utiliza un modelo personalizado ('best.pt') que solo contiene dos clases (oreja y no oreja).

scikit-image (local_binary_pattern de skimage.feature) (scikit-image development team, 2024; Van Der Walt et al., 2014)

- LBP es una técnica de descripción de texturas que captura patrones locales de intensidad. Se usa para calcular el patrón binario local (LBP) sobre la imagen en escala de grises. La característica (LBP) se combina con la imagen en color para formar una imagen de cuatro canales (R, G, B, LBP) que será la entrada a la red VGG19.

PIL (Python Imaging Library) y torchvision.transforms

- Se emplean para transformar la imagen a formato de tensor (definido como estructura de datos que generaliza el concepto de matrices a n dimensiones (Goodfellow et al., 2016)) y normalizarla antes de pasarla por el modelo entrenado.

CAPÍTULO 5

CONJUNTO DE IMÁGENES

El capítulo de conjunto de imágenes, detalla la información de las imágenes empleadas en la presente investigación, el cual se articula en torno a dos ejes principales: un banco de imágenes propio y una selección de conjuntos públicos de referencia. La pieza central de la etapa es la construcción del Conjunto de Imágenes de Orejas Humanas Naturales y Artificiales (CIOHNA), que puede ser utilizado en sistemas de detección de ataque de presentación (PAD) en biometría de la oreja, con el fin de evaluar el desempeño del sistema en ambientes controlados y no controlados, ya que en la investigación realizada sólo se encontraron conjuntos públicos de imágenes de orejas humanas naturales.

Muestras Artificiales: Compuestas por representaciones en dibujo, impresiones 3D en materiales como silicón y filamento rígido, así como imágenes realistas generadas mediante modelos avanzados de Inteligencia Artificial (DALL-E y Meta IA).

Muestras Naturales: Conjuntos públicos de referencia y captura propia utilizados para el análisis comparativo y el entrenamiento base del modelo. Incluyen escenarios controlados como los propuestos en los *datasets* AMI y IIT Delhi (I y II), caracterizados por condiciones de iluminación uniforme y posturas estandarizadas. Escenarios no controlados como el *dataset In-the-wild* (ITWE) y EarVN1.0, que aportan variabilidad en términos de oclusión, rotación y ruido ambiental. Captura propia de imágenes obtenidas mediante dispositivos de uso común como cámaras web y teléfonos móviles, para validar la operación en escenarios de "mundo real".

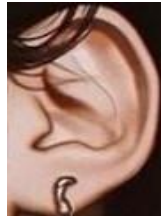
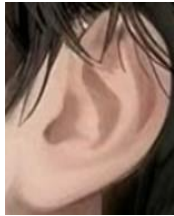
5.1. Muestras Artificiales - Instrumentos de Ataque

El conjunto integra 3,895 muestras de orejas humanas con resolución mínima de 16×21 píxeles, resolución máxima de 4000×3000 píxeles y promedio de ancho $\times$ alto de 1184×1650 píxeles, que incluye dibujos, fotografías de orejas impresas en silicón, e imágenes generadas con herramientas de inteligencia artificial (tanto realista como hiperrealistas) diseñadas para simular ataques de presentación. A continuación, se presenta la información de las categorías que integran a las muestras artificiales.

5.1.1. Dibujo

Para el trabajo de tesis una imagen considerada en la categoría de *dibujo*, es una representación simplificada o exagerada de la realidad, utilizando las formas geométricas, colores y rasgos del objeto. La mayoría de las imágenes consideradas son dibujos generados con herramientas de inteligencia artificial.

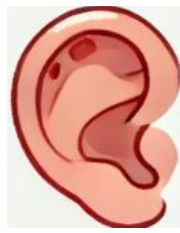
Conjunto compuesto de 703 imágenes de dibujo con resolución mínima de 38×60 píxeles, resolución máxima de 809×1347 píxeles y promedio de ancho $\times$ alto de 149×234 píxeles, descargadas de diferentes fuentes de internet como la página de Pinterest (2024) y la herramienta ChatGPT (OpenAI, 2024). La colección cuenta con muestras en diferentes situaciones, como son: oclusión, rotación, cambios en la intensidad luminosa, escala y uso o ausencia de accesorios. En la Figura 5.1 se observan ejemplos de las muestras artificiales en la categoría de dibujo.



a) ejemplos de imágenes con oclusión



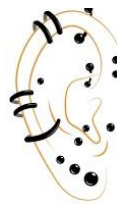
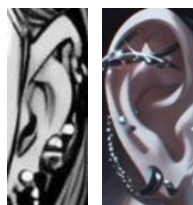
b) ejemplos de imágenes con rotación



c) ejemplos de imágenes con cambios en la intensidad luminosa



d) ejemplos de imágenes con variación de tamaño



e) ejemplos de imágenes con accesorios

Figura 5.1 Ejemplos de la distribución de las imágenes de orejas humanas en dibujo.

5.1.1.1. Internet

El grupo se integra de 563 imágenes de orejas humanas descargadas desde diferentes sitios entre ellos Pinterest (2024). En la Figura 5.2 se pueden observar algunos ejemplos.



Figura 5.2 Ejemplos de imágenes de orejas humanas descargadas desde internet.

Como se observa en los ejemplos de la Figura 5.2, el subconjunto se distingue por presentar variabilidad no estructurada, lo que introduce desafíos significativos para la detección y clasificación automática.

5.1.1.2. Generador de Imágenes (Image Generator)

ChatGPT es una herramienta de inteligencia artificial desarrollada por OpenAI (2024). Está diseñada para generar respuestas similares a las humanas en conversaciones escritas. Entre sus habilidades cuenta con aplicaciones que permiten generar imágenes de cualquier tipo.

El grupo está conformado por 240 muestras de imágenes de orejas humanas en dibujo, generadas desde la página de ChatGPT (OpenAI, 2024) con Generador de imágenes (*imagen generator*) especializado en generar y refinar imágenes con una combinación de tono profesional y amigable. Algunos ejemplos se observan en la Figura 5.3.



Figura 5.3 Ejemplos de imágenes de orejas humanas artificiales generadas por IA (OpenAI, 2024).

5.1.2. Impresión 3D: Filamento Rígido y Silicón

El grupo de impresión 3D está integrado por un total de 1,646 imágenes con resolución mínima de 843×1311 píxeles, resolución máxima de 4000×3000 píxeles y promedio de ancho $\times$ alto de 2188×3053 píxeles, capturadas mediante los dispositivos móviles Redmi Note 11 y OPPO Find N3 Flip. Las muestras se dividen en dos categorías según el material de fabricación: filamento rígido y silicón.

1. **Filamento Rígido:** Corresponde a modelos fabricados mediante impresión por deposición de material fundido (FDM). Las muestras presentan una estructura sólida y una superficie con patrones de líneas de capa visibles, además de una reflectancia especular más alta (brillo plástico) y colores artificiales no realistas. Son útiles para validar la detección basada puramente en geometría y forma (Soares et al., 2024).
2. **Silicón:** Las muestras emulan las propiedades de los tejidos blandos. A diferencia del filamento, el silicón ofrece una textura suave y homogénea, con una coloración y opacidad que simulan la dispersión de luz en la piel humana. Lo que añade un nivel de complejidad mayor al acercarse a la apariencia orgánica (Wagner et al., 2024).

La adquisición de imágenes de ambos materiales, se realizó contemplando variaciones controladas de rotación, iluminación y escala. En la Figura 5.4 se presentan ejemplos de imágenes de orejas humanas en impresión 3D.



a) ejemplos de imágenes con rotación



b) ejemplos de imágenes con cambios en la intensidad luminosa



c) ejemplos de imágenes con variación de tamaño

Figura 5.4 Ejemplos de imágenes de orejas humanas en impresión 3D: filamento rígido y silicón.

5.1.2.1. Filamento Rígido

Se tomaron 530 muestras de impresiones 3D en filamento rígido, algunos ejemplos se observan en la Figura 5.5. Corresponde a modelos fabricados mediante impresión por deposición de material fundido (Soares et al., 2024).



Figura 5.5 Ejemplos de imágenes en impresión 3D con filamento rígido

5.1.2.2. Silicón

Se tomaron 1,116 muestras de impresiones 3D en silicón, algunos ejemplos se observan en la Figura 5.6.

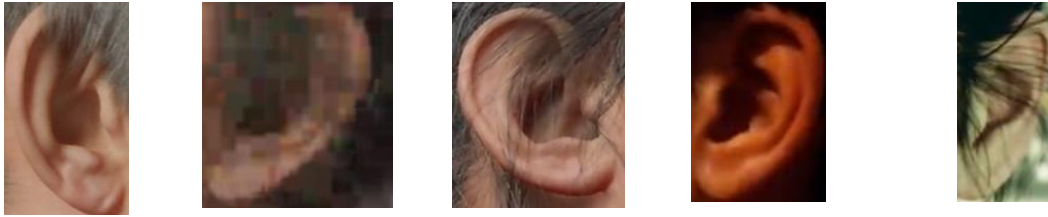


Figura 5.6 Ejemplos de imágenes en impresión 3D con silicón

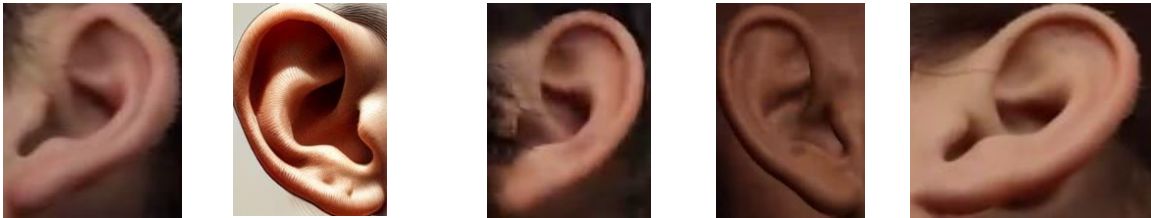
5.1.3. Imágenes sintéticas con apariencia realista

El conjunto de las imágenes sintéticas con apariencia realista está integrado por la generación de 1,116 muestras con dos herramientas de inteligencia artificial: Meta humano (*MetaHuman Creator*) y DALL-E de ChatGPT, con resolución mínima de 16×21 pixeles, resolución máxima de 1258×712 pixeles y promedio de ancho $\times$ alto de 606×674 pixeles en diferentes condiciones, como son: oclusión parcial por cabello u otros elementos, cambios de rotación o pose, variaciones en la intensidad de iluminación, diferencias de tamaño y presencia de accesorios cerca de la región de la oreja.

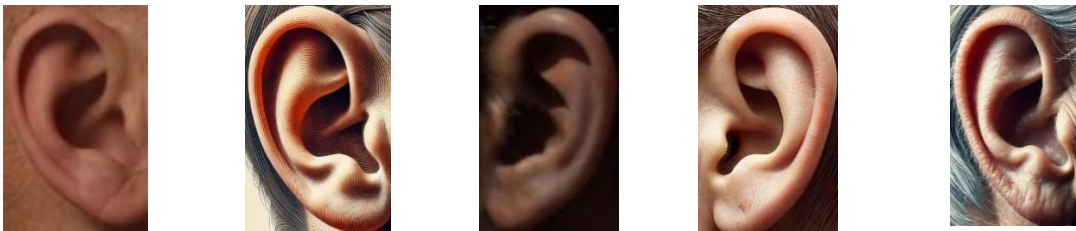
En la Figura 5.7 se observan ejemplos de imágenes realistas generadas con IA.



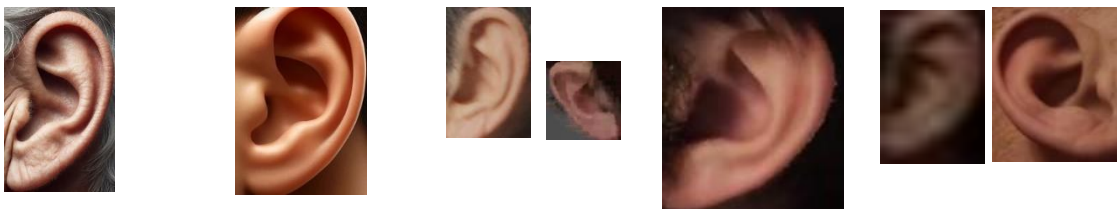
a) ejemplos de imágenes con oclusión



b) ejemplos de imágenes con diferente rotación



c) ejemplos de imágenes con cambios en la intensidad luminosa



d) ejemplos de imágenes con variación de tamaño



e) ejemplos de imágenes con accesorios

Figura 5.7 Ejemplos de imágenes de orejas humanas realistas generadas por IA (Epic Games, 2024).

5.1.3.1. MetaHuman Creator

MetaHuman Creator es una herramienta gratuita desarrollada por *Epic Games* (2024) que está disponible en internet, que permite crear imágenes de personas. Es una aplicación que ofrece a los usuarios la capacidad de diseñar y caracterizar humanos con detalles faciales y corporales. Con *MetaHuman Creator*, se puede personalizar el peinado, los rasgos faciales, la altura, las proporciones corporales y más de un avatar digital (Epic Games, 2024).

Se generaron 876 imágenes de orejas humanas realistas con la IA *MetaHuman Creator*, algunos ejemplos se observan en la Figura 5.8.



Figura 5.8 Ejemplos de imágenes de orejas humanas realistas generadas con *MetaHuman Creator*. (Epic Games, 2024).

5.1.3.2. DALL-E

En el grupo de DALL-E se generaron 240 imágenes sintéticas de orejas humanas con apariencia realista mediante la herramienta disponible en ChatGPT (OpenAI, 2024). DALL-E es una herramienta capaz de interpretar instrucciones detalladas y crear representaciones visuales que corresponden al texto que recibe como entrada.

A continuación, se presentan dos ejemplos de prompts utilizados para la generación de imágenes sintéticas con apariencia realista:

1. **Oreja humana con apariencia realista:** “Genera una imagen realista de una oreja humana vista de lado, con iluminación natural, fondo neutro y alto nivel de detalle en la piel.”
2. **Oreja humana de persona adulta:** “Genera una imagen realista de una oreja humana correspondiente a una persona adulta mayor de 25 años, con tono de piel claro, vista de perfil, fondo simple y con definición en los pliegues de la oreja.”

En la Figura 5.9 se presentan ejemplos de imágenes sintéticas generadas con DALL-E.

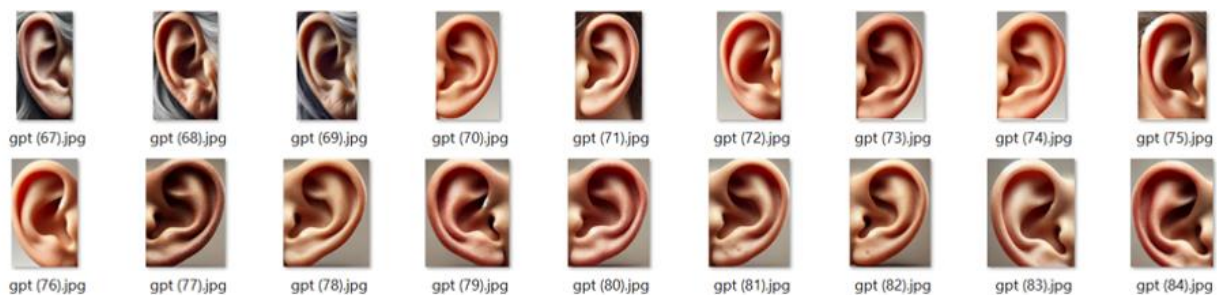
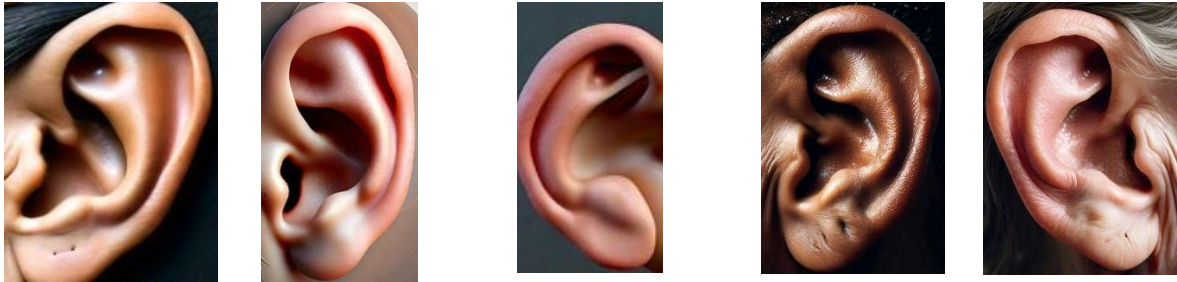


Figura 5.9 Ejemplos de imágenes sintéticas de orejas humanas con apariencia realista generadas mediante DALL-E en ChatGPT (OpenAI, 2024).

5.1.4. Imágenes hiperrealistas generadas con Meta IA Generativa

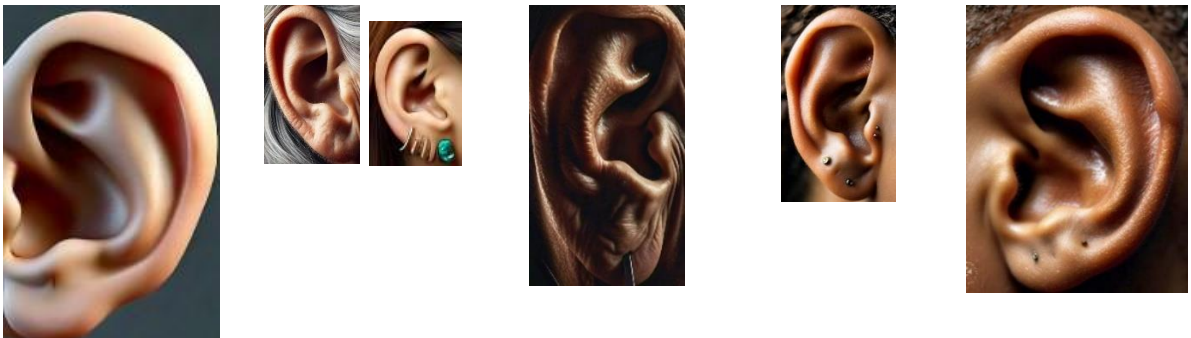
En el conjunto de imágenes hiperrealistas se generaron 330 muestras de imágenes de orejas humanas hiperrealistas mediante Meta IA Generativa de WhatsApp (2024) para crear imágenes, videos y textos. Algunos ejemplos se observan en la Figura 5.10.



b) ejemplos de imágenes con diferente perspectiva



c) ejemplos de imágenes con cambios en la intensidad luminosa



d) ejemplos de imágenes con variación de tamaño



e) ejemplos de imágenes con accesorios

Figura 5.10 Ejemplos de imágenes de orejas humanas hiperrealistas creadas con Meta IA Generativa de WhatsApp (Meta, 2024).

La Tabla 5.1 presenta un resumen de la creación y recopilación de imágenes de orejas humanas artificiales. Los datos se distribuyen en cuatro grupos principales según la naturaleza y método de obtención: dibujo, impresión 3D, imágenes realistas generadas con inteligencia artificial (IA) e imágenes hiperrealistas generadas con IA.

Tabla 5.1 Compendio del conjunto imágenes de orejas humanas artificiales

#	Subgrupo	# imágenes	Fuente	Características
Grupo Dibujo – 803 muestras				
1	Internet	563	Internet	Imágenes descargadas mediante el navegador; incluyen variaciones como oclusión, rotación, cambios en intensidad luminosa, tamaño, y accesorios.
2	Generador de imágenes	240	ChatGPT (<i>Image Generator</i>) (OpenAI, 2024)	Generadas con inteligencia artificial; incluyen características como oclusión, rotación, cambios en iluminación, tamaño, y accesorios.
Grupo Impresión 3D – 1,646 muestras				
3	Filamento Rígido	530	Fotografías tomadas con Redmi Note 11 y OPPO Find N3 Flip	Imágenes de orejas impresas en 3D con filamento rígido, con variaciones en rotación, iluminación y tamaño.
4	Silicón	1,116		Imágenes de orejas impresas en 3D en silicón, con variaciones en rotación, iluminación y tamaño.
Grupo IA - Imágenes sintéticas con apariencia realista – 1,116 muestras				
5	<i>MetaHuman Creator</i>	876	<i>MetaHuman Creator (Epic Games, 2024)</i>	Generadas con detalles realistas, permitiendo personalización facial y corporal, con variaciones de iluminación, tamaño, accesorios, etc.
6	DALL-E	240	ChatGPT con DALL-E (OpenAI, 2024)	Imágenes generadas a partir de descripciones de texto, con variaciones en iluminación, rotación, tamaño, y accesorios.
Grupo IA - Imágenes hiperrealistas				
7	Meta IA Generativa	330	Meta IA Generativa en WhatsApp (Meta, 2024)	Imágenes hiperrealistas generadas con IA, incluyendo variaciones en perspectiva, iluminación, tamaño, y accesorios.
	Total	3,895 imágenes		

La combinación de técnicas tradicionales (dibujos e impresiones 3D) y tecnología avanzada (generación por IA) permitió abordar una brecha existente en la disponibilidad de datos de orejas humanas artificiales. La construcción del conjunto de imágenes constituye una contribución de la presente tesis, al abordar una de las limitaciones identificadas en el estado del arte: la falta de datos específicos para PAD en biometría de la oreja (Toprak & Toygar, 2021).

5.2. Orejas Humanas Naturales

La segunda colección cuenta 31,876 imágenes provenientes de conjuntos públicos en ambientes controlados como *AMI* y *IIT Delhi* hasta imágenes obtenidas en entornos reales (*EarVN1.0*, *In-the-Wild*) o con dispositivos de uso común como cámaras web. En conjunto presentan diferentes características y condiciones, con resolución mínima de 8 x 22 píxeles, resolución máxima de 5616 × 3744 píxeles y con promedio de ancho x alto de

713 × 730 píxeles. La diversidad es importante para el desarrollo y validación de algoritmos de reconocimiento que necesiten manejar tanto ambientes controlados como no controlados.

5.2.1. Análisis Matemático de Imágenes (AMI)

El conjunto AMI (Gonzalez et al., 2012) contiene imágenes de orejas humanas capturadas en condiciones controladas. Específicamente diseñado para aplicaciones biométricas, las imágenes muestran orejas, con iluminación uniforme y fondos simples. Se recopilaron en la Universidad de Las Palmas, contiene 700 imágenes de orejas humanas naturales de 100 adultos caucásicos de ambos sexos de entre 19 y 65 años, con resolución mínima de 492 × 702 píxeles, resolución máxima de 492 × 702 píxeles y con promedio de ancho x alto de 492 × 702 píxeles. Todas las imágenes se tomaron con la misma iluminación y con una posición de cámara fija. Se capturaron ambas orejas (Booysens & Viriri, 2022). En la Figura 5.11 se muestran ejemplos de las imágenes.

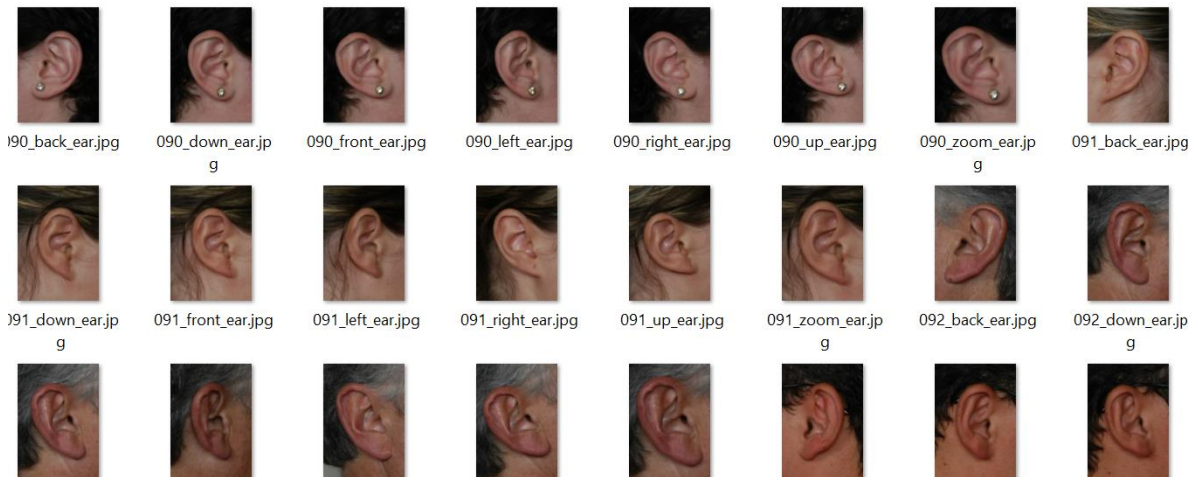


Figura 5.11 Ejemplos de imágenes del conjunto AMI.

5.2.2. EarVN1.0

El conjunto EarVN1.0 (Truong Hoang, 2019) incluye diferentes ángulos, condiciones de iluminación. El conjunto incluye oclusiones parciales y variaciones naturales en las orejas humanas (Truong Hoang, 2019). Incluye 28,412 imágenes de 164 sujetos asiáticos masculinos y femeninos, se capturó la oreja izquierda y derecha, con resolución mínima de 13 × 17 píxeles, resolución máxima de 436 × 581 píxeles y con promedio de ancho x alto de 89 × 131 píxeles. Se recopiló durante 2018 y se formó a partir de condiciones sin restricciones, incluidos los sistemas de cámara y las condiciones de iluminación.

Las imágenes presentan variaciones como oclusión, cambios en rotación, cambios en la intensidad luminosa, tamaño o presencia de accesorios (Booysens & Viriri, 2022) (ver Figura 5.12).



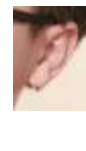
a) ejemplos de imágenes con oclusión



b) ejemplos de imágenes con diferente rotación



c) ejemplos de imágenes con cambios en la intensidad luminosa



d) ejemplos de imágenes con variación de tamaño



e) ejemplos de imágenes con accesorios

Figura 5.12 Ejemplos de imágenes de orejas humanas naturales del conjunto EarVN1.0 (Truong Hoang, 2019).

5.2.3. IIT Delhi

El conjunto *IIT Delhi* (2004) es uno de los más antiguos y ampliamente utilizados para estudios de biometría. Sus imágenes se capturan en condiciones controladas, generalmente son sujetos estáticos y en posiciones ideales para aplicaciones de reconocimiento. *IIT Delhi* está integrada por 1,779 imágenes de 121 adultos distintos, tanto hombres como mujeres. Las imágenes se tomaron en un ambiente controlado, sin oclusiones significativas presentes, y sólo se capturó la oreja derecha de la persona. Las imágenes están en escala de grises con un tamaño de 272×204 píxeles (Booysens & Viriri, 2022). En la Figura 5.13 se muestran ejemplos de las imágenes.

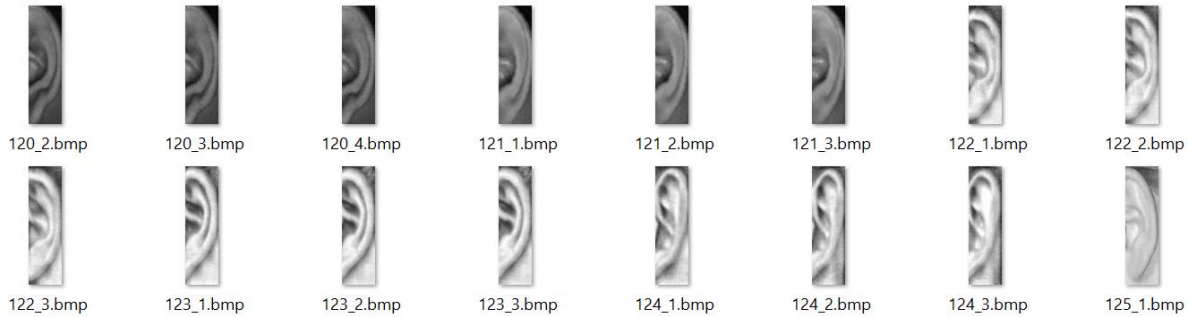


Figura 5.13 Ejemplos de imágenes del conjunto IITD.

5.2.4. ITWE (In-the-Wild-Ear)

El conjunto se caracteriza por contener orejas humanas naturales en escenarios no controlados. Incluye imágenes tomadas en diferentes contextos, como exteriores e interiores, con variaciones de oclusión, rotación, cambios en la intensidad luminosa, tamaño y accesorios. Ideal para entrenar sistemas robustos de reconocimiento de orejas en entornos reales (Zhou & Zaferiou, 2017).

Fue creado para la evaluación de reconocimiento y tiene 2,663 imágenes en total con resolución mínima de 288×91 píxeles, resolución máxima de 5616×3744 píxeles y promedio de ancho $\times$ alto de 713×730 píxeles, incluyendo 231 sujetos masculinos y femeninos (Booysens & Viriri, 2022). En la Figura 5.14 se muestran ejemplos de las imágenes.



a) ejemplos de imágenes con oclusión



b) ejemplos de imágenes con diferente rotación



c) ejemplos de imágenes con cambios en la intensidad luminosa

Figura 5.14 Ejemplos de imágenes de orejas humanas naturales del conjunto ITWE.



d) ejemplos de imágenes con variación de tamaño



e) ejemplos de imágenes con accesorios

Figura 5.14 Ejemplos de imágenes de orejas humanas naturales del conjunto ITWE (continuación).

5.2.5. Cámara web

El conjunto de imágenes obtenidas con la cámara web se compone de 285 muestras con resolución mínima de 33×69 píxeles, resolución máxima de 150×170 píxeles y promedio de ancho x alto de 140×160 píxeles, que presentan variaciones significativas en escala y niveles de oclusión, características que lo hacen idóneo para evaluar aplicaciones en entornos no controlados, como el reconocimiento biométrico de personas. La adquisición de las imágenes se realizó mediante un sensor USB2.HD UVC WebCam, con el consentimiento informado de los participantes. Las imágenes son utilizadas exclusivamente con fines de investigación académica.

La Figura 5.15 ilustra ejemplos obtenidos durante la validación de la etapa de localización. En dichas imágenes, los marcadores visuales denotan la posición del trago: el punto azul corresponde a la oreja izquierda, mientras que el punto amarillo a la oreja derecha. La población de estudio consistió en 15 sujetos (10 hombres y 5 mujeres).

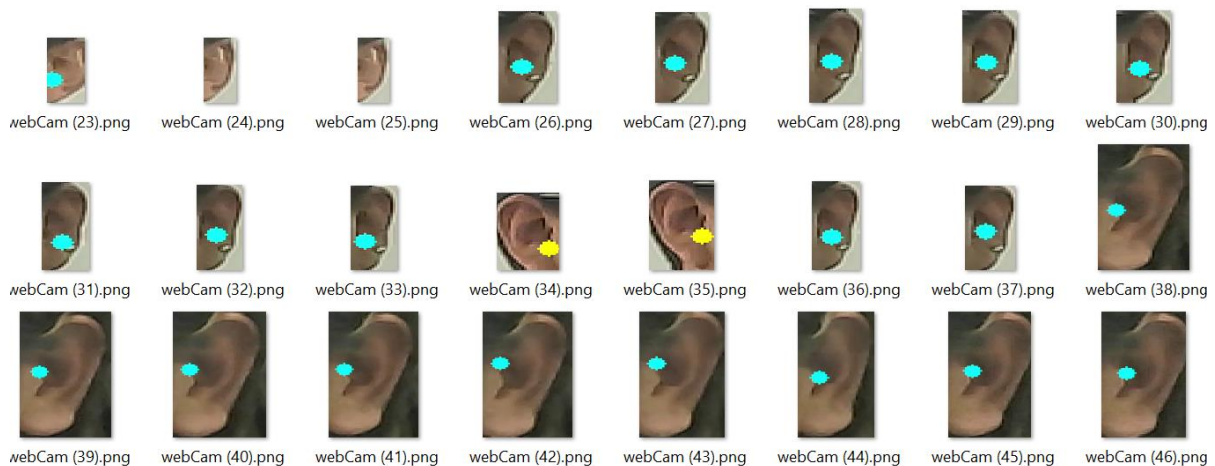


Figura 5.15 Ejemplos de imágenes tomadas con la cámara web.

La selección de los conjuntos de imágenes de orejas humanas naturales fue básica para el desarrollo y validación del proyecto de tesis. Los conjuntos varían en sus condiciones de captura (controladas o no controladas), las condiciones de iluminación, oclusión, rotación y otros factores. En la Tabla 5.2 se presenta un resumen de los conjuntos de imágenes de orejas humanas naturales utilizados en el proyecto de tesis. Cada conjunto aporta características únicas, desde los entornos controlados a los de ambiente real, pasando por colecciones con variabilidad demográfica, diferentes condiciones de iluminación, presencia de accesorios, oclusiones parciales y variaciones en el ángulo de captura.

Tabla 5.2 Compendio del conjunto imágenes de orejas humanas naturales

#	Conjunto	# imágenes	Fuente	Características
1	AMI	700	(Gonzalez et al., 2012)	Imágenes en condiciones controladas, iluminación uniforme y fondo simple.
2	EarVN1.0	28,412	(Truong Hoang, 2019)	Imágenes con diferentes niveles de iluminación, rotación, oclusión y accesorios.
3	IIT Delhi	1,779	(IIT Delhi Ear Database, 2004)	Uno de los conjuntos más antiguos, ampliamente utilizado, condiciones controladas, sujetos estáticos.
4	ITWE	2,663	(Zhou & Zaferiou, 2017)	Escenarios no controlados, variaciones de iluminación, tamaño, accesorios, oclusión y rotación.
5	Cámara web	285	USB2.HD UVC WebCam	Imágenes que presentan variaciones en ángulo, iluminación y distancia.
	TOTAL	33,839 imágenes		

La diversidad en las condiciones de captura, la demografía de los participantes y la variedad de factores ambientales presente en la colección de imágenes fue básico para diseñar, entrenar y evaluar el modelo propuesto capaz de funcionar en distintas situaciones.

En este capítulo se describió la conformación del conjunto de imágenes CIOHNA utilizado en la presente investigación, integrado por muestras de orejas humanas naturales y artificiales. Las imágenes artificiales incluyeron dibujos, impresiones 3D en distintos materiales e imágenes generadas mediante herramientas de inteligencia artificial, mientras que las imágenes naturales provinieron de conjuntos públicos de referencia y capturas propias. La integración permitió construir un banco de imágenes con diversidad visual, variabilidad de escala, diferencias de iluminación, rotación, oclusión y distintas condiciones de adquisición.

Con base en este conjunto de imágenes, el siguiente capítulo presenta la configuración experimental y el proceso seguido para la evaluación del modelo propuesto, incluyendo el preprocesamiento aplicado, la generación de la entrada utilizada por la red neuronal y las métricas empleadas para analizar su desempeño en la clasificación de orejas naturales y artificiales.

CAPÍTULO 6

PRUEBAS Y RESULTADOS

En el presente capítulo se detalla el entorno experimental y los resultados obtenidos en la validación de la propuesta de tesis. El proceso de evaluación se divide en: la verificación del método de localización de la región de interés y la validación del Modelo 4C en tareas de reconocimiento y detección de ataques de presentación (*PAD*).

Para el análisis de la localización, se establecen tres casos de prueba, describiendo la composición de los conjuntos de datos empleados tanto en la fase de entrenamiento como de validación. Posteriormente, se presentan los resultados cuantitativos obtenidos en cada caso.

Asimismo, se somete a prueba el Modelo 4C con seis casos de estudio, incluyendo un análisis comparativo frente a métodos del estado del arte, con especial énfasis ante intentos de vulneración. Finalmente, el capítulo concluye con una discusión sobre los resultados de las pruebas.

6.1. Ambiente de pruebas

Las pruebas se realizaron en un equipo con las características que se describen en la Tabla 6.1.

Tabla 6.1 Especificaciones del equipo de cómputo utilizado en las pruebas

Componente	Descripción
Procesador	13th Gen Intel(R) Core (TM) i7-13650HX 2.60 GHz
RAM instalada	16.0 GB (15.6 GB usable)
Tipo de sistema	Sistema operativo de 64 bits, procesador basado en x64
Modelo GPU	NVIDIA GeForce RTX 4060 Laptop GPU
Memoria GPU dedicada (VRAM)	8 GB

6.2. Evaluación del método de localización

Para la evaluación del desempeño de la etapa de localización se utilizaron las métricas clásicas establecidas en el Capítulo 2 (Marco Teórico) apartado 2.6.2 (Métricas de clasificación) considerando tres casos de prueba:

- 1) **Entrenamiento de YOLOv5s.** El objetivo es presentar los resultados del entrenamiento y validación de la red YOLOv5s.

- 2) **Comparativa de localización de orejas humanas.** El objetivo es evaluar el desempeño de MediaPipe y YOLOv5s (por separado) para localizar orejas humanas en ambientes controlados y no controlados.
- 3) **Método de localización propuesto.** El objetivo es evaluar el modelo propuesto (la combinación de MediaPipe y YOLOv5s) para la localizar orejas humanas de objetos diferentes (no orejas) en ambientes controlados y no controlados.

6.2.1. Entrenamiento de YOLOv5s

El primer caso tiene como objetivo evaluar YOLOv5s durante el entrenamiento. El conjunto de imágenes para el caso uno está formado por 3,988 imágenes dividido en dos grupos: orejas humanas y no orejas humanas.

El grupo de orejas humanas está integrado por 2,400 imágenes de conjuntos públicos que contienen rostros en ambientes reales, que presentan diferentes perspectivas del mismo (de frente, perfil y con diferentes rotaciones), con cambios en la escala, iluminación y traslape. También contiene imágenes de oreja alineada y recortada. El grupo de no orejas humanas contiene 1,588 imágenes de objetos similares a una oreja humana, así como; objetos circulares, plantas, animales y frutas (ver Tabla 6.2). En la Figura 6.1 pueden ver muestras de los conjuntos de imágenes.

Tabla 6.2 Distribución de los conjuntos de imágenes para el entrenamiento de YOLOv5s

IMÁGENES DE OREJA HUMANA				
#	Imágenes	Origen	Muestras	Total
1	Naturales (Conjuntos públicos)	In-the-wild Ear Database (Zhou & Zaferiou, 2017)	500	2,400
		IIT Delhi Ear Database (IIT Delhi Ear Database, 2004)	400	
		AMI Ear Database (Gonzalez et al., 2012)	350	
		EarVN1.0 Dataset (Truong Hoang, 2019)	200	
2	Artificiales	Generadas mediante Meta Human (Epic Games, 2024)	500	
		Descargadas del internet	280	
		Tomadas con el celular Redmi Note 11	170	
IMÁGENES DE NO OREJA HUMANA				
1	Plantas, frutas, hongos	Descargadas del internet	464	1,588
2	Animales y accesorios		1,124	



Figura 6.1 Muestras del conjunto de imágenes de entrenamiento YOLOv5s
a) oreja humana y b) no oreja humana.

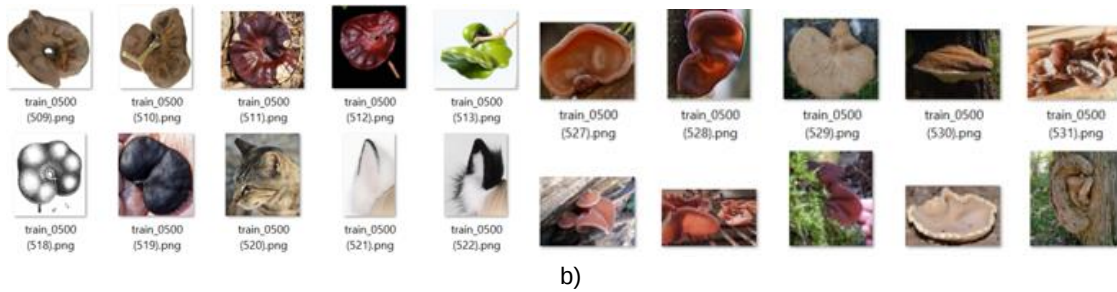


Figura 6.1 Muestras del conjunto de imágenes de entrenamiento YOLOv5s
a) oreja humana y b) no oreja humana (continuación)

La Tabla 6.3 presenta los resultados obtenidos durante el entrenamiento del modelo YOLOv5s para la localización de orejas humanas y la identificación de muestras correspondientes a la clase No oreja. Se reportan las métricas de Precisión, *Recall* y F1-score para cada clase, con el propósito de evaluar el comportamiento del modelo ante regiones de interés positivas y objetos que no corresponden a orejas.

Tabla 6.3 Resultados del entrenamiento de YOLOv5s.

Clases	Precisión (%)	Recall (%)	F1-Score (%)
Oreja	96.90	100	98.40
No oreja	53.80	45.20	49.10

Como se observa en la Tabla 6.3, la clase Oreja obtuvo un desempeño alto, con Precisión de 96.90 %, *Recall* de 100 % y F1-score de 98.40 %. Esto indica que el modelo logró localizar correctamente la totalidad de las orejas presentes en el conjunto evaluado, manteniendo además una baja proporción de detecciones incorrectas para esta clase. En contraste, la clase No oreja presentó valores menores, con Precisión de 53.80 %, *Recall* de 45.20 % y F1-score de 49.10 %. Estos resultados muestran que el modelo tiene mayor dificultad para discriminar objetos que no son orejas, especialmente cuando presentan formas, bordes o texturas visualmente similares a una oreja humana. La Figura 6.2 muestra ejemplos de la validación realizada durante el entrenamiento de YOLOv5s.

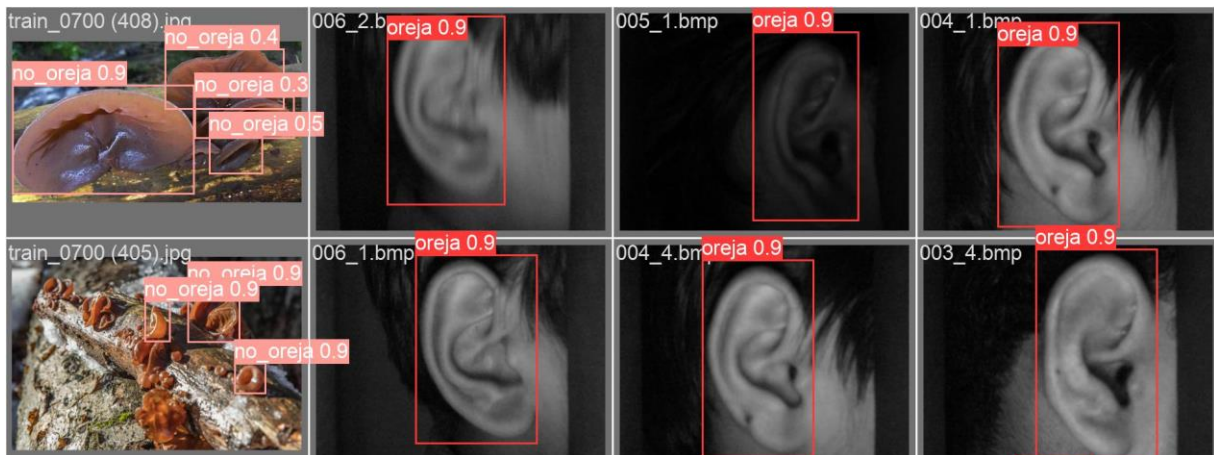


Figura 6.2 Muestras de la validación dentro de entrenamiento de YOLOv5s.

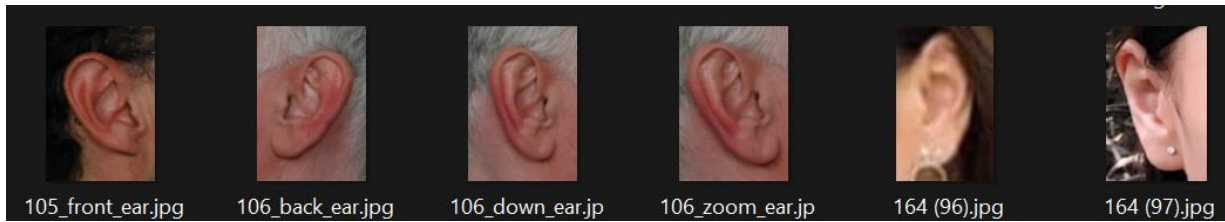
6.2.2. Comparativa de localización de orejas humanas

El objetivo de esta prueba es evaluar de forma separada el desempeño de MediaPipe y YOLOv5s para la localización del área de interés, utilizando el mismo conjunto de imágenes y bajo las mismas condiciones experimentales.

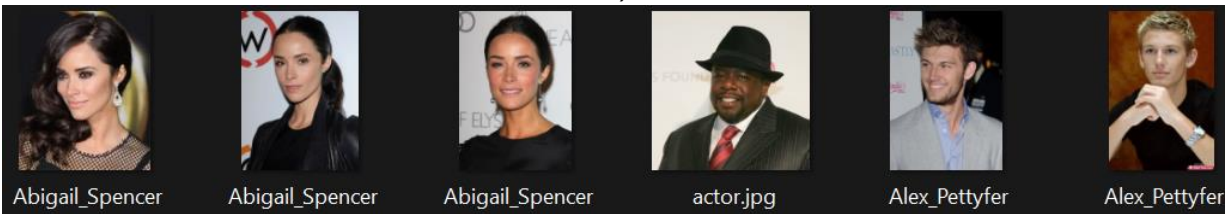
El *dataset* está integrado por muestras que contienen una oreja humana, tomadas de conjuntos públicos y descargadas del internet, donde la mitad está en un ambiente controlado y la otra mitad en ambiente no controlado, ver Tabla 6.4. Algunos ejemplos se ilustran en la Figura 6.3.

Tabla 6.4 Distribución de los conjuntos de imágenes utilizados en la validación

IMÁGENES DE OREJA HUMANA					
#	Imágenes	Origen	Muestras	Total	
1	Controlado	AMI Ear Database (Gonzalez et al., 2012)	45	180	360
2		IIT Delhi (<i>IIT Delhi Ear Database</i> , 2004)	45		
3		EarVN1.0 (Truong Hoang, 2019)	45		
4		Tomadas con el celular	45		
5	No controlado	In-the-wild Ear Database	150	180	
6		Descargadas de internet	30		



a)



b)

Figura 6.3 Muestras del conjunto de imágenes para la validación a) oreja humana ambiente controlado b) oreja humana ambiente no controlado

El resultado de las pruebas de MediaPipe y YOLOv5s en ambientes controlados y no controlados se muestran en la Tabla 6.5, donde la métrica a obtener es Exactitud.

Tabla 6.5 Resultados de evaluar por separado a MediaPipe y YOLOv5s para imágenes que contienen una oreja humana en ambientes controlados y no controlados.

Ambiente	Muestras	MediaPipe			YOLOv5s		
		VP	FN	Exactitud (%)	VP	FN	Exactitud (%)
Controlado	180	0	180	0	169	11	93.89
No controlado	180	179	1	99.44	155	25	86.11

Con respecto al conjunto de imágenes donde existe una oreja humana en un ambiente no controlado, se observa que el resultado de MediaPipe supera al resultado de YOLOv5s, ya que sólo se equivoca una vez obteniendo 99.44 % de exactitud con 0.66 % de error.

En cambio, en ambiente controlado YOLOv5s obtiene el valor más alto al localizar 169 orejas de 180 con 93 % de Exactitud y 11 Falsos Positivos con un error del 7 %. El resultado se esperaba, debido a que MediaPipe no logra ubicar ninguno de los puntos necesarios para calcular la región donde potencialmente puede estar presente una oreja humana. Algunos ejemplos se pueden ver en la Figura 6.4 y 6.5

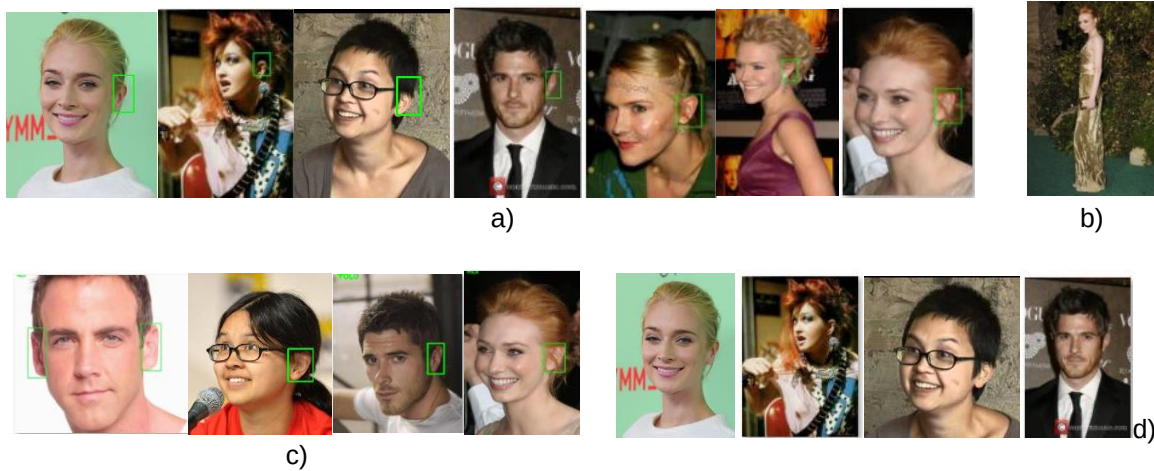


Figura 6.4 Muestras de imágenes en la evaluación en ambiente no controlado a) MediaPipe localiza la oreja b) MediaPipe no localiza la oreja, c) YOLOv5s localiza la oreja y d) YOLOv5s no localiza oreja.

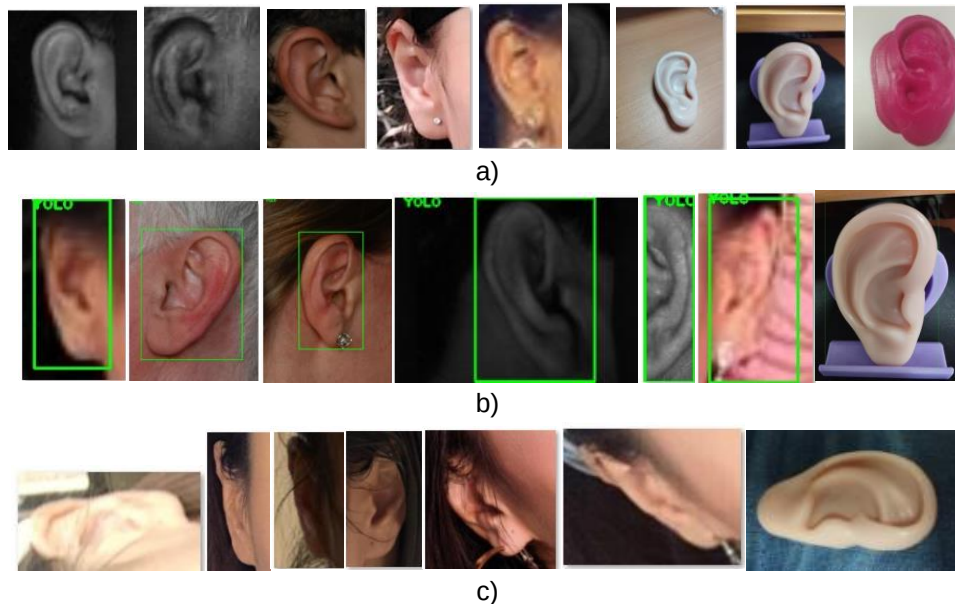


Figura 6.5 Muestras de imágenes en la evaluación en ambiente controlado a) MediaPipe no localiza la oreja, b) YOLOv5s no localiza la oreja y c) YOLOv5s no localiza la oreja.

En la Figura 6.6 se muestran ejemplos de imágenes donde YOLOv5s no es capaz de detectar el área de interés, mientras MediaPipe lo hace de manera correcta.

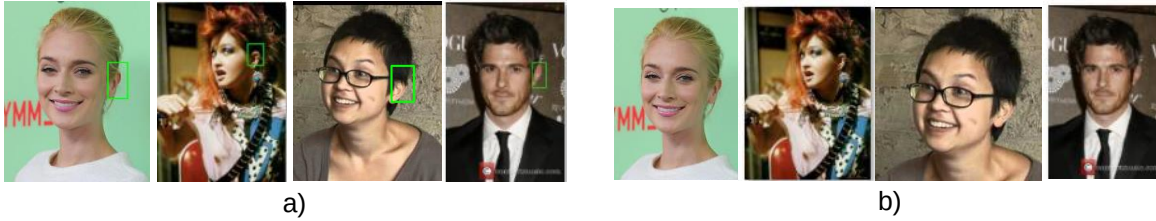


Figura 6.6 Muestras del resultado de la localización de la oreja humana a) MediaPipe y b) YOLOv5s.

6.2.3. Método de localización propuesto

El método propuesto busca localizar y delimitar automáticamente la región de interés en cada imagen del conjunto, segmentando el área donde existe una alta probabilidad de presencia de una oreja humana. El objetivo de la prueba es evaluar de forma general el desempeño del método con imágenes que contienen una oreja en ambiente controlado y no controlado, así como con imágenes que no contienen una oreja.

La Tabla 6.6 presenta la distribución del conjunto de 500 imágenes empleado para la evaluación del método propuesto, organizado en dos grupos: imágenes de oreja humana e imágenes de no oreja humana. Las imágenes de oreja humana incluyen 360 muestras obtenidas en condiciones controladas y no controladas, procedentes de bases públicas, capturas propias y recursos descargados de internet. Por su parte, las imágenes de no oreja humana incorporan 140 muestras provenientes de animales, plantas, hongos y otros elementos.

Tabla 6.6 Distribución del conjunto de imágenes utilizado para la validación del método propuesto

IMÁGENES DE OREJA HUMANA					
#	Imágenes	Origen	Muestras	Total	
1	Controlado	AMI Ear Database (Gonzalez et al., 2012)	45	180	360
2		IIT Delhi (<i>IIT Delhi Ear Database</i> , 2004)	45		
3		EarVN1.0 (Truong Hoang, 2019)	45		
4		Tomadas con el celular	45		
5	No controlado	In-the-wild Ear Database	150	180	
6		Descargadas de internet	30		
IMÁGENES DE NO OREJA HUMANA					
7	Animales	Descargadas de internet		30	140
8	Plantas y hongos			30	
9	Otros			80	

Algunos ejemplos de las imágenes utilizadas para la validación del método propuesto se muestran en la Figura 6.7. Las muestras se organizan en tres grupos: a) imágenes de oreja humana en ambiente controlado, b) imágenes de oreja humana en ambiente no controlado y c) imágenes correspondientes a la clase No oreja humana.

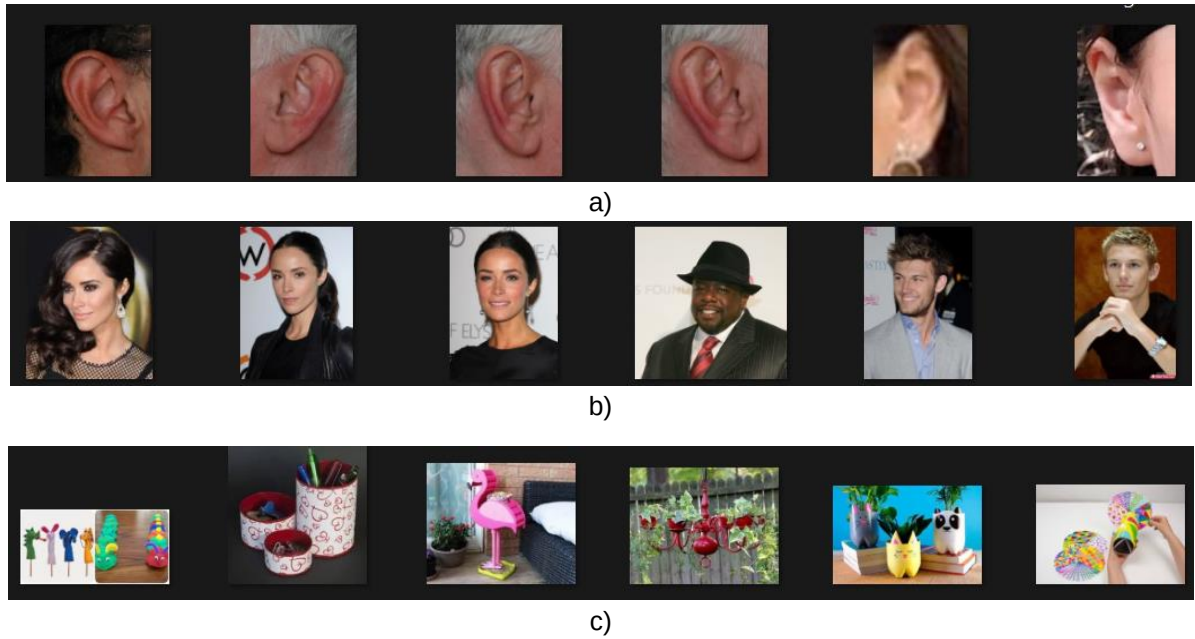


Figura 6.7 Muestras del conjunto de imágenes para la validación a) oreja humana ambiente controlado b) oreja humana ambiente no controlado y c) no oreja humana.

La Figura 6.8 se observa en el grupo a) imágenes donde la oreja aparece de forma clara, con menor variación de fondo, posición e iluminación, lo que facilita la localización del área de interés. En el grupo b) se muestran imágenes en condiciones no controladas, donde la oreja puede aparecer con variaciones de escala, orientación, iluminación, fondo, distancia o incluso parcialmente visible dentro de una imagen facial completa. Finalmente, el grupo c) incluye objetos que no corresponden a orejas humanas. Los resultados son mostrados en la Tabla 6.7.

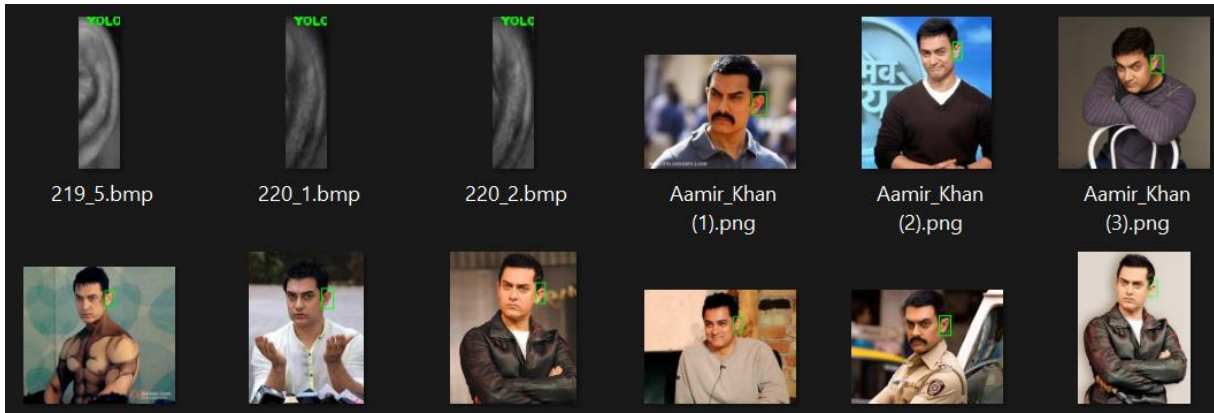
Tabla 6.7 Resultados de aplicar el método de localización propuesto

Método	Exactitud (%)	Precisión (%)	Recall (%)	F1-Score (%)
Propuesto (MediaPipe y YOLOv5s)	97.00	99.00	96.00	97.00

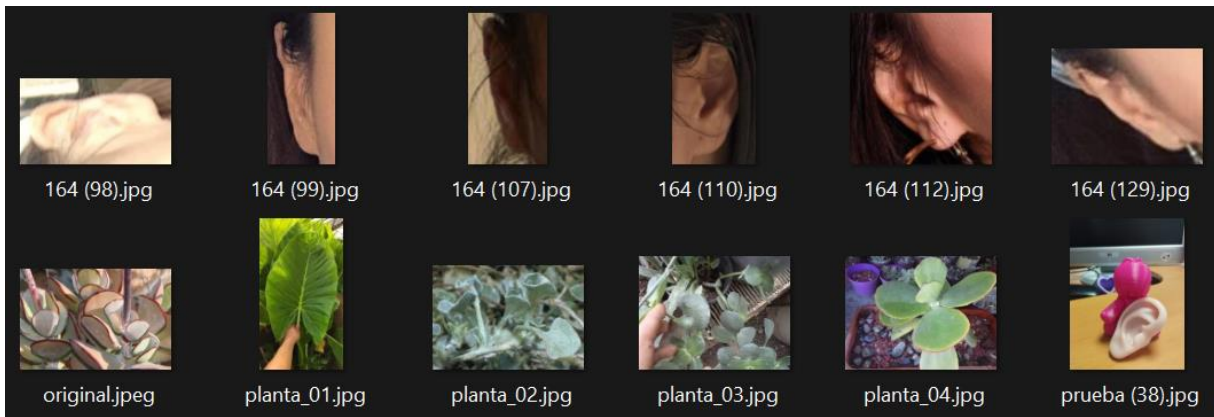
De acuerdo con los valores reportados en la Tabla 6.7, el método propuesto, basado en la integración de MediaPipe y YOLOv5s, alcanzó Exactitud de 97.00 %, lo que indica que se localiza correctamente la región de interés en la mayoría de las imágenes evaluadas. Asimismo, obtuvo Precisión de 99.00 %, mostrando que las regiones detectadas como oreja corresponden en gran medida a la clase correcta. El Recall de 96.00 % indica que el método logró identificar la mayor parte de las orejas presentes en el conjunto de evaluación.

En general, los resultados muestran una mejora en la localización del área de interés al combinar ambos métodos. Esto se debe a que MediaPipe presenta mejor desempeño en imágenes de ambiente no controlado, mientras que YOLOv5s en ambientes controlados.

Algunos ejemplos de los resultados obtenidos durante la evaluación del método propuesto se presentan en la Figura 6.8. La imagen ilustra muestras en las que el método logró localizar correctamente la región de interés, así como ejemplos clasificados dentro del grupo No oreja.



a)



b)

Figura 6.8 Muestras del resultado de la evaluación del método propuesto para la localización del área de interés a) objeto localizado correctamente b) imágenes establecidas como no oreja.

En la Figura 6.8-a se observan imágenes donde el método identifica y delimita la zona correspondiente a la oreja humana, tanto en imágenes de orejas aisladas como en rostros capturados en ambientes no controlados. En la Figura 6.8-b se presentan imágenes establecidas como No oreja, incluyendo objetos, plantas y muestras donde la oreja no aparece de forma clara o no corresponde a la región objetivo. Estos ejemplos permiten observar que el método propuesto puede localizar adecuadamente la oreja cuando está presente y, al mismo tiempo, distinguir imágenes que no contienen una oreja humana válida para el proceso de localización.

Los resultados obtenidos muestran que el método de localización propuesto puede identificar la región de interés de forma consistente en imágenes adquiridas en condiciones controladas y no controladas. La correcta localización de la oreja es un requisito previo para el desempeño global del sistema, ya que errores en dicha etapa pueden propagarse a las fases posteriores de caracterización y clasificación.

6.3. Evaluación experimental del Modelo 4C

En esta sección se presentan las pruebas experimentales realizadas para evaluar el desempeño del Modelo 4C propuesto, el cual integra información cromática RGB y un cuarto canal de textura basado en LBP. La evaluación se organiza de manera progresiva con el propósito de analizar, primero, el desempeño general del modelo; posteriormente, su comparación con arquitecturas de referencia; después, su comportamiento desde la perspectiva de seguridad biométrica mediante métricas PAD; y finalmente, su respuesta ante condiciones específicas que pueden afectar el reconocimiento de orejas naturales y artificiales.

La evaluación se estructuró en seis bloques principales: diseño experimental, evaluación de desempeño clásico, análisis de errores mediante matriz de confusión, evaluación mediante métricas PAD, análisis visual mediante mapas de características y mapas de activación, y evaluaciones parciales de robustez. Esta organización permite presentar los resultados de forma ordenada, relacionando las métricas cuantitativas con la interpretación visual del comportamiento del modelo.

Es importante señalar que la generación del cuarto canal LBP y la modificación de la arquitectura VGG19 para aceptar entradas de cuatro canales fueron descritas previamente en el Capítulo 4. Por tanto, en este apartado solo se retoman los elementos necesarios para interpretar los resultados experimentales.

6.3.1. Diseño experimental

Con el fin de garantizar la trazabilidad y reproducibilidad de las pruebas, se describen los conjuntos de imágenes utilizados, la distribución de las clases, la entrada evaluada y la configuración experimental aplicada durante el entrenamiento y la validación del modelo. La evaluación considera dos clases: Artificial y Natural, correspondientes a orejas artificiales y orejas humanas naturales, respectivamente.

Para mantener condiciones comparables entre modelos, los experimentos se realizaron con conjuntos balanceados de entrenamiento y validación. Asimismo, se mantuvieron constantes los principales parámetros de entrenamiento, con el objetivo de que las diferencias observadas en el desempeño pudieran atribuirse principalmente a la arquitectura evaluada y al tipo de entrada utilizada, ya sea RGB tradicional de tres canales o RGB con canal adicional de textura LBP.

6.3.1.1. Conjunto de imágenes de entrenamiento

Para el entrenamiento del modelo se utilizaron imágenes pertenecientes a dos clases: Artificial y Natural. El conjunto de entrenamiento estuvo conformado por 5,400 imágenes, distribuidas de manera balanceada en 2,700 imágenes por clase. Esta distribución permitió entrenar el modelo bajo una condición equilibrada, evitando que una clase tuviera mayor representación que la otra durante el aprendizaje.

La clase Artificial incluyó imágenes de orejas generadas o capturadas a partir de diferentes fuentes, tales como dibujos, muestras de filamento rígido, muestras de silicón y orejas realistas generadas mediante herramientas digitales.

Por su parte, la clase Natural integró imágenes provenientes de conjuntos públicos y capturas propias, incluyendo AMI Ear Database, EarVN1.0, IIT Delhi Ear Database, In-the-Wild Ear Database e imágenes obtenidas mediante cámara web. La distribución del conjunto de entrenamiento se presenta en la Tabla 6.8.

Tabla 6.8 Distribución de los conjuntos de imágenes para el entrenamiento

IMÁGENES CLASE ARTIFICIAL				
#	Imágenes	Origen	Muestras	Total
1	Dibujo	Descargadas del internet y Pinterest (Pinterest, 2024)	418	2,700
2	Filamento rígido	Tomadas con el celular Redmi Note 11 y celular OPPO	530	
3	Silicón	Find N3 Flip	876	
4	Realistas	Generadas mediante Meta Human (Epic Games, 2024)	876	
IMÁGENES CLASE NATURAL				
1	AMI Ear Database (Gonzalez et al., 2012)		536	2,700
2	EarVN.1(Truong Hoang, 2019)		508	
3	IIT Delhi Ear Database (IIT Delhi Ear Database, 2004)		655	
4	In-the-Wild Ear Database (Zhou & Zaferiou, 2017)		726	
5	Cámara web		275	

6.3.1.2. Conjunto de imágenes de validación

Para la validación se utilizó un conjunto diferente compuesto por 1,440 imágenes, distribuidas en 720 imágenes de oreja artificial y 720 imágenes de oreja natural. Al igual que en el conjunto de entrenamiento, la distribución fue balanceada para permitir una comparación directa entre ambas clases.

En la clase Artificial se incluyeron imágenes de dibujos, muestras de silicón e imágenes realistas generadas mediante herramienta de inteligencia artificial. En la clase Natural se utilizaron muestras públicas, específicamente *AMI Ear Database*, *EarVN1.0* e *In-the-Wild Ear Database*. La distribución del conjunto de validación se muestra en la Tabla 6.9.

Tabla 6.9 Distribución de los conjuntos de imágenes para la validación

IMÁGENES CLASE ARTIFICIAL				
#	Imágenes	Origen	Muestras	Total
1	Dibujo	Descargadas del internet	240	720
2	Silicón	Tomadas con el celular Redmi Note 11 y celular OPPO Find N3 Flip	240	
3	Realistas	Generadas mediante Chat GPT(OpenAI, 2024)	240	
IMÁGENES CLASE NATURAL				
1	AMI Ear Database (Gonzalez et al., 2012)		100	720
2	EarVN.1(Truong Hoang, 2019)		300	
3	In-the-Wild Ear Database (Zhou & Zaferiou, 2017)		320	

La conformación del conjunto de entrenamiento permitió incluir diferentes apariencias de orejas artificiales y naturales, con el propósito de proporcionar al modelo ejemplos representativos de ambas clases. En particular, la diversidad de materiales y fuentes en la clase artificial resulta relevante para la tarea PAD, ya que permite exponer al modelo a variaciones asociadas a posibles ataques de presentación

6.3.1.3. Entrada evaluada: comparación entre 3C y 4C

Para analizar la colaboración del canal de textura LBP, se compararon dos tipos de entrada. La primera corresponde al modelo de referencia de tres canales (3C), compuesto únicamente por la información RGB de la imagen. La segunda corresponde al Modelo 4C propuesto, en el cual se incorpora un cuarto canal de textura calculado mediante el operador LBP.

En el Modelo 3C, cada imagen se representa mediante los canales rojo, verde y azul. En el Modelo 4C, cada región de interés de la oreja se representa como una imagen compuesta por los canales R, G, B y LBP. De esta manera, la red convolucional recibe información cromática y de textura, lo que permite evaluar si el canal LBP aporta información discriminativa adicional para diferenciar orejas naturales de orejas artificiales.

6.3.1.4. Hiperparámetros y configuración experimental

La configuración experimental utilizada durante el entrenamiento y validación se resume en la Tabla 6.10. Para todos los experimentos se utilizó un tamaño de lote de 32, un máximo de 50 épocas, parada temprana con paciencia de 5, tasa de aprendizaje de $1e-05$, optimizador SGD con momentum de 0.9 y función de pérdida CrossEntropyLoss. El entrenamiento se ejecutó con una semilla de aleatoriedad de 42 y utilizando dispositivo cuda, con el fin de aprovechar la aceleración por GPU. El número total de muestras de entrenamiento fue de 5,400 y el de validación fue de 1,440.

Tabla 6.10 Hiperparámetros y configuración adicional utilizados en la experimentación

HIPERPARÁMETROS	
1	Lote (Batch Size): 32
2	Épocas: 50
3	Paciencia (Parada Temprana): 5
4	Tasa de aprendizaje: $1e-05$
5	Optimizador: SGD con momentum=0.9
6	Criterio de pérdida: CrossEntropyLoss
7	Modelo: 3C/ 4C modificado para entrada de 4 canales
8	Transformaciones Modelo 3C: Compose (Resize (size= (224, 224), interpolation=bilinear, max_size=None, antialias=True) ToTensor () Normalize (mean= [0.485, 0.456, 0.406], std= [0.229, 0.224, 0.225]))
9	Transformaciones de Modelo 4C: Compose (Resize (size= (224, 224), interpolation=bilinear, max_size=None, antialias=True) ToTensor () Normalize (mean=[0.5341, 0.3783, 0.3592, 0.5819], std=[0.2021, 0.1622, 0.1559, 0.2029]))
10	img_lbp=local_binary_pattern (img_gray, P=8, R=1, method= default)
11	lbp_norm = cv2.normalize(img_lbp, None, 0, 255, cv2.NORM_MINMAX).astype('uint8')
CONFIGURACIÓN ADICIONAL	
1	Semilla: 42
3	Número de muestras de entrenamiento: 5400
4	Número de muestras de validación: 1440
5	Número de clases: 2 ['Artificial', 'Natural']
6	Dispositivo: cuda

6.3.2. Evaluación de desempeño con entrada 3C y 4C

La primera evaluación cuantitativa se realizó mediante métricas clásicas de clasificación: Exactitud, Precisión, Recall y F1-score. Estas métricas permiten analizar el desempeño general de los modelos y su comportamiento por clase.

Se evaluaron tres arquitecturas de aprendizaje profundo basadas en redes neuronales convolucionales: AlexNet, ResNet50 y VGG19. Cada arquitectura fue probada con entrada 3C y 4C, con el propósito de determinar si la incorporación del canal de textura LBP mejora el desempeño respecto al uso exclusivo de información RGB. La Tabla 6.11 presenta los resultados obtenidos para cada arquitectura y tipo de entrada.

Tabla 6.11 Comparativa de los resultados de desempeño entre el Modelo 3C y 4C

Modelo	Entrada	Exactitud (%)	Clase Artificial (%)			Clase Natural (%)		
			Precisión	Recall	F1-score	Precisión	Recall	F1-score
AlexNet	3C	95.00	94.00	97.00	95.00	97.00	94.00	95.00
	4C	95.00	93.00	97.00	95.00	97.00	93.00	95.00
ResNet50	3C	94.00	97.00	90.00	94.00	91.00	97.00	94.00
	4C	98.00	98.00	98.00	98.00	98.00	98.00	98.00
VGG19	3C	96.39	97.00	96.00	96.00	96.00	97.00	96.00
	4C	98.33	99.00	98.00	98.00	98.00	99.00	98.00

Los resultados muestran que la incorporación del cuarto canal produce mejoras relevantes en dos de las tres arquitecturas evaluadas. En el caso de ResNet50, la exactitud aumenta de 94.00 % con entrada 3C a 98.00 % con entrada 4C. En VGG19, la exactitud incrementa de 96.39 % a 98.33 %. En contraste, AlexNet mantiene una exactitud de 95.00 % tanto en la versión 3C como en la versión 4C.

Desde el punto de vista de las métricas por clase, la arquitectura VGG19 con entrada 4C muestra un desempeño balanceado, con Precisión de 99.00 %, Recall de 98.00 % y F1-score de 98.00 % para la clase Artificial; y Precisión de 98.00 %, Recall de 99.00 % y F1-score de 98.00 % para la clase Natural. Estos valores indican que el modelo mantiene un desempeño alto tanto en la detección de orejas humana artificiales como naturales.

En conjunto, los resultados sugieren que el canal LBP proporciona información complementaria a la representación RGB, particularmente en arquitecturas con mayor capacidad de extracción de características, como ResNet50 y VGG19. Esta mejora es relevante para la tarea PAD, ya que la discriminación entre orejas humanas naturales y artificiales puede depender de variaciones locales de textura que no siempre son evidentes en los canales RGB.

6.3.2.1. Integración final del Modelo 4C

A partir de los resultados de la Tabla 6.11, se seleccionó la arquitectura VGG19 para integrar en el Modelo 4C para los análisis posteriores. Debido a que dicha configuración obtuvo la mayor exactitud global, con 98.33 %, y presentó métricas equilibradas para ambas clases. La selección de VGG19 con entrada 4C permite continuar la evaluación con el modelo que mostró mejor desempeño general y mayor estabilidad en la clasificación de orejas naturales y artificiales.

Por ello, las secciones posteriores se centran en el análisis de errores, métricas PAD, mapas de activación y pruebas parciales correspondientes a esta configuración.

6.3.3. Análisis de errores mediante matriz de confusión

Para complementar las métricas clásicas de desempeño, se analiza la matriz de confusión del Modelo 4C. La matriz permite observar la distribución de aciertos y errores entre las clases Artificial y Natural, y sirve como base para el cálculo de las métricas PAD.

En el conjunto de validación se utilizaron 720 imágenes de la clase Artificial y 720 imágenes de la clase Natural. A partir de los valores reportados para APCER y BPCER, se identificaron 17 errores en muestras artificiales y 7 errores en muestras naturales. Bajo esta convención, las muestras artificiales corresponden a presentaciones de ataque, mientras que las muestras naturales corresponden a presentaciones bona fide. La Figura 6.9 muestra la matriz de confusión del modelo propuesto.

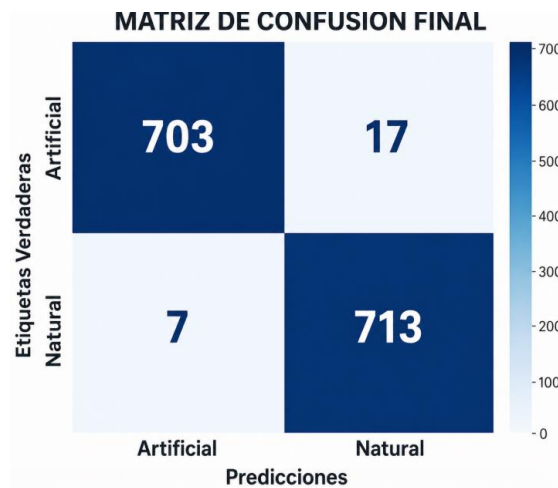


Figura 6.9 Matriz de confusión del modelo propuesto para la clasificación de orejas naturales y artificiales.

A partir de la matriz se observa que el modelo clasificó correctamente 703 de 720 imágenes artificiales y 713 de 720 imágenes naturales. Los principales errores se concentran en la clase Artificial, donde 17 imágenes artificiales fueron clasificadas como naturales. Este tipo de error es especialmente relevante en un escenario PAD, ya que representa casos en los que una presentación artificial podría ser aceptada como natural.

6.3.4. Evaluación del Modelo 4C mediante métricas PAD actuales

Además de las métricas clásicas de clasificación, el Modelo 4C fue evaluado mediante métricas PAD, debido a que el objetivo del trabajo de tesis es la caracterización de la oreja humana para diferenciar entre artificiales y naturales. En este contexto, las muestras artificiales se consideran presentaciones de ataque (PA), mientras que las muestras naturales corresponden a presentaciones bona fide. Las métricas PAD permiten analizar el desempeño desde una perspectiva de seguridad biométrica. En particular, se consideran APCER, BPCER y ACER descritas en el Capítulo 2 (Marco Teórico) apartado 2.6.3 (Métricas PAD).

La Tabla 6.12 presenta la comparación entre el modelo de referencia con entrada RGB de tres canales y el modelo propuesto con entrada RGB + LBP.

Tabla 6.12 Comparación del desempeño entre el modelo de referencia (RGB) y el modelo propuesto (4C) según métricas PAD actuales

Modelo	Entrada	Tasas de error (%)		
		APCER	BPCER	ACER
Línea base	3C (RGB)	4.03	3.19	3.61
Propuesto	4C (RGB +LBP)	2.36	0.97	1.67

Los resultados muestran que el modelo propuesto reduce las tres tasas de error PAD respecto al modelo de referencia. El modelo RGB de tres canales reporta un ACER de 3.61 %, con APCER de 4.03 % y BPCER de 3.19 %. En contraste, el Modelo 4C alcanza un ACER de 1.67 %, con APCER de 2.36 % y BPCER de 0.97 %.

El valor de APCER del Modelo 4C se obtuvo a partir de los errores cometidos sobre la clase Artificial:

$$APCER = \frac{17}{720} \times 100 = 2.36 \%$$

Este resultado indica que 17 de las 720 presentaciones artificiales fueron clasificadas incorrectamente como naturales. Por su parte, el valor de BPCER se calculó a partir de los errores cometidos sobre la clase Natural:

$$BPCER = \frac{7}{720} \times 100 = 0.97 \%$$

Este resultado indica que 7 de las 720 presentaciones naturales fueron clasificadas incorrectamente como artificiales. Finalmente, el ACER se calculó como el promedio entre APCER y BPCER:

$$ACER = \frac{2.36 + 0.97}{2} = 1.67 \%$$

Los resultados muestran que la incorporación del canal LBP reduce tanto los errores asociados a seguridad como los errores asociados a usabilidad. La disminución de APCER es particularmente relevante, ya que representa una reducción en la proporción de orejas artificiales clasificadas como naturales.

En el escenario binario evaluado, y considerando que no se reportan casos donde no se localiza la oreja, los valores de APCER y BPCER pueden relacionarse con las métricas complementarias FGR y FFR descritas en el Capítulo 2 (Marco teórico) apartado 2.6.3 (Métricas PAD). Bajo esta condición experimental, APCER equivale a FGR y BPCER equivale a FFR. Por tanto:

$$FGR = 2.36 \%, \quad FFR = 0.97 \%, \quad \text{y} \quad HTER = \frac{FGR + FFR}{2} = \frac{2.36 + 0.97}{2} = 1.67 \%$$

La interpretación de estas equivalencias debe limitarse al escenario experimental descrito, debido a que se trata de una clasificación binaria entre orejas artificiales y naturales.

6.3.5. Análisis visual del aprendizaje del modelo

Con el propósito de complementar los resultados cuantitativos, se realizó un análisis visual del comportamiento del modelo mediante mapas de características y mapas de activación. Este análisis permite observar qué regiones y patrones de la imagen son resaltados por el modelo durante el proceso de clasificación.

El análisis visual no sustituye las métricas cuantitativas, pero permite interpretar de manera cualitativa el efecto del canal LBP en la representación aprendida por la red. En particular, se busca observar si el Modelo 4C atiende con mayor claridad a estructuras y texturas propias de la oreja, en comparación con el modelo de referencia basado únicamente en RGB.

6.3.5.1. Mapas de características

Los mapas de características corresponden a las salidas generadas por los filtros de la primera capa convolucional del modelo. Dichas representaciones permiten analizar qué patrones locales de la imagen son resaltados durante las etapas iniciales del procesamiento, tales como bordes, contornos, cambios de intensidad y texturas elementales. La Figura 6.10 muestra una comparación entre los mapas de características generados por el Modelo 3C y el Modelo 4C.

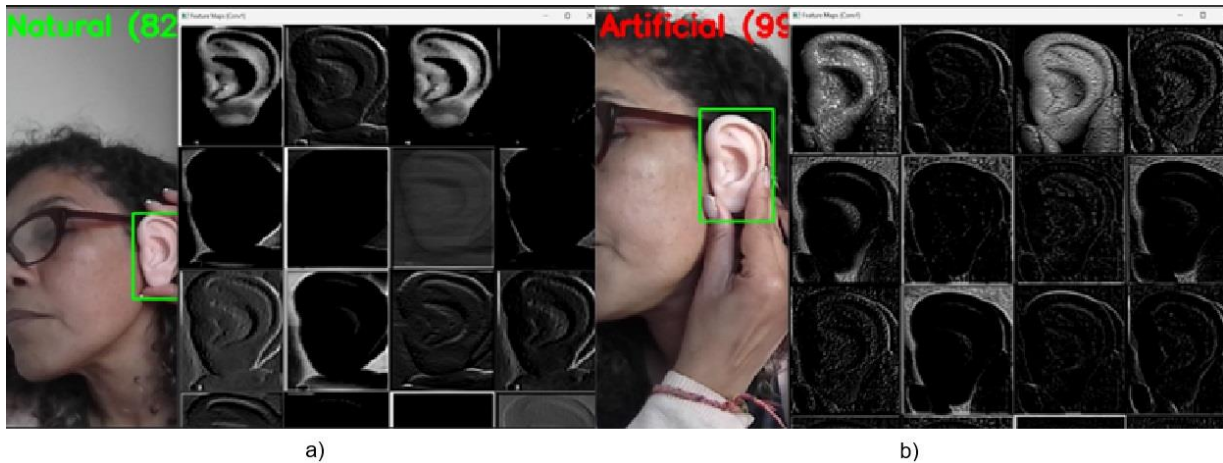


Figura 6.10 Visualización de mapas de características de la primera capa convolucional. (a) Modelo 3C y (b) Modelo 4C.

En la comparación visual se observa que el Modelo 4C genera mapas con mayor presencia de información estructural asociada a la oreja. La silueta y regiones internas aparecen con mayor definición en varios canales, lo que sugiere que la red aprovecha la información de textura para reforzar la representación inicial de la imagen. Por el contrario, en algunos mapas del modelo 3C se observan canales con menor activación o con información menos localizada.

La diferencia visual es consistente con los resultados cuantitativos obtenidos. Como se mostró en la Tabla 6.12, el Modelo 4C alcanzó Exactitud de 98.33 %, superando al modelo 3C, que obtuvo 96.39 %. Asimismo, el modelo propuesto presentó mejores tasas de error PAD, reduciendo el APCER de 4.03 % a 2.36 % y el ACER de 3.61 % a 1.67 %.

En conjunto, estos resultados indican que la incorporación del canal LBP ayuda a mejorar la representación inicial de patrones locales de textura.

6.3.5.2. Mapas de activación mediante Grad-CAM

Además de los mapas de características, se analizaron mapas de activación mediante Grad-CAM (Selvaraju et al., 2020), con el objetivo de identificar las regiones de la imagen que contribuyen con mayor peso a la decisión del modelo. A diferencia de los mapas de características de capas iniciales, Grad-CAM permite interpretar la atención del modelo en etapas más cercanas a la clasificación final (Selvaraju et al., 2020). La Figura 6.11 muestra una comparación cualitativa entre los mapas de activación del Modelo 3C y del Modelo 4C.

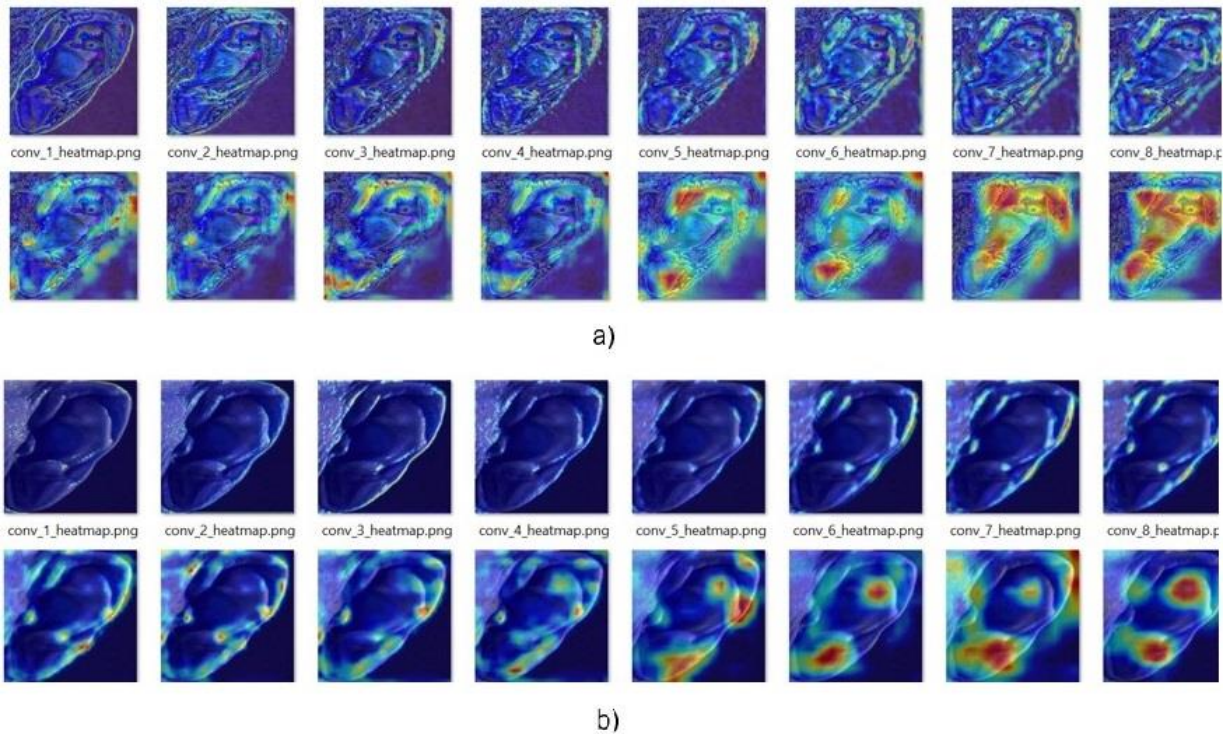


Figura 6.11 Visualización de mapas de activación mediante Grad-CAM a) Modelo 3C y b) Modelo 4C.

En la imagen se observa que en el Modelo 3C, la activación tiende a distribuirse de forma dispersa, en contraste, el Modelo 4C muestra activaciones concentradas en regiones estructurales de la oreja, como el hélix, antihélix, concha y contornos internos.

La comparación entre ambos mapas sugiere que el Modelo 4C no solo mejora el desempeño cuantitativo, sino que también modifica la forma en que la red atiende la imagen. En lugar de depender únicamente de información RGB.

6.3.6. Evaluaciones parciales del Modelo 4C

Con el fin de analizar el desempeño del Modelo 4C ante condiciones que pueden afectar la correcta clasificación de oreja humana, se diseñaron cinco conjuntos de prueba independientes. Cada conjunto fue organizado en dos clases: Natural y Artificial, manteniendo una distribución balanceada para permitir una comparación objetiva entre condiciones.

La Tabla 6.13 muestra los conjuntos utilizados en las pruebas parciales del Modelo 4C, considerando cinco condiciones específicas: uso de accesorios, variación morfológica,

cambios de iluminación, oclusión parcial y variación de escala. Estas condiciones permiten analizar el comportamiento del modelo ante escenarios que pueden presentarse en aplicaciones reales y que podrían afectar la visibilidad, textura, forma o tamaño de la oreja.

Tabla 6.13 Distribución de los conjuntos evaluados en las pruebas parciales del Modelo 4C

#	Conjunto	Condición	Clase	No. de muestras	Descripción
1	Accesorios	Uso de accesorios	Natural / Artificial	120	Evalúa el modelo para reconocer orejas cuando se añaden pendientes, audífonos o piercings que alteran la silueta.
2	Diferentes	Variación morfológica	Natural / Artificial	100	Contiene orejas de sujetos/estilos no vistos durante el entrenamiento (p. ej. curvas atípicas) para medir la generalización a ejemplos “fuera de distribución”.
3	Iluminación	Cambios de iluminación	Natural / Artificial	120	Incluye imágenes con exposiciones altas/bajas, dominantes de color y contrastes extremos para probar la invariancia luminosa.
4	Oclusión	Oclusión parcial	Natural / Artificial	120	Introduce cubrimientos parciales (cabello, marcos de lentes) o recortes que esconden partes de la oreja.
5	Tamaño	Variación de escala	Natural / Artificial	120	Presenta orejas en diferentes escalas y resoluciones (pequeñas o ampliadas) para validar la sensibilidad a cambios de tamaño en el detector/recortador previo.

En total, se evaluaron conjuntos con entre 100 y 120 muestras, distribuidas entre las clases Natural y Artificial. El conjunto Diferentes permite medir la generalización del modelo ante orejas no vistas durante el entrenamiento; mientras que los conjuntos de Accesorios, Iluminación, Oclusión y Tamaño evalúa la estabilidad del modelo frente a condiciones visuales específicas. De esta manera, las pruebas parciales permiten identificar en qué condiciones el Modelo 4C mantiene su desempeño y en cuáles no.

En la Tabla 6.14 se presentan los resultados obtenidos al evaluar el modelo en los cinco conjuntos de prueba parciales. Cada conjunto fue evaluado de manera independiente con las clases Artificial y Natural, utilizando las métricas de clasificación establecidas en el Capítulo 2. (Marco Teórico) correspondientes a Exactitud, Precisión, *Recall* y *F1-score*.

En la columna de comentarios se reportan observaciones derivadas de los errores de clasificación, donde FP corresponde a falsos positivos y FN a falsos negativos.

Tabla 6.14 Resultado de las métricas aplicadas a las pruebas parciales

Conjunto	Exactitud (%)	Clase Artificial (%)			Clase Natural (%)			Comentario
		Precisión	Recall	F1-score	Precisión	Recall	F1-score	
Accesorios	95.00	97.00	93.00	95.00	94.00	97.00	95.00	Genera cuatro FN y dos FP.
Diferentes	95.00	100	90.00	95.00	91.00	100.00	95.00	Precisión alta, y pierde 10 % de recall en artificiales.
Iluminación	97.00	98.00	97.00	97.00	97.00	98.00	98.00	Tres errores totales; evidencia a invariancia a cambios de luz/contraste.
Oclusión	94.00	100.00	88.00	94.00	90.00	100.00	94.00	12 % de las orejas artificiales, parcialmente ocluidas se clasifican como naturales (7 FN).
Tamaño	97.00	100.00	93.00	97.00	94.00	100.00	97.00	Errores en imágenes artificiales muy pequeños.

Los resultados muestran que el modelo obtuvo valores de Exactitud entre 94.00 % y 97.00 %. Los mejores resultados se observaron en los conjuntos de Iluminación y Tamaño, ambos con Exactitud de 97.00 %. En el conjunto de Iluminación, el modelo alcanzó valores equilibrados para ambas clases, con Precisión, Recall y F1-score entre 97.00 % y 98.00 %, lo que sugiere un comportamiento estable ante las variaciones de luz y contraste consideradas en esta prueba. En el conjunto de Tamaño, la Exactitud también fue de 97.00 %; sin embargo, el Recall de la clase Artificial fue de 93.00 %, lo que indica que algunos errores se concentraron en imágenes artificiales con variaciones de escala.

En los conjuntos Accesorios y Diferentes, el modelo obtuvo Exactitud 95.00 %. Para Accesorios, se reportaron cuatro falsos negativos y dos falsos positivos, lo cual indica que ciertos elementos como pendientes, audífonos o piercings pueden modificar la apariencia visual de la oreja y afectar la clasificación. En el conjunto Diferentes, la Precisión para la clase Artificial fue de 100 %, pero el Recall disminuyó a 90.00 %, lo que significa que el modelo no clasificó correctamente el 10 % de las muestras artificiales de ese conjunto. Este resultado indica una reducción en la recuperación de la clase Artificial cuando se presentan variaciones morfológicas no observadas durante el entrenamiento.

El conjunto con menor desempeño fue Oclusión, con Exactitud de 94.00 %. En este caso, el Recall de la clase Artificial fue de 88.00 %, lo que representa el valor más bajo entre las pruebas parciales. Esto indica que algunas orejas artificiales parcialmente ocluidas fueron clasificadas como naturales, generando siete falsos negativos. Este resultado es relevante desde la perspectiva de detección de ataques de presentación, ya que los falsos negativos en la clase Artificial implican que una muestra artificial no fue detectada correctamente.

6.4. Comparativa mediante métricas PAD históricas

La Tabla 6.15 presenta una comparación del Modelo 4C con trabajos del estado del arte que reportan métricas PAD históricas para la detección de orejas reales y falsas. En particular, se consideran las tasas FFR (*False Fake Rate*), FGR (*False Genuine Rate*) y HTER (*Half Total Error Rate*), las cuales permiten analizar la tasa de error asociada a la clasificación de muestras naturales y artificiales. en escenarios de detección de ataques de presentación.

Tabla 6.15 Comparativa del modelo propuesto con el estado del arte mediante métricas PAD históricas

Referencia	Técnicas	Imágenes	FFR (%)	FGR (%)	HTER (%)
(Toprak & Toygar, 2020)	25 IQM	AMI	2.00	0.00	1.00
		UBEAR	20.00	0.00	10.00
(Toprak & Toygar, 2021)	PSNR, GPE, SSIM, RRED, HLF1, ReLU	AMI	0.00	0.00	0.00
		UBEAR	0.00	0.00	0.00
		IITD	0.00	0.00	0.00
		USTB	0.00	0.00	0.00
Modelo 4C	LBP y VGG19	CIOHNA	0.97	2.36	1.67

Los resultados muestran que Toprak y Toygar (2020) reportan un HTER de 1.00 % en AMI y de 10.00 % en UBEAR, mientras que Toprak y Toygar (2021) reportan valores de 0.00 % para FFR, FGR y HTER en las bases AMI, UBEAR, IITD y USTB. Por su parte, el Modelo 4C, evaluado sobre el conjunto CIOHNA, obtuvo FFR de 0.97 %, FGR de 2.36 % y HTER de 1.67 %.

Si bien, los trabajos previos reportan valores de error menores, la comparación debe interpretarse considerando la naturaleza de los conjuntos utilizados. En este trabajo, CIOHNA se conforma como un conjunto compuesto de imágenes de orejas humanas naturales y artificiales, integrando muestras provenientes de escenarios controlados y no controlados. Las muestras naturales incluyen conjuntos públicos como AMI e IIT Delhi, así como conjuntos con mayor variabilidad visual, como In-the-Wild Ear Database y EarVN1.0, además de capturas propias obtenidas con cámaras web y teléfonos móviles. Esto permite evaluar el modelo en condiciones cercanas a escenarios reales, donde pueden presentarse variaciones de iluminación, rotación, escala, ruido ambiental y oclusión.

Asimismo, las muestras artificiales de CIOHNA no se limitan a un solo tipo de ataque, sino que incluyen dibujos, impresiones 3D elaboradas con materiales como silicón y filamento rígido, así como imágenes realistas generadas mediante herramientas de inteligencia artificial. Por lo tanto, el HTER de 1.67 % obtenido por el Modelo 4C debe analizarse no solo como una tasa de error, sino también en función de la variedad del conjunto de evaluación. A diferencia de evaluaciones realizadas sobre conjunto de imágenes en condiciones controladas, CIOHNA incorpora una mayor variabilidad en el origen, tipo de muestra y condiciones de captura. En este sentido, los resultados sugieren que el Modelo 4C mantiene un desempeño competitivo en un escenario de evaluación más diverso, lo cual fortalece su aplicabilidad en sistemas PAD de biometría de la oreja.

6.5. Discusión de los resultados

Los resultados experimentales obtenidos permiten analizar la propuesta en el mismo orden en que fue evaluada: primero la etapa de localización del ROI, posteriormente el desempeño del Modelo 4C y, finalmente, su comparación frente a métricas y trabajos relacionados con PAD en biometría de la oreja.

En la evaluación de localización, MediaPipe presentó mejor comportamiento en imágenes no controladas, mientras que YOLOv5s fue más adecuado en imágenes controladas donde MediaPipe no logró ubicar puntos faciales. A partir de esta evaluación, el método propuesto integró ambos enfoques y alcanzó Exactitud de 97.00 %, Precisión de 99.00 %, Recall de 96.00 % y F1-score de 97.00 %, lo que muestra que la combinación MediaPipe + YOLOv5s permite localizar la región de interés de manera más consistente antes de la clasificación PAD.

Posteriormente, la evaluación del Modelo 4C mostró que la incorporación del canal de textura en una entrada RGB + LBP mejora el desempeño frente a entradas RGB. En particular, para VGG19 se observó un incremento de Exactitud de 96.39 % a 98.33 % al pasar de 3C a 4C. Asimismo, la comparación con AlexNet y ResNet50 permitió identificar que VGG19 obtuvo el desempeño más equilibrado entre las clases Artificial y Natural, por lo que fue seleccionada como arquitectura final del Modelo 4C.

Desde la perspectiva de seguridad biométrica, en las métricas PAD el Modelo 4C redujo el APCER de 4.03 % a 2.36 %, el BPCER de 3.19 % a 0.97 % y el ACER de 3.61 % a 1.67 % respecto a la línea base 3C. Estos resultados indican que el uso del canal LBP no solo mejora la exactitud general, sino que también reduce los errores asociados a la aceptación de orejas artificiales como naturales y al rechazo de orejas naturales como artificiales.

Las pruebas parciales permitieron identificar condiciones específicas que afectan el desempeño del modelo. Aunque el Modelo 4C mantuvo exactitudes entre 94.00 % y 97.00 % en los conjuntos evaluados, la oclusión fue la condición crítica, debido a que redujo el *Recall* de la clase Artificial a 88.00 %. Este resultado es relevante en un contexto PAD, ya que una oreja artificial parcialmente ocluida puede ocultar rasgos texturales importantes y aumentar el riesgo de clasificación incorrecta.

Finalmente, la comparación con trabajos del estado del arte mediante métricas PAD históricas muestra que el Modelo 4C obtiene un HTER de 1.67 % en el conjunto CIOHNA. Aunque algunos trabajos previos reportan tasas de error menores, la comparación debe interpretarse con cautela, ya que los protocolos, bases de datos, tipos de ataque y condiciones de captura no son equivalentes. En este trabajo, la evaluación integra orejas humanas naturales y artificiales procedentes de fuentes diversas, incluyendo imágenes controladas, no controladas, generadas digitalmente y capturadas con distintos dispositivos. Por ello, los resultados sugieren que la integración de localización del ROI y clasificación PAD mediante el Modelo 4C constituye una alternativa viable para apoyar a sistemas biométricos basados en oreja frente a presentaciones artificiales.

CAPÍTULO 7

CONCLUSIONES

Las conclusiones que se presentan a continuación se sustentan en los experimentos y métricas reportadas en el Capítulo 6 (Pruebas y Resultados) y deben considerarse dentro del alcance de los conjuntos de imágenes descritos en el Capítulo 5.

7.1. Conclusiones generales

La presente tesis se centra en la caracterización de la oreja humana para clasificarla como natural o artificial, aborda la detección de ataques de presentación (PAD) basados en la imagen del biométrico, motivado por la disponibilidad creciente de medios de suplantación (p. ej., impresiones, pantallas, prótesis y contenido sintético). A partir del análisis del estado del arte, se identificó la limitada información sobre la caracterización de una oreja humana natural y una imagen generada, capaz de operar en condiciones realistas (*in-the-wild*) y de discriminar entre muestras genuinas y artefactos artificiales. Con el objetivo de atender la problemática, se propuso un modelo que incorpora información de textura local y un flujo completo de procesamiento integrando la localización de la oreja, la caracterización de la información y un modelo de clasificación basado en aprendizaje profundo, conformando un sistema alineado a condiciones controladas y no controladas.

En los experimentos realizados, el enfoque híbrido de localización (MediaPipe + YOLOv5s) mostró un desempeño del 97 % de exactitud para extraer la ROI de oreja en imágenes en entornos controlados y no controlados, incluyendo imágenes provenientes de y capturas móviles/internet, lo cual facilita que la etapa de PAD opere sobre una región consistente.

Los resultados obtenidos en la experimentación permiten concluir, que el enfoque híbrido de localización basado en MediaPipe + YOLOv5s constituye una solución competitiva, al lograr 97 % de exactitud en la localización del ROI en entornos controlados y no controlados, incluyendo imágenes provenientes de cuatro conjuntos públicos (AMI, EarVN1.0, IITD e ITWE) y capturas móviles/internet. Asimismo, la integración de un canal de textura basado en LBP como cuarto canal de entrada (representación 4C) aportó información de utilidad para resaltar texturas.

En particular, el modelo 4C alcanzó una exactitud global del 98 % en el escenario de PAD evaluado, mostrando mejoras frente a la configuración 3C (RGB) bajo las condiciones reportadas. Además, al evaluar el comportamiento desde la perspectiva de seguridad, la incorporación del canal LBP permitió reducir el *APCER* del modelo base (3C) de 4.03 % a 2.36 %, mostrando una disminución en el riesgo de aceptar ataques como oreja natural (*bona fide*).

En conjunto, los resultados sugieren que una solución integral —localización de ROI + PAD mediante fusión temprana RGB+LBP (Modelo 4C) — puede contribuir a fortalecer la seguridad de sistemas biométricos basados en oreja, especialmente en contextos modernos de autenticación remota y escenarios no controlados.

7.2. Objetivos alcanzados

El desarrollo de la investigación permitió cumplir con el objetivo general y los objetivos específicos planteados en la tesis. La propuesta del Modelo 4C establecida en el Capítulo 4 (Propuesta de Solución), integra una técnica de extracción de textura (LBP) con una arquitectura de aprendizaje profundo (VGG19), definidas en el Capítulo 2 (Marco Teórico) apartado 2.4 y 2.5 respectivamente, obteniendo resultados competitivos con respecto a la literatura existente. En la Tabla 7.1 se muestra el cumplimiento del objetivo general y los específicos.

Tabla 7.1 Tabla de evidencia de los logros sobre objetivos

Objetivo General	Evidencia
Proponer un modelo computacional para la diferenciación automática entre orejas humanas naturales y artificiales (<i>anti-spoofing</i>) en imágenes adquiridas en entornos no controlados, mediante una arquitectura híbrida que fusione características de textura local con representaciones profundas.	Propuesta del Modelo 4C para reconocer orejas humanas naturales o artificiales, descrita en el Capítulo 4 (Propuesta de Solución).
Objetivos Específicos	
Analizar las limitaciones de las técnicas actuales referentes a la caracterización e identificación de la oreja, para seleccionar las adecuadas al problema según su desempeño.	Se realizó un análisis de las técnicas existentes de biometría de la oreja y detección de ataques de presentación. Mencionadas en el Capítulo 3 (Estado del Arte)
Implementar redes profundas con los mejores resultados documentados, para validar su desempeño ante escenarios complejos.	Se implementaron tres redes profundas y se seleccionó VGG19 por mostrar mejores resultados. Los resultados se muestran en el Capítulo 6 (Pruebas y Resultados), apartado 6.3 (Evaluación del Modelo 4C).
Construir un banco de imágenes de orejas humanas naturales y artificiales adquiridas en ambientes no controlados para seleccionar las imágenes de orejas a procesar.	Se construyó un banco de imágenes con orejas naturales y artificiales denominado CIOHNA (Conjunto de Imágenes de Orejas Humanas Naturales y Artificiales), descrito en el Capítulo 5 (Conjunto de Imágenes).
Formular el modelo de caracterización mediante la combinación de una técnica tradicional de descripción de textura y una red neuronal de aprendizaje profundo.	Integración de LBP con la red neuronal VGG19, generando el Modelo 4C, descrito en el Capítulo 4 (Propuesta de Solución).
Evaluar el modelo propuesto con las métricas estándar empleadas en el reconocimiento y en la detección de suplantación de orejas.	Se evaluó el desempeño del modelo propuesto mediante un protocolo experimental y métricas estándar de PAD, establecidas en el Capítulo 2 (Marco Teórico) apartado 2.6 (Métricas de evaluación). Los resultados de la evaluación del modelo se muestran en el Capítulo 6 (Pruebas y Resultados), apartado 6.3 (Evaluación experimental del Modelo 4C).

7.3. Productos académicos obtenidos

Los productos académicos principales se sintetizan en:

- **Creación de un conjunto de imágenes diversificado:** Se generó un Conjunto de Imágenes de Orejas Humanas Naturales y Artificiales (CIOHNA), que incluye:
 - ✓ Imágenes de cuatro conjuntos públicos (AMI, EarVN.0, IITD e ITWE)
 - ✓ Representaciones de orejas artificiales creadas mediante inteligencia artificial (IA).
 - ✓ Fotografías de orejas tomadas con dispositivos móviles.
 - ✓ Imágenes descargadas de internet.
 - ✓ Orejas humanas reales capturadas directamente desde cámaras web, abarcando tanto ambientes controlados como entornos reales.
- **Método para localización de ROI:** Diseño e implementación de un método para la localización de la oreja en imágenes adquiridas en entornos controlados y no controlados, mediante la combinación de MediaPipe y YOLOv5s.
- **Modelo 4C (RGB+LBP):** Se introduce el Modelo 4C, una solución híbrida que incorpora un canal adicional de textura basado en LBP junto con información RGB, adaptado a VGG19, orientado a capturar texturas y reducir el riesgo de aceptación de ataques que imitan apariencia cromática.
- **Integración de un flujo completo (ROI + PAD):** Se propone un sistema aplicable en escenarios reales para la detección de ataques de presentación que integra explícitamente la localización de la oreja, evitando el supuesto frecuente de “orejas ya recortadas/alineadas” de trabajos previos.

7.4. Actividades académicas realizadas

A continuación, se presentan los productos académicos realizados, resultado del proyecto de tesis.

- ✓ Elaboración y envío del artículo: “A Texture-Based Deep Learning Approach for Classifying Artificial and Natural Ears Under Uncontrolled Environment”. Se envió a la revista: *Expert Systems with Applications*, indexada en *Journal Citation Reports* (JCR) con un factor de impacto de 7.5. Se encuentra en el primer cuartil (Q1).

A Texture-Based Deep Learning Approach for
Classifying Artificial and Natural Ears Under
Uncontrolled Environment

Elydi López Hernández*, Andrea Magalán Salazar*, Ondina Osorio Vergara
Villalga^{a,b}

^aCentro Nacional de Investigación y Desarrollo Tecnológico,
Departamento de Computación,
Interior Politécnico Pabellón 5/10, Polanco, P.C. 04501,
Ciudad de México, México

^bUniversidad Autónoma de Ciudad Juárez,
Instituto de Ingeniería y Tecnología,
Avenida del Charro 450 Norte, Partido Romero, P.C. 32310,
Juárez, Chihuahua, México

*Corresponding author
Email addresses: elydi@ciqa.mx (Elydi López Hernández),
andrea.magalán@ciqa.mx (Andrea Magalán Salazar), osorio@ciqa.mx (Ondina Osorio Vergara Villalga)



Móstoles, Comunidad de Madrid, España, 20/02/2025
Asunto: aceptación de estancia doctoral

Dra. Andrea Magadán Salazar
Profesora-Investigadora
CENIDET
PRESENTE

Apreciable Doctora, es para mí un placer y un gusto informarle que ha sido aprobada la estancia académica para la alumna de doctorado en Ciencias Computacionales EYDI LÓPEZ HERNÁNDEZ, que solicitaron en el Departamento de Informática y Estadística de la Universidad Rey Juan Carlos (URJC) de Madrid, España, de la cual soy profesor e investigador.

La estancia tendrá como título "ESTUDIO DE LOS MODELOS DE APRENDIZAJE PROFUNDO PARA EL ANÁLISIS DE DATOS ESPECTRALES" y es parte de la tesis "MODELO DE DIFERENCIACIÓN DE OREJAS HUMANAS NATURALES, MEDIANTE IMÁGENES" de la estudiante de doctorado Eydi López. El periodo considerado para llevarla a cabo es del 01 de septiembre al 14 de octubre de 2025. Al final de la estancia se tiene programado redactar un reporte técnico.

Aprovecho para comentarle que estoy convencido de que la estancia de EYDI LÓPEZ HERNÁNDEZ será de gran valía para nuestra universidad, así como su centro de investigación, y redituará a su vez en una formación académica de excelencia para la tesis.

Sin más que agregar, aprovecho para enviarle un cordial y afable saludo.

ATENTAMENTE

Dr. Gerardo Reyes Salgado
Profesor
URJC, España

FIRMADO POR: GERARDO REYES SALGADO	FECHA FIRMA: 12/02/2025
Documento firmado digitalmente - Universidad Rey Juan Carlos - C/Alcalá 39, 28032 Móstoles Universidad Rey Juan Carlos	



Ciudad Juárez, Chihuahua a 14 de febrero de 2025

DRA. ANDREA MAGADÁN SALAZAR
TecnM/CENIDET
PRESENTE

Asunto: Carta de aceptación de estancia

Por medio de la presente le mando un cordial saludo. Además, aprovecho el medio para informarle que con mucho gusto recibirá a la M.C. Eydi López Hernández para realizar una estancia académica en el Campus del Instituto de Ingeniería y Tecnología de la Universidad Autónoma de Ciudad Juárez.

El proyecto para desarrollar en la estancia académica se titula "Modelo de Diferenciación de orejas humanas naturales, mediante imágenes". La estancia se realizará del 17 de febrero al 28 de marzo de 2025 y el plan de trabajo propuesto me parece adecuado para alcanzar los objetivos de la estancia.

Vale la pena recordar que la Universidad no puede proporcionar financiamiento de ningún tipo. Sin embargo, sin ningún problema se tendrá acceso a las instalaciones de la UACJ. Aprovecho la oportunidad para mencionarle que estoy convencido de que la estancia de Eydi será de gran valía para nuestras instituciones.

Sin otro particular por el momento, me despido de Usted, quedando a sus órdenes, en caso de existir alguna duda o aclaración.

ATENTAMENTE

"POR UNA VIDA CIENTÍFICA
POR UNA CIENCIA VITAL"

DR. OSSLAN OSIRIS VERGARA VILLEGAS
Catedrático del Departamento de Ingeniería Industrial y Manufactura
Director del Laboratorio de Visión por Computadora y Realidad Aumentada
Universidad Autónoma de Ciudad Juárez



Artículo
Method for Human Ear Localization in Controlled and Uncontrolled Environments

Eydi López-Hernández¹, Andrea Magadán-Salazar^{1,*}, Raúl Pinto-Elián¹, Nimrod González-Franco¹ and Miguel A. Zamiga-García²

¹ Tecnología Nacional de México/CENIDET, Cuernavaca 62000, México, México; eydi@cenidet.mx (E.L.); andrea@cenidet.mx (A.M.); raul@cenidet.mx (R.P.E.); nimrod@cenidet.mx (N.G.F.); miguel@cenidet.mx (M.A.Z.G.)

² FC Energy Solutions, Norman, OK 73071, USA; angel@fcenergy.com

* Correspondence: andrea@cenidet.mx

Abstract: One of the fundamental stages in recognizing people by their ears, which most works omit, is localizing the area of interest. The sets of images used for experiments generally contain only the ear, which is not appropriate for application in a real environment, where the visual field may contain part of the entire face, a human body or objects other than the ear. Therefore, determining the exact area where the ear is located is complicated, mainly in uncontrolled environments. This paper proposes a method for ear localization in controlled and uncontrolled environments using MediaPipe, a tool for face localization, and YOLOv8n architecture for detecting the ear. The proposed method first determines whether there are ears that indicate that a face exists in an image, and then, using the MediaPipe facial mesh, the points where an ear potentially exists are obtained. The extracted points are employed to determine the ear length based on the proportions of the human body proposed by Leonardo Da Vinci. Once the dimensions of the ear are obtained, the definition of the area of interest is carried out. If the required elements are not found, the model uses the YOLOv8n architecture module, trained to recognize ears in controlled environments. We employed four datasets for testing: (i) the real face Database, (ii) ITD Data Ear Database, (iii) ARI Ear Database, and (iv) EarVNI. Also, we used images from the Internet and some acquired using a Redmi Note 11 cell phone camera. An accuracy of 97% with an error of 3% was obtained with the proposed method, which is a competitive measure considering that some were conducted in controlled and uncontrolled environments, unlike state-of-the-art methods.

Keywords: localization of the area of interest; ear biometric systems; MediaPipe; YOLOv8n

MSC: 68D10

Citation: López-Hernández, E.; Magadán-Salazar, A.; Pinto-Elián, R.; González-Franco, N.; Zamiga-García, M.A. Method for Human Ear Localization in Controlled and Uncontrolled Environments. *Mathematics* **2024**, *12*, 1042. <https://doi.org/10.3390/math12091042>
Academic Editor: Jonathan Buckley

- ✓ Realización de estancia en el extranjero dentro de la Universidad Rey Juan Carlos (España), del 01 de septiembre al 14 de octubre de 2025. Donde se realizaron actividades de evaluación de técnicas de aprendizaje profundo para el análisis de datos obtenidos de imágenes en escenarios diferentes a los estudiados, bajo la dirección del Dr. Gerardo Reyes Salgado.

- ✓ Realización de estancia nacional en la Universidad Autónoma de Ciudad Juárez, del 17 de febrero al 28 de marzo de 2025. Donde se trabajó bajo la dirección del Dr. Osslan Osiris Vergara Villegas en la conclusión del proyecto de tesis titulado "Modelo de Diferenciación de orejas humanas naturales, mediante imágenes."

- ✓ Publicación del artículo "Method for Human Ear Localization in Controlled and Uncontrolled Environments", en la revista MDPI **mathematics**, indizada en el JCR. Se encuentra en el primer cuartil (Q1). Artículo aceptado y publicado el 01 de abril del 2024.

- ✓ Participación en el Coloquio de Inteligencia Artificial 2024 con la ponencia “Reconocimiento de identidad mediante imágenes de la oreja”. Que se llevó a cabo en Chilpancingo Gro el 19 de abril del 2024.



- ✓ Participación en la exposición de posters, Escuela de Inteligencia Computacional y Robótica 2023 realizada del 10 al 14 de octubre.



- ✓ Participación en el evento del 8 marzo 2023 “Orgullosa de ser mujer”, que se realizó en las instalaciones del CENIDET.



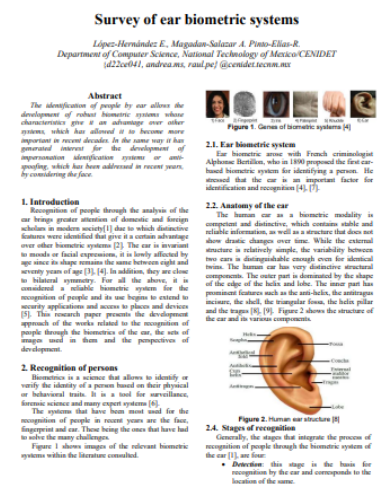
- ✓ Participación en la Feria Nacional de Ciencia, Tecnología e Innovación. “Soy mujer y hago ciencia”. Llevada a cabo el día 11 de febrero del 2023 en curso en Jojutla Morelos.



- ✓ Participación con un poster en el evento 9JCyTA (9ª Jornada de Ciencia y Tecnología Aplicada), el cual tiene por título “Reconocimiento de orejas humanas reales mediante análisis de imágenes”. Del 16 al 18 de noviembre del 2022.



- ✓ Elaboración del artículo “Survey of ear biometric systems” para el ICMEAE 2022 (2022 International Conference on Mechatronics, Electronics and Automotive Engineering); el cual fue aceptado y presentado el día 16 de diciembre del 2022.



7.5. Trabajo futuro

Con base en las limitaciones identificadas y en los resultados obtenidos, se proponen las siguientes líneas de trabajo futuro:

- Evaluar el desempeño del esquema ROI+PAD bajo condiciones que incluyan oclusión severa (>60 %), variaciones extremas de iluminación y degradación por movimiento, con el fin de estimar su robustez fuera del escenario experimental base.
- Ampliar el conjunto de ataques de presentación incorporando artefactos físicos no convencionales y materiales con propiedades ópticas/texturales distintas (p. ej., impresiones con acabados variables, prótesis con diferentes texturas), para incrementar la cobertura de casos atípicos.
- Explorar estrategias de mejora del rendimiento en oclusión mediante modelos complementarios o combinaciones ligeras (ensemble) y comparar su costo computacional frente a la ganancia en métricas de PAD, evitando incrementar de manera innecesaria la complejidad del sistema.
- Realizar análisis explicables comparativos (por ejemplo, Grad-CAM bajo distintos accesorios/oclusiones) para identificar regiones de atención del modelo y guiar posibles ajustes en preprocesamiento, aumento de datos y diseño del canal de textura.

REFERENCIAS

- Abaza, A., Ross, A., Hebert, C., Harrison, M. A. F., & Nixon, M. S. (2013). A survey on ear biometrics. *ACM Computing Surveys*, 45(2), 1–35. <https://doi.org/10.1145/2431211.2431221>
- Abdullakutty, F., Elyan, E., & Johnston, P. (2021). A review of state-of-the-art in Face Presentation Attack Detection: From early development to advanced deep learning and multi-modal fusion methods. *Information Fusion*, 75, 55–69. <https://doi.org/10.1016/J.INFFUS.2021.04.015>
- Aggarwal, C. C. (2018). *Neural Networks and Deep Learning*. Springer. <https://doi.org/doi.org/10.1007/978-3-319-94463-0>
- Aiadi, O., Khaldi, B., & Saadeddine, C. (2022). MDFNet: an unsupervised lightweight network for ear print recognition. *Journal of Ambient Intelligence and Humanized Computing*, 14(10), 1. <https://doi.org/10.1007/S12652-022-04028-Z>
- Alagarsamy, S. B., & Murugan, K. (2020). Ear recognition system using adaptive approach Runge-Kutta (AARK) threshold segmentation with cart classifier. *Multimedia Tools and Applications*, 79(15–16), 10445–10459. <https://doi.org/10.1007/s11042-019-7418-6>
- Alkababji, A. M., & Mohammed, O. H. (2021). Real time ear recognition using deep learning. *Telkomnika (Telecommunication Computing Electronics and Control)*, 19(2), 523–530. <https://doi.org/10.12928/TELKOMNIKA.v19i2.18322>
- Alqaralleh, E., & Toygar, Ö. (2018). Ear Recognition Based on Fusion of Ear and Tragus Under Different Challenges. *International Journal of Pattern Recognition and Artificial Intelligence*, 32(9). <https://doi.org/10.1142/S0218001418560098>
- Ametefe, D. S., Sarnin, S. S., Ali, D. M., & Zaheer, M. Z. (2022). Fingerprint Liveness Detection Schemes: A Review on Presentation Attack. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 10(2), 217–240. <https://doi.org/10.1080/21681163.2021.2012826>
- Anagnostopoulos, N. A. (2013). *Ear Recognition*. Springer London. <https://doi.org/10.1007/978-1-4471-5370-2>
- Bansal, M., & Madasu, H. (2021). Ear-based authentication using information sets and information modelling. *Soft Computing*, 25(16), 11123–11138. <https://doi.org/10.1007/s00500-021-05858-3>
- Benouis, M., & Regouid, M. (2020). Shifted 1D-LBP based ear recognition system. *2020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP)*, 193–197. <https://doi.org/10.1109/CCSSP49278.2020.9151650>

- Benzaoui, A., Khaldi, Y., Bouaouina, R., Amrouni, N., Alshazly, H., & Ouahabi, A. (2023). A Comprehensive survey on ear recognition: Databases, approaches, comparative analysis, and open challenges. *Neurocomputing*, 537, 236–270. <https://doi.org/10.1016/J.NEUCOM.2023.03.040>
- Bertillon, A. (1853-1914). A. du texte. (1890). *La photographie judiciaire, avec un appendice sur la classification et l'identification anthropométriques / par Alphonse Bertillon*. <https://gallica.bnf.fr/ark:/12148/bpt6k30435611>
- Bokade, G. U., & Kanphade, R. D. (2022). An ArmurMimus multimodal biometric system for Khosher authentication. *Concurrency and Computation: Practice and Experience*, 34(18). <https://doi.org/10.1002/cpe.7011>
- Booyens, A., & Viriri, S. (2022). Ear Biometrics Using Deep Learning: A Survey. In *Applied Computational Intelligence and Soft Computing* (Vol. 2022). Hindawi Limited. <https://doi.org/10.1155/2022/9692690>
- Bouaouina, R., Benzaoui, A., Doghmane, H., & Brik, Y. (2024). Boosting the Performance of Deep Ear Recognition Systems Using Generative Adversarial Networks and Mean Class Activation Maps. *Applied Sciences*, 14(10), 4162. <https://doi.org/10.3390/APP14104162>
- Chaves, D., Saikia, S., Fernández-Robles, L., Alegre, E., & Trujillo, M. (2018). A systematic review on object localisation methods in images. In *RIAI - Revista Iberoamericana de Automatica e Informatica Industrial* (Vol. 15, Number 3, pp. 231–242). Universitat Politècnica de Valencia. <https://doi.org/10.4995/riai.2018.10229>
- da Vinci, L. (1490). *Study of the proportions of the human body, known as the Vitruvian Man*. <https://www.britannica.com/topic/Vitruvian-man>
- da Vinci, L. (1883). *The literary works of Leonardo da Vinci* (J. P. Richter, Ed. & Tran.). <https://archive.org/details/literaryworksof02leon/page/70/mode/2up>
- Dai, G., Fan, J., Yan, S., & Li, R. (2022). Research on Detecting Potato Sprouting Based on Improved YOLOV5. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2022.3192406>
- Dodge, S., Mounsef, J., & Karam, L. (2018). Unconstrained ear recognition using deep neural networks. *IET Biometrics*, 7(3), 207–214. <https://doi.org/10.1049/iet-bmt.2017.0208>
- Emeršič, Ž., Meden, B., Peer, P., & Štruc, V. (2020). Evaluation and analysis of ear recognition models: performance, complexity and resource requirements. *Neural Computing and Applications*, 32(20), 15785–15800. <https://doi.org/10.1007/s00521-018-3530-1>
- Epic Games, I. (2024). *MetaHuman Creator*. <https://metahuman.unrealengine.com/>

- Espinoza-Vicuña, C. E., & Nuñez-Medrano, Y. (2023). Dyslexia recognition by handwriting using Convolutional Neural Networks. *Proceedings of the LACCEI International Multi-Conference for Engineering, Education and Technology, 2023-July*. <https://doi.org/10.18687/laccei2023.1.1.1346>
- Fatemifar, S., Rahimzadeh Arashloo, S., Awais, M., & Kittler, J. (2019). *Spoofing Attack Detection by Anomaly Detection*. IEEE Signal Processing Society.
- Freire-Obregón, D., De Marsico, M., Barra, P., Lorenzo-Navarro, J., & Castrillón-Santana, M. (2023). Zero-shot ear cross-dataset transfer for person recognition on mobile devices. *Pattern Recognition Letters*, 166, 143–150. <https://doi.org/10.1016/j.patrec.2023.01.012>
- Freire-Obregón, D., Neves, J., Emeršič, Ž., Meden, B., Castrillón-Santana, M., & Proença, H. (2025). Synthesizing multilevel abstraction ear sketches for enhanced biometric recognition. *Image and Vision Computing*, 154, 105424. <https://doi.org/10.1016/J.IMAVIS.2025.105424>
- Fukuda, R., Yokoyanagi, Y., Prathom, C., & Okada, Y. (2025). Toward Personal Identification Using Multi-Angle-Captured Ear Images: A Feasibility Study. *Applied Sciences (Switzerland)*, 15(6), 3329. <https://doi.org/10.3390/app15063329>
- Ganapathi, I. I., Ali, S. S., & Prakash, S. (2020). Geometric statistics-based descriptor for 3D ear recognition. *The Visual Computer*, 36(1), 161–173. <https://doi.org/10.1007/s00371-018-1593-8>
- Ganapathi, I. I., Ali, S. S., Prakash, S., Bakshi, S., & Werghi, N. (2024). B3D-EAR: Binarized 3D descriptors for ear-based human recognition. *Expert Systems with Applications*, 249, 123580. <https://doi.org/10.1016/J.ESWA.2024.123580>
- Ganesan, K., A, C., Ganapathi, I. I., Javed, S., & Werghi, N. (2023). Multimodal hybrid features in 3D ear recognition. *Applied Intelligence*, 53, 11618–11635. <https://doi.org/10.1007/s10489-022-04071-5>
- Gonzalez, E., Alvarez, L., & Mazorra, L. (2012). *AMI ear database [Data set]*. University of Las Palmas de Gran Canaria. https://webctim.ulpgc.es/research_works/ami_ear_database/
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. <https://www.deeplearningbook.org>
- Guyon, I., & Elisseeff, A. (2006). An introduction to feature extraction. *Studies in Fuzziness and Soft Computing*, 207, 1–25. https://doi.org/10.1007/978-3-540-35488-8_1/SAVE-RESEARCH
- Hassaballah, M., Alshazly, H. A., & Ali, A. A. (2019). Ear recognition using local binary patterns: A comparative experimental study. *Expert Systems with Applications*, 118, 182–200. <https://doi.org/10.1016/j.eswa.2018.10.007>

- Hassaballah, M., Alshazly, H. A., & Ali, A. A. (2020). Robust local oriented patterns for ear recognition. *Multimedia Tools and Applications*, 79(41–42), 31183–31204. <https://doi.org/10.1007/s11042-020-09456-7>
- Hurley, D. J., Arbab-Zavar, B., & Nixon, M. S. (2008). The Ear as a Biometric. *European Signal Processing Conference*, 131–150. https://doi.org/10.1007/978-0-387-71041-9_7
- IIT Delhi Ear Database. (2004, January). https://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Ear.htm
- International Organization for Standardization. (2023a). *Information technology - Biometric presentation attack detection - Part 1: Framework (ISO/IEC Standard No. 30107-1)*. <https://www.iso.org/standard/83828.html>
- International Organization for Standardization. (2023b). *Information technology - Biometric presentation attack detection - Part 3: Testing and reporting (ISO/IEC Standard No. 30107-3)*. <https://www.iso.org/standard/79520.html>
- International Telecommunication Union Radiocommunication Sector. (2011). *Recommendation ITU-R BT.601-7: Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios*. <https://www.itu.int/rec/R-REC-BT.601-7-201103-I/en>
- Jocher, G., & Ultralytics. (2023). Ultralytics YOLOv5 [Software]. In *GitHub, Inc.* GitHub. <https://github.com/ultralytics/yolov5>
- Kamboj, A., Rani, R., & Nigam, A. (2021). CG-ERNet: a lightweight Curvature Gabor filtering based ear recognition network for data scarce scenario. *Multimedia Tools and Applications*, 80(17), 26571–26613. <https://doi.org/10.1007/s11042-020-10264-2>
- Kamboj, A., Rani, R., & Nigam, A. (2022a). A comprehensive survey and deep learning-based approach for human recognition using ear biometric. In *The Visual Computer* (Vol. 38, Number 7, pp. 2383–2416). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s00371-021-02119-0>
- Kamboj, A., Rani, R., & Nigam, A. (2022b). *EIQA: ear image quality assessment using deep convolutional neuralnetwork* (Vols. 47, 245). Sādhana. <https://doi.org/10.1007/s12046-022-02017-8>
- Khalidi, Y., & Benzaoui, A. (2021). A new framework for grayscale ear images recognition using generative adversarial networks under unconstrained conditions. *Evolving Systems*, 12(4), 923–934. <https://doi.org/10.1007/s12530-020-09346-1>
- Khiarak, J. N. (2021). *Transfer learning using deep neural networks for Ear Presentation Attack Detection: New Database for PAD*. <https://github.com/Jalilnkh/KartalOI-EAR-PAD>

- Kohavi, R., & Provost, F. (1998). Glossary of Terms. *Machine Learning*, 30(3), 271–274.
- Lei, Y., Pan, D., Feng, Z., & Qian, J. (2023). Lightweight YOLOv5s Human Ear Recognition Based on MobileNetV3 and Ghostnet. *Applied Sciences* 2023, Vol. 13, Page 6667, 13(11), 6667. <https://doi.org/10.3390/AP13116667>
- Li, S., Li, Y., Li, Y., Li, M., & Xu, X. (2021). YOLO-FIRI: Improved YOLOv5 for Infrared Image Object Detection. *IEEE Access*, 9, 141861–141875. <https://doi.org/10.1109/ACCESS.2021.3120870>
- Losardo, D. R. J., Murcia, D. M., Tamaris, V. L., & Hurtado De Mendoza, W. (2015). Canon de las proporciones humanas y el Hombre de Vitruvio. *Revista de La Asociación Médica Argentina*, 128, 17–22.
- M N, S., & Prasanna, K. (2024). Transfer Learning with Optimization Enabled Person Recognition using Ear Images. *International Journal of Intelligent Systems and Applications in Engineering*, 12(16s), 66–76. <https://www.ijisae.org/index.php/IJISAE/article/view/4784>
- MediaPipe. (2022). <https://mediapipe.dev/>
- Meta. (2024). *Meta | Social Metaverse Company*. <https://about.meta.com/>
- Mo, S., Lu, P., & Liu, X. (2022). AI-Generated Face Image Identification with Different Color Space Channel Combinations. *Sensors*, 22(21). <https://doi.org/10.3390/s22218228>
- Muhammad, U., Laaksonen, J., Romaissa Beddiar, D., & Oussalah, M. (2024). Domain Generalization via Ensemble Stacking for Face Presentation Attack Detection. *International Journal of Computer Vision*, 132(12), 5759–5782. <https://doi.org/10.1007/S11263-024-02152-1/FIGURES/7>
- Nicholson, P. J. (2019). Art and occupation: Leonardo da Vinci, the proportions of the human figure (after vitruvius), c 1490. In *Occupational Medicine* (Vol. 69, Number 2, pp. 86–88). Oxford University Press. <https://doi.org/10.1093/occmed/kqy166>
- Nourmohammadi-Khiarak, J., & Pacut, A. (2019). An Ear Anti-spoofing Database With Various Attacks. *2019 International Conference on Biometrics (ICB)*.
- Ojala, T., Pietika, M., & Ma, T. (2002). Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971–987. <https://doi.org/10.1109/TPAMI.2002.1017623>
- Omara, I., Li, F., Zhang, H., & Zuo, W. (2016). A novel geometric feature extraction method for ear recognition. *Expert Systems with Applications*, 65, 127–135. <https://doi.org/10.1016/j.eswa.2016.08.035>

- OpenAI. (2024). *ChatGPT (Versión de Feb 2024) [Modelo de lenguaje a gran escala]*. OpenAI. <https://openai.com/>
- OpenCV Team. (2022). *Open Source Computer Vision Library (Versión 4.6.0) [Software]*. OpenCV. <https://opencv.org/>
- Oyebiyi, O. G., Abayomi-Alli, A., Arogundade, O., Tale, Qazi, A., Imoize, A. L., & Awotunde, J. B. (2023). A Systematic Literature Review on Human Ear Biometrics: Approaches, Algorithms, and Trend in the Last Decade. In *Information (Switzerland)* (Vol. 14, Number 3). MDPI. <https://doi.org/10.3390/info14030192>
- Park, S. Y., Kim, J. H., Chang, J. H., Park, J. M., Choi, C. H., & Kim, J. I. (2022). Quantitative evaluation of radiodermatitis following whole-breast radiotherapy with various color space models: A feasibility study. *PLOS ONE*, 17(3)(3), e0264925. <https://doi.org/10.1371/JOURNAL.PONE.0264925>
- Pinterest. (2024). *Pin page*. <https://mx.pinterest.com/pin/539165386656700280/>
- Pooshideh, M., Beheshti, A., Qi, Y., Farhood, H., Simpson, M., Gatland, N., & Soltany, M. (2024). Presentation Attack Detection: A Systematic Literature Review. *ACM Computing Surveys*, 57(1). <https://doi.org/10.1145/3687264>
- PyTorch Foundation. (2024). *PyTorch: An open source machine learning framework [Software]*. PyTorch Foundation. <https://pytorch.org/>
- Qin, Z., Zhao, P., Zhuang, T., Deng, F., Ding, Y., & Chen, D. (2022). A survey of identity recognition via data fusion and feature learning. *Information Fusion*, 91, 694–712. <https://doi.org/10.1016/j.inffus.2022.10.032>
- Ramos-Cooper, S., Gomez-Nieto, E., & Camara-Chavez, G. (2022). VGGFace-Ear: An Extended Dataset for Unconstrained Ear Recognition†. *Sensors*, 22(5). <https://doi.org/10.3390/s22051752>
- Raveane, W., Galdámez, P. L., & Arrieta, M. A. G. (2019). Ear Detection and Localization with Convolutional Neural Networks in Natural Images and Videos. *Processes 2019, Vol. 7, Page 457*, 7(7), 457. <https://doi.org/10.3390/PR7070457>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016, June 8). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. <http://arxiv.org/abs/1506.02640>
- Sajadi, S., & Fathi, A. (2020). Genetic algorithm based local and global spectral features extraction for ear recognition. *Expert Systems with Applications*, 159. <https://doi.org/10.1016/j.eswa.2020.113639>

- Sarangi, P. P., Mishra, B. S. P., & Dehuri, S. (2019). Fusion of PHOG and LDP local descriptors for kernel-based ear biometric recognition. *Multimedia Tools and Applications*, 78(8), 9595–9623. <https://doi.org/10.1007/s11042-018-6489-0>
- scikit-image development team. (2024). *scikit-image: Image processing in Python [Software]*. scikit-image. <https://scikit-image.org/>
- Scikit-learn. (2026). *Feature extraction*. https://scikit-learn.org/stable/modules/feature_extraction.html
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization*. <http://gradcam.cloudcv.org>
- Sepas-Moghaddam, A., Pereira, F., & Correia, P. L. (2018). Ear recognition in a light field imaging framework: A new perspective. *IET Biometrics*, 7(3), 224–231. <https://doi.org/10.1049/iet-bmt.2017.0204>
- Shaheed, K., Szczuko, P., Kumar, M., Qureshi, I., Abbas, Q., & Ullah, I. (2024). Deep learning techniques for biometric security: A systematic review of presentation attack detection systems. *Engineering Applications of Artificial Intelligence*, 129, 107569. <https://doi.org/10.1016/J.ENGAPPAI.2023.107569>
- Sharma, D., & Selwal, A. (2021). FinPAD: State-of-the-art of fingerprint presentation attack detection mechanisms, taxonomy and future perspectives. *Pattern Recognition Letters*, 152, 225–252. <https://doi.org/10.1016/J.PATREC.2021.10.013>
- Sharma, D., & Selwal, A. (2023). A survey on face presentation attack detection mechanisms: hitherto and future perspectives. *Multimedia Systems*, 29(3), 1527. <https://doi.org/10.1007/S00530-023-01070-5>
- Simonyan, K., & Zisserman, A. (2015, September 4). Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations (ICLR 2015)*. <http://arxiv.org/abs/1409.1556>
- Singh, R. P., Dash, R., & Mohapatra, R. K. (2023). LBP and CNN feature fusion for face anti-spoofing. *Pattern Analysis and Applications*, 26(2), 773–782. <https://doi.org/10.1007/s10044-023-01132-4>
- Sivanarain, K., & Viriri, S. (2020, November 25). Ear Recognition based on Local Texture Descriptors. *2020 2nd International Multidisciplinary Information Technology and Engineering Conference, IMITEC 2020*. <https://doi.org/10.1109/IMITEC50163.2020.9334147>
- Soares, T., Fernandes, C., Barbosa, C., Vaz, M. A. P., Reis, T., & Figueiral, M. H. (2024). Surface Roughness of Polyetheretherketone Printed by Fused Deposition Modeling: A Pilot

- Study Investigating the Impact of Print Layer Thickness and Polishing Method. *Applied Sciences (Switzerland)*, 14(7). <https://doi.org/10.3390/app14073096>
- Szeliski, R. (2011). *Computer Vision: Algorithms and Applications* (D. Gries & F. B. Schneider, Eds.). Springer. <https://doi.org/10.1007/978-1-84882-935-0>
- Temoshok, D., Fenton, J. L., Choong, Y.-Y., Lefkovitz, N., Regenscheid, A., Galluzzo, R., & Richer, J. P. (2025). *NIST Special Publication NIST SP 800-63B-4 Digital Identity Guidelines Authentication and Authenticator Management*. <https://doi.org/10.6028/NIST.SP.800-63B-4>
- Toprak, İ., & Toygar, Ö. (2020). Ear anti-spoofing against print attacks using three-level fusion of image quality measures. *Signal, Image and Video Processing*, 14(2), 417–424. <https://doi.org/10.1007/s11760-019-01570-w>
- Toprak, İ., & Toygar, Ö. (2021). Detection of spoofing attacks for ear biometrics through image quality assessment and deep learning. *Expert Systems with Applications*, 172. <https://doi.org/10.1016/j.eswa.2021.114600>
- Truong Hoang, V. (2019). *EarVN1.0 [Data set]*. Mendeley Data. Mendeley Data. <https://doi.org/10.17632/YWS3V3MWX3.3>
- Van Der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). Scikit-image: Image processing in python. *PeerJ*, 2014(1). <https://doi.org/10.7717/PEERJ.453>
- Wagner, M., Fugger, O., Foschum, F., & Kienle, A. (2024). Development of silicone-based phantoms for biomedical optics from 400 to 1550 nm. *Biomedical Optics Express*, 15(11), 6561–6572. <https://doi.org/10.1364/BOE.533481>
- Wang, Z., Yang, J., & Zhu, Y. (2021). Review of Ear Biometrics. *Archives of Computational Methods in Engineering*, 28(1), 149–180. <https://doi.org/10.1007/s11831-019-09376-2>
- Xu, X., Liu, C., Liu, Y., & Lu, L. (2024). Anti-Software Attack Ear Identification System Using Deep Feature Learning and Blockchain Protection. *Symmetry* 2024, Vol. 16, Page 85, 16(1), 85. <https://doi.org/10.3390/SYM16010085>
- Xu, X., Liu, Y., Liu, C., & Lu, L. (2023). A Feature Fusion Human Ear Recognition Method Based on Channel Features and Dynamic Convolution. *Symmetry* 2023, Vol. 15, Page 1454, 15(7), 1454. <https://doi.org/10.3390/SYM15071454>
- Youbi, Z., Boubchir, L., & Boukrouche, A. (2019). Human ear recognition based on local multi-scale LBP features with city-block distance. *Multimedia Tools and Applications*, 78(11), 14425–14441. <https://doi.org/10.1007/s11042-018-6768-9>

Zhou, Y., & Zaferiou, S. (2017). Deformable Models of Ears in-the-Wild for Alignment and Recognition. *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 626–633. <https://doi.org/10.1109/FG.2017.79>