



Técnicas de visión artificial aplicadas a tecnologías asistivas para personas con discapacidad visual: una revisión sistemática de la literatura

Computer vision techniques applied to assistive technologies for individuals with visual impairment: a systematic review literature

Brian Delfino Flores Méndez

TecNM/Instituto Tecnológico Superior de Xalapa, Xalapa, México

257002563@xalapa.tecnm.mx

ORCID: 0009-0005-4169-2378

María Salomé Alejandre Apolinar ✉

TecNM/Instituto Tecnológico Superior de Xalapa, Xalapa, México

salome.aa@xalapa.tecnm.mx

ORCID: 0009-0008-0232-8217

Hugo Amores Pérez

TecNM/Instituto Tecnológico Superior de Xalapa, Xalapa, México

hugo.ap@xalapa.tecnm.mx

ORCID: 0009-0009-0260-4872

Virginia Lagunes Barradas

TecNM/Instituto Tecnológico Superior de Xalapa, Xalapa, México

virginia.lb@xalapa.tecnm.mx

ORCID: 0000-0003-3312-7211

Miguel Ángel Hidalgo Reyes

TecNM/Instituto Tecnológico Superior de Xalapa, Xalapa, México

miguel.hr@xalapa.tecnm.mx

ORCID: 0000-0001-8303-6030



<https://doi.org/10.36825/RITI.14.34.008>

Recibido: Junio 08, 2026

Aceptado: Septiembre 25, 2026

Resumen: La discapacidad visual constituye una de las principales barreras para la inclusión social, afectando a más de 285 millones de individuos que padecen debilidad visual o ceguera, y los dispositivos de apoyo suelen ser costosos o poco accesibles. Por tal motivo, en el presente estudio se llevó a cabo una Revisión Sistemática de la Literatura (RSL) cuyo objetivo principal es identificar los modelos de inteligencia artificial más eficientes en el área de reconocimiento de texto, colores y objetos como frutas, billetes y expresiones faciales. Esta búsqueda

incluye artículos que se publicaron entre 2020 y 2026, explorando las principales bases de datos científicas indexadas como Redalyc, Biblioteca Digital ACM, Consensus, IEEE Xplore, Springer Natura Link y Google Académico. Durante la búsqueda, se identificaron y analizaron 46 artículos donde se identificó que las arquitecturas basadas en Redes Neuronales Convolucionales (CNN) y modelos de aprendizaje profundo son las más eficaces para procesar información visual a través de imágenes o videos en tiempo real sin gran necesidad de recursos elevados.

Palabras clave: *Inteligencia Artificial, Discapacidad Visual, Dispositivos Móviles, Computo en la Nube.*

Abstract: Visual impairment constitutes one of the primary barriers to social inclusion, affecting over 285 million individuals experiencing low vision or blindness, while assistive devices are frequently costly or lack accessibility. Consequently, the present study conducted a Systematic Literature Review (SLR) with the main objective of identifying the most efficient artificial intelligence models in the domain of text, color, and object recognition—specifically concerning items such as fruits, banknotes, and facial expressions. This research encompasses articles published between 2020 and 2026, exploring prominent indexed scientific databases including Redalyc, the ACM Digital Library, Consensus, IEEE Xplore, Springer Nature Link, and Google Scholar. Throughout the search process, 46 articles were identified and analyzed, revealing that architectures based on Convolutional Neural Networks (CNN) and deep learning models are the most effective for processing visual information through images or real-time video without demanding substantial computational resources.

Keywords: *Artificial Intelligence, Visual Impairment, Mobile Devices, Cloud Computing.*

1. Introducción

La visión es el sentido más dominante y la discapacidad visual afecta a más de 285 millones de personas alrededor de todo el mundo, donde menos de la mitad todavía no han recibido tratamiento médico, esta condición abarca desde la deficiencia visual moderada o baja visión hasta la ceguera total generando un impacto drástico en el neurodesarrollo durante la primera infancia, lo que limita la autonomía y capacidad de interacción social de las personas que la padecen. [1], [2]. Demográficamente, se reconoce como un problema de salud mundial de alta carga que afecta de manera desproporcionada a mujeres, personas de edad avanzada y poblaciones de bajos recursos, quienes enfrentan barreras importantes en actividades vitales como el manejo de finanzas o la navegación [3]. Una de la causa más común que genera ceguera o pérdida de visión irreversible en el mundo está relacionada con la Degeneración Macular Asociada con la Edad (DMAE) [4].

En 1990 se registraron aproximadamente 3,640,180 casos y un aumento en 2021 a 8,057,521 en casos mundiales de Años de Vida Ajustados por Discapacidad (AVAD) y DMAE, estos datos fueron obtenidos a través de un estudio en 2021 donde se analizó la carga de la enfermedad y lesión del GBD (*Global Burden of Disease*, o en español Carga Mundial de Morbilidad) abarcando 204 países y regiones, llegando a la conclusión de que entre 2020 y 2021 la carga mundial de enfermedad aumentó nuevamente después de 3 décadas, principalmente por la pandemia de COVID-19 que ocasionó sobrecargo de trabajo en los sistemas de servicios de salud [5].

Aunque existen algunas tecnologías diseñadas para apoyar a personas con discapacidad visual, muchas requieren de dispositivos especializados, por tal motivo, en esta RSL se tiene como propósito identificar las principales tecnologías de visión artificial asistivas que puedan utilizarse como medio para mejorar la autonomía e inclusión social en personas con discapacidad visual.

2. Estado del arte: Tecnologías Asistivas de Visión Artificial para Personas con Discapacidad Visual

La discapacidad visual es una condición que abarca la pérdida o disminución de las funciones visuales, la cual no puede ser corregida a través de medios habituales. La Organización Mundial de la Salud (OMS) la clasifica, según la agudeza visual mejor corregida, en deficiencia visual moderada, grave (agrupadas bajo el término “baja visión”) y ceguera total[2],[3]. Para mitigar las barreras de accesibilidad que genera esta condición, la visión artificial se ha posicionado como una herramienta tecnológica fundamental, desarrollando métodos que permiten a las computadoras ver, comprender e interpretar el contenido de imágenes digitales y videos [6].

El desarrollo de tecnologías asistivas visuales tienen fundamento en avances de análisis de imágenes médicas donde técnicas de aprendizaje automático y una Red Neuronal Convolucional (CNN) han transformado la comprensión y el análisis en patologías oculares [7]. Un paso importante antes de la asistencia tecnológica empieza con el diagnóstico preciso de condiciones que causan pérdida de visión donde se destacan implementaciones con inteligencia artificial multimodal para el cribado temprano de glaucoma a partir de imágenes oculares [8], así como de métodos eficientes para seleccionar e incrustar características visuales para la clasificación general de enfermedades oculares, destacando que en el área de la salud se requiere una gran precisión en los modelos de IA [9]. Adicionalmente, el uso de redes neuronales de fusión densa explicable y el aprendizaje profundo han logrado identificar lesiones en retinas mediante Tomografía de Coherencia Óptica [10].

Por otro lado, la robustez de las herramientas de asistencia depende de la evolución sobre el análisis de aplicaciones de visión artificial y la detección de objetos [11]. Donde actualmente el avance de las tecnologías asistivas para personas con discapacidad visual se ha fundamentado en la convergencia de la inteligencia artificial y el procesamiento de imágenes buscando dotar a los usuarios de mayor autonomía en su vida cotidiana [6].

En el área específica de la asistencia visual, provee los datos del entorno necesarios para que el usuario pueda adquirir independencia y capacidades de reconocimiento espacial [7]. El procesamiento y la interpretación de estos datos visuales se logran mediante arquitecturas avanzadas de aprendizaje profundo, donde históricamente las CNN han sido la técnica predilecta para la extracción de características, empleando una estructura multicapa inspirada en la corteza visual que procesa datos en cuadrícula (imágenes 2D) para detectar patrones espaciales y clasificar objetos con alta precisión [12].

Sin embargo, el paradigma de extracción ha evolucionado con la adopción de Vision Transformers (ViT), originalmente introducidos para traducciones automáticas mediante mecanismos de atención, aunque actualmente han sido adaptados para procesar secuencias de imágenes mediante auto-atención lo que permite capturar dependencias a largo plazo y contextos espaciales globales de manera más eficaz que un procesamiento de píxel por píxel [13]. El rendimiento de ambas arquitecturas es ampliamente potenciado por el aprendizaje por transferencia (*Transfer Learning*), donde se reutilizan habilidades adquiridas por modelos preentrenados con masivas bases de datos y aplicándolas a nuevas tareas afines, mediante el cual se acelera el entrenamiento y mejora la extracción de características espaciales sin necesidad de construir conjuntos de datos etiquetados [14, 15]. Además, para maximizar la robustez en entornos dinámicos, se emplea la fusión de características (*Ensemble Learning*), donde se combinan representaciones extraídas por múltiples redes (como ResNet y DenseNet) incrementando el poder discriminativo general del modelo [16, 17].

Más allá de la detección de obstáculos, la comprensión integral del entorno necesita una interpretación de elementos textuales y contextuales, por lo que se integra el Reconocimiento Óptico de Caracteres (OCR) como una tecnología impulsada por inteligencia artificial capaz de leer texto curvo o degradado en escenarios reales, y Modelos de Lenguaje Visual (VLM) como sistemas multimodales interactivos que pueden procesar imagen y texto de manera simultánea para generar respuestas descriptivas con un desempeño comparable al humano [18].

Finalmente, para que estos componentes de software conformen una herramienta de asistencia viable, deben traducirse en estímulos comprensibles y operar de forma autónoma, como el sistema TTS que actúa como mecanismo principal de salida donde convierte la información procesada visualmente en palabras habladas logrando una interfaz auditiva accesible [19]. Ecosistema tecnológico que converge a través del Internet de las Cosas (IoT) y los sistemas embebidos, como computadoras compactas de bajo consumo (por ejemplo, Raspberry Pi). Esta integración de IoT, procesamiento embebido y comunicaciones inalámbricas permite desplegar algoritmos pesados de inteligencia artificial directamente en dispositivos portátiles garantizando la emisión de alertas en tiempo real y una mejor autonomía del usuario [20, 21].

3. Materiales y métodos

La presente RSL se basa en el modelo de Barbara Kitchenham y Stuart M. Charters [22] para la creación de una RSL en el área de ingeniería de software, método que es mencionado en la RSL: Revisión de la literatura para el conteo de unidades formadoras de colonias en microorganismos mediante visión artificial [23].

3.1. Fase 1: Planificación

Pregunta de investigación

En esta primera etapa de análisis, se definieron preguntas de investigación para llevar a cabo la RSL con el objetivo de delimitar el alcance de la investigación con técnicas avanzadas de búsqueda para la recolección de artículos en el área de tecnologías asistivas basadas en visión artificial.

Como pregunta principal se plantea:

- ¿Qué modelos de visión artificial se utilizan en tecnologías asistivas orientadas a personas con discapacidad visual?

Como preguntas secundarias de apoyo a la pregunta principal se plantean las siguientes:

- ¿Qué técnicas de visión artificial se utilizan en tecnologías asistivas orientadas a personas con discapacidad visual?
- ¿Cuáles son los avances y limitaciones de la visión artificial en el reconocimiento de objetos?
- ¿Qué dispositivos se emplean como apoyo para personas con discapacidad visual?

Objetivos específicos

Además, en base a las preguntas anteriores se plantean los siguientes objetivos de este estudio:

- Identificar algoritmos y arquitecturas eficientes de visión artificial.
- Clasificar modelos y técnicas de visión artificial utilizados en desarrollo de tecnologías asistivas.
- Clasificar los tipos de dispositivos empleados como apoyo para personas con discapacidad visual
- Analizar limitaciones de aplicaciones con visión artificial en entornos asistivos.

Identificación de bases de datos

Se escogieron bases de datos científicas relevantes para realizar la investigación dando prioridad a las que poseen reconocimiento extenso y publicaciones indexadas. A continuación, se presenta en la Tabla 1 las fuentes de información seleccionadas junto a su criterio de elección.

Tabla 1. Bases de datos para consultar y recopilar artículos científicos.

Base de datos	Criterio
Google académico	Cobertura multidisciplinaria y acceso a artículos con alto índice de citación.
Biblioteca Digital ACM	Fuente especializada en informática y tecnología, contiene publicaciones revisadas por pares.
Redalyc	Plataforma de acceso abierto con enfoque en revistas científicas latinoamericanas
Consensus	Motor de búsqueda basado en IA, prioriza estudios indexados y resultados con evidencia científica. Se utilizó como auxiliar en la recopilación de artículos.
IEEE Xplore	Base reconocida en ingeniería y computación, contiene artículos de alto impacto y revisiones metodológicas rigurosas.
Springer Link	Editorial científica internacional, contiene publicaciones revisadas por pares y estudios recientes de alta calidad.

Fuente: Elaboración propia.

Para realizar la investigación y recopilación de artículos, se plantearon las siguientes palabras clave presentadas en la Tabla 2.

Tabla 2. Palabras clave y sus sinónimos.

Palabra clave	Sinónimos
<i>Artificial intelligence</i>	<i>Intelligent systems</i>
	<i>Machine Learning</i>
	<i>Computer vision</i>
	<i>Image processing</i>
	<i>Machine visión</i>
<i>Object detection</i>	<i>Object recognition</i>
	<i>Visual object identification</i>
<i>Visual impairment</i>	<i>Visual deficiency</i>
	<i>Partial or total blindness</i>

	<i>blindness</i>
<i>Assistive technologies</i>	<i>Assistive devices</i> <i>Support devices</i>
<i>Mobile applications</i>	<i>Mobile software</i> <i>Mobile app</i>

Fuente: Elaboración propia.

Se formularon diferentes cadenas de búsqueda a partir de las palabras clave anteriores y se consultaron en bases de datos científicas indexadas, las cuales se presentan a continuación en la Tabla 3 las más relevantes.

Tabla 3. Cadenas de búsqueda para la consulta y recopilación de artículos.

Código de cadena de búsqueda	Cadena de búsqueda
CB01	("artificial intelligence" OR "intelligent systems" OR "machine learning" OR "computer vision" OR "image processing" OR "machine vision") AND ("object detection" OR "object recognition" OR "visual object identification") AND ("visual impairment" OR "vision loss" OR "blindness") AND ("assistive technologies" OR "support technologies" OR "assistive devices" OR "support devices")
CB02	("computer vision" OR "image processing") AND ("assistive technology" OR "assistive device") AND ("visual impairment" OR "blindness")
CB03	("machine vision" OR "Image processing" OR "object detection" OR "Object recognition") AND ("assistive system" OR "assistive technology") AND ("visually impaired" OR "blindness")
CB04	"Artificial intelligence" AND "Computer vision" AND "Object detection" AND "Visual impairment" AND "Assistive technologies"
CB05	("Mobile applications" OR "Mobile app" OR "Mobile software") AND ("Artificial intelligence" OR "Machine learning" OR "Computer vision") AND ("Visual impairment" OR "Blindness")

Fuente: Elaboración propia

A continuación, se presenta en la Tabla 4 los resultados obtenidos en la consulta de cada cadena de búsqueda, cabe mencionar que los resultados presentados son los resultados totales antes de aplicar algún criterio de inclusión o exclusión.

Tabla 4. Resultados obtenidos en la consulta de artículos.

Base de datos	CB01	CB02	CB03	CB04	CB05	Total
Google académico	3,710	6,600	8,510	1,150	8,920	28,890
Biblioteca Digital ACM	374	456	470	126	567	1,993
Redalyc	2	1	33	1	10	47
IEEE Xplore	0	455	758	27	424	1,664
Springer Link	708	365	984	45	673	2,775
Total	4,794	7,877	10,755	1,349	10,594	35,369

Fuente: Elaboración propia.

3.2. Fase 2: Revisión de los estudios relevantes

Criterios de inclusión y exclusión

A continuación, en la Tabla 5 se definieron criterios de inclusión y exclusión con el objetivo de delimitar los artículos recabados y así obtener mejor calidad en los mismos.

Tabla 5. Criterios de inclusión y exclusión.

Criterio	Descripción
Tipo de estudios	Artículos científicos, artículos de divulgación, tesis
Fechas	2020 – 2026
Idioma	Inglés y español
Área	Visión artificial, tecnologías móviles asistivas, accesibilidad, discapacidad visual
Inclusión	Implementaciones, prototipos, pruebas con usuarios, frameworks/librerías
Exclusión	Artículos sin aplicaciones reales, RSL's, tecnologías de apoyo sin vision por computadora, opiniones, conferencias
Relevancia	Inclusión de artículos empíricos que aborden el desarrollo, implementación o evaluación de modelos de visión artificial

Fuente: Elaboración propia.

Aplicando la fecha de publicación como primer filtro de exclusión se nota una considerable reducción en los artículos publicados, resultados que se presentan a continuación en la Tabla 6.

Tabla 6. Resultados obtenidos después de aplicar las fechas de publicación como filtro de exclusión.

Base de datos	CB01	CB02	CB03	CB04	CB05	Total
Google académico	2,890	4,320	5,720	955	6,630	20,515
Biblioteca Digital ACM	321	342	363	115	486	1,627
Redalyc	1	1	16	0	8	26
IEEE Xplore	0	357	620	25	323	1,325
Springer Natura Link	556	227	628	33	427	1,871
Total	3,768	5,247	7,347	1,128	7,874	25,364

Fuente: Elaboración propia.

Como siguiente filtro, en la Tabla 7 se excluirán otras revisiones sistemáticas, además de incluir solo artículos con idioma de publicación en español e inglés.

Tabla 7. Resultados obtenidos al excluir revisiones sistemáticas y conferencias, e incluir artículos en español e inglés.

Base de datos	CB01	CB02	CB03	CB04	CB05	Total
Google académico	2,620	888	1,310	189	1,030	6,037
Biblioteca Digital ACM	60	46	58	8	29	201
Redalyc	0	0	12	0	6	28
IEEE Xplore	0	26	42	2	22	92
Springer Natura Link	54	31	181	18	176	460
Total	2,734	991	1,603	217	1,263	6,808

Fuente: Elaboración propia.

Como últimos filtros se incluirán artículos que aborden el área de visión artificial, tecnologías móviles asistivas, accesibilidad y discapacidad visual, posteriormente se seleccionarán algunos artículos que tengan mayor relevancia y aporte significativo a las preguntas de investigación y objetivos de este análisis sistemático, obteniendo los siguientes resultados de artículos que serán analizados, se presentan a continuación en la Tabla 8. Algo importante de mencionar, es que dentro de los resultados obtenidos en Google académico pueden aparecer artículos publicados en alguna de las otras bases de datos debido a que es un buscador y no una base de datos.

Tabla 8. Resultados finales para análisis de la literatura.

Base de datos	CB01	CB02	CB03	CB04	CB05	Total
Google académico	6	3	9	4	6	28
Biblioteca Digital ACM	1	0	1	0	1	3

Redalyc	0	0	1	0	0	1
IEEE Xplore	0	1	1	0	0	2
Springer Natura Link	3	1	4	1	3	12
Total	10	5	16	5	10	46

Fuente: Elaboración propia.

A continuación, en la Figura 1 se muestra de manera gráfica el proceso de búsqueda y selección de la literatura de acuerdo a la metodología PRISMA. Se muestran los registros identificados, los criterios de exclusión aplicados en cada etapa (cribado y elegibilidad) y finalmente los estudios incluidos en esta revisión donde se evaluó la calidad metodológica de los estudios seleccionados considerando criterios como los objetivos, la relevancia de los resultados y la actualidad de la información.

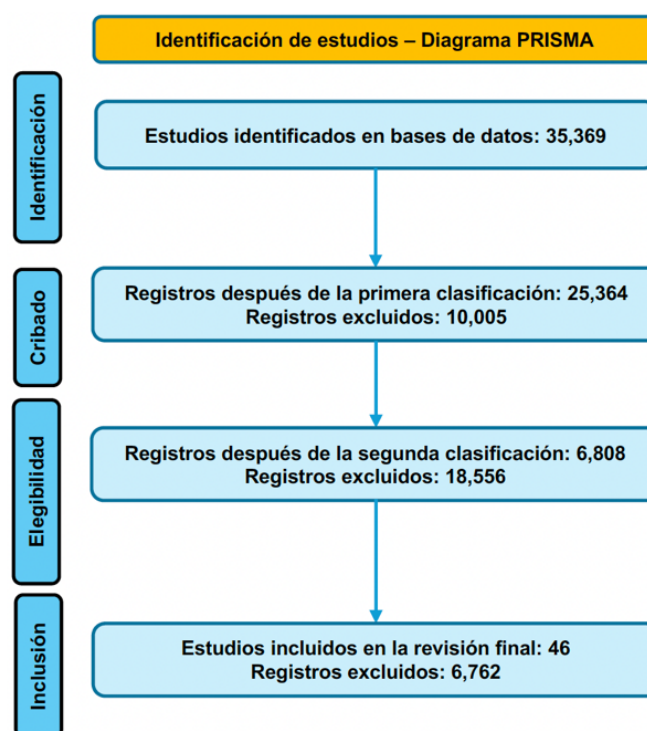


Figura 1. Diagrama PRISMA que muestra el flujo de selección de los estudios. Fuente: Elaboración propia.

4. Resultados

4.1. Fase 3: Síntesis y análisis de estudios relevantes

A continuación, se presentan los hallazgos de la literatura seleccionada, estructurados con base en las preguntas y objetivos de investigación planteados para este estudio.

La discapacidad se define como la ausencia o disfunción de un órgano o sistema causada por una enfermedad, lesión u otra condición de salud donde se requiere atención médica y rehabilitación para acercar a la persona al estado normativo del cuerpo [24]. La discapacidad visual se define como la pérdida o disminución significativa de las funciones visuales que no puede ser corregida mediante intervenciones habituales como el uso de lentes, afectando a más de 285 millones de individuos [25]. Esta condición abarca desde la deficiencia visual moderada o baja visión hasta la ceguera total generando un impacto drástico en el neurodesarrollo durante la primera infancia lo que limita las oportunidades de educación, independencia, aprendizaje y adaptación social [2]. Demográficamente, se reconoce como un problema de salud mundial de alta carga que afecta de manera desproporcionada a mujeres, personas de edad avanzada y poblaciones de bajos ingresos, enfrentando barreras importantes en actividades vitales como el manejo de finanzas o la navegación en interiores y exteriores [3].

Para mejorar la autonomía, la literatura evidencia la integración de Inteligencia Artificial en diversas plataformas de hardware donde se observan aplicaciones basadas en inteligencia artificial desplegadas en teléfonos

inteligentes con el objetivo de reconocer el entorno físico con altas tasas de precisión, funcionando como asistentes personales en tiempo real [26], donde se consolidan como la herramienta más accesible donde se pueden ejecutar aplicaciones móviles con asistencia por voz (TTS) que funcionan con o sin conexión a internet, como la aplicación “CopReader” para identificar papel moneda y “Vivid” para reconocer colores, textos y expresiones faciales [27]. Otra alternativa destacable es el Single Shot Detector (SSD), un modelo de una sola etapa que alcanza grandes niveles de precisión comparables con arquitecturas mucho más densas al prescindir de una red de propuesta de regiones [11]. Frecuentemente, se acopla con arquitecturas ligeras como MobileNet creando MobileNet-SSD para mantener una alta tasa de aciertos con un bajo costo computacional [19, 28].

Además del uso de móviles, existe un gran avance en el diseño de dispositivos vestibles (wearables), como gafas inteligentes equipadas con cámaras y sensores de profundidad (Vuzix Blade 2 y OrCam MyEye), bastones inteligentes con tecnología ultrasónica o láser, y prendas cognitivas como “Blind's Apron” [29, 30]. Para los bastones, fue necesario desarrollar diferentes prototipos físicos cuyo objetivo consistía en realizar una estimación precisa de la distancia hacia los obstáculos u objetos de interés, un punto crítico que necesitaba garantizar la movilidad segura [31]. Estos periféricos físicos se combinaban frecuentemente con sistemas embebidos de bajo costo, como una “Raspberry Pi” o “Intel Edison”, principalmente al emplear IoT para procesar los datos visuales y comunicarlos vía retroalimentación háptica (vibración) o auditiva [20, 32].

Además del reconocimiento de objetos, los resultados revelan un incremento dentro del enfoque hacia la inclusión cultural y cognitiva, principalmente en el diseño y desarrollo de dispositivos interactivos impulsados por IA capaces de soportar múltiples idiomas para facilitar que el usuario reciba la retroalimentación de su entorno en su lengua materna [33].

En el ámbito del desarrollo de tecnologías asistivas, las arquitecturas basadas en CNN y aprendizaje profundo son estándar para el reconocimiento del entorno donde modelos detectores de objetos de una sola etapa, como la familia YOLO en sus variantes más recientes como YOLOv8, YOLOv9, YOLO11 y YOLO26, dominan la literatura al ofrecer un excelente equilibrio entre velocidad de procesamiento en tiempo real y alta precisión [34, 35, 36].

En la versión más reciente de la familia YOLO, se destaca YOLOE-26 como una importante solución para mitigar limitaciones de entornos dinámicos en el mundo real, ya que a diferencia de modelos anteriores limitados a categorías previamente definidas, en YOLOE-26 se introduce la “segmentación de instancias de vocabulario abierto”, logrando que puede ser guiado dinámicamente mediante texto (fácilmente integrable con comandos de voz del usuario) para buscar objetos específicos, inusuales o en evolución que no estaban en la base de datos original del modelo. Además de demostrar en evaluaciones de las variantes más ligeras (YOLOE-26n y YOLOE-26s), ofrecen un gran equilibrio entre alta precisión de segmentación y latencia computacional muy baja gracias a su optimización para hardware de borde [36].

Además, como se muestra en la Figura 2, comparado con versiones anteriores YOLO26 ha demostrado una optimización superior entre precisión y velocidad (*accuracy-speed trade-off*) incluso con modelos basados en Transformadores, presentado en la Figura 3, destacando su gran ventaja en el diseño libre de Supresión de No-Máximos liderando la eficiencia del flujo de trabajo completo [36].

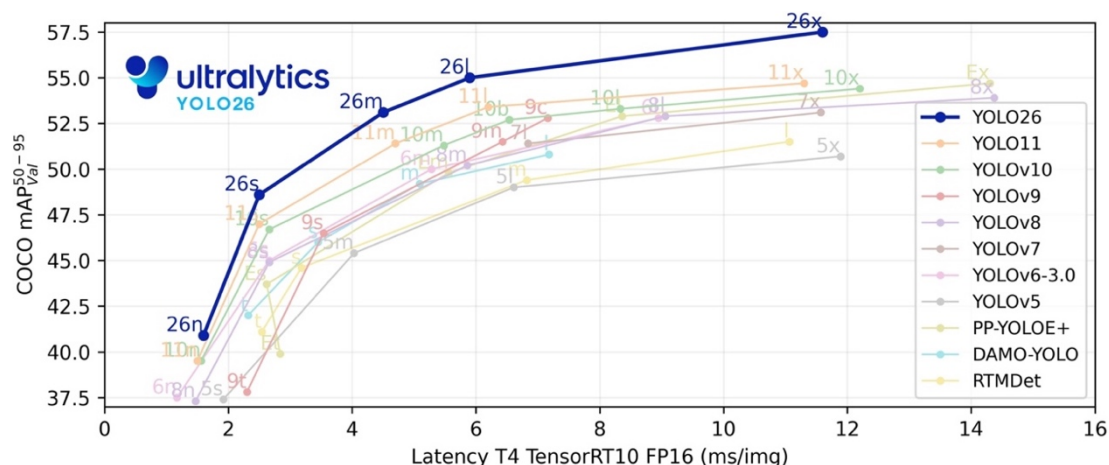


Figura 2. Comparación de la familia YOLO entre latencia pura de inferencia (milisegundos por imagen) y precisión de detección (métrica estándar COCO mAP). Fuente: <https://docs.ultralytics.com/models/yolo26>.

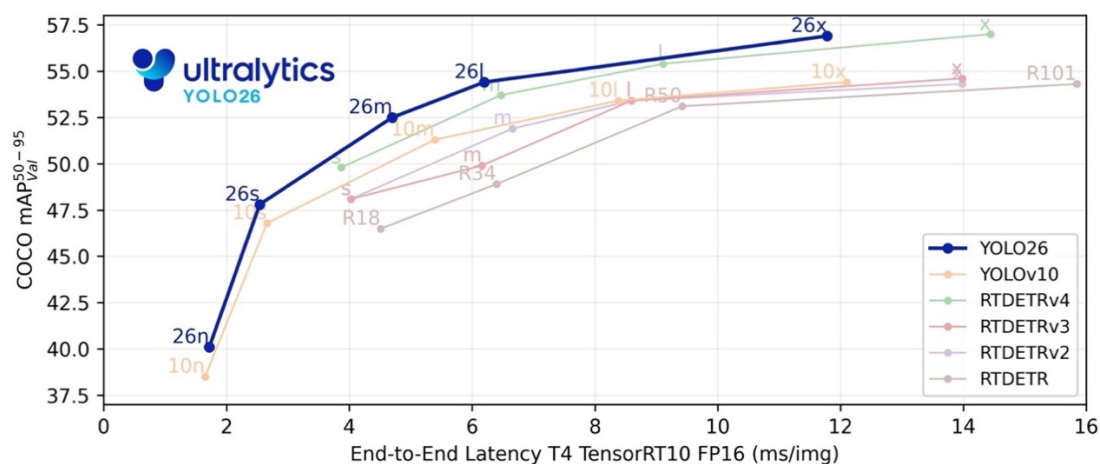


Figura 3. Comparación de YOLOv10, YOLO26 y variantes basadas en Transformadores (variantes RT-DETR), entre latencia de extremo a extremo y precisión de detección. Fuente: <https://docs.ultralytics.com/models/yolo26>.

Por otro lado, para aplicaciones móviles que necesitan eficiencia energética y un bajo consumo de memoria, algoritmos ligeros como SSD integrados con MobileNetV2 han logrado una precisión superior al 95% en la detección de objetos cotidianos y billetes [28, 37].

Para tareas altamente especializadas como la lectura de letreros, los enfoques híbridos utilizan algoritmos de Reconocimiento Óptico de Caracteres acoplados a Modelos de Lenguaje Visual (VLM) de última generación, logrando así interpretar textos complejos en escenarios desordenados [18]. Asimismo, algunos investigadores proponen modelos detectores de dos etapas como Faster R-CNN, y Transformadores de Visión que proporcionan diversas características de contexto global pero también son aplicados en ecosistemas Edge-Cloud [15].

Las tecnologías actuales presentan avances significativos que permiten tasas de detección, precisión y recuperación cercanas al 99% bajo condiciones controladas de experimentación [38]. Sin embargo, la implementación de Visión Artificial en la vida diaria de personas con discapacidad visual enfrenta diversas limitaciones en escenarios dinámicos del mundo real, donde las variaciones drásticas en la iluminación, las oclusiones parciales del objeto, el movimiento de la cámara o el ruido de fondo disminuyen considerablemente la confiabilidad y precisión del modelo [32, 39].

Además, el despliegue de redes neuronales profundas y densas presentan barreras de capacidad computacional y un alto consumo de batería al intentar ejecutarse directamente sobre dispositivos móviles o embebidos de bajo costo [35]. A nivel de usuario, la usabilidad a largo plazo no solo requiere detecciones numéricas exactas, sino también de una retroalimentación instantánea con un diseño que preserve la privacidad de la información captada por las cámaras en espacios públicos [40].

En la siguiente tabla, Tabla 9, se presenta una síntesis de los artículos que abordan visión artificial, así como los dispositivos empleados, principalmente abordan la detección de objetos en dispositivos móviles.

Tabla 9. Resultados finales para análisis de la literatura.

Autor(es)	Año	Modelo/Arquitectura de Visión Artificial	Reconocimiento abordado	Dispositivo Utilizado
Joshi, R. C., <i>et al.</i>	2020	Redes Neuronales Convolucionales	Reconocimiento multiobjeto, evasión de obstáculos y segmentación	Procesador Digital de Señales integrado con sensores ultrasónicos
Ashiq, F., <i>et al.</i>	2022	Redes Neuronales Convolucionales	Reconocimiento y seguimiento de objetos	Sistemas de asistencia portátiles
Alzahrani, N. y Al-Baity, H. H.	2023	Mask R-CNN	Detección y segmentación de objetos con salida de voz	Software de asistencia

Islam, R. B., <i>et al.</i>	2023	CNN, YOLO y ORB-SLAM	Detección de objetos y descripción del entorno	Gafas inteligentes y sistemas multisensoriales
Bolaños-Fernández, C. y Bacca-Cortes, E. B.	2024	SqueezeNet, MobileNet, SSD y YOLO	Identificación de moneda colombiana con retroalimentación de audio	Teléfonos inteligentes
Naz, S. y Jabeen, F.	2024	R-CNN, Faster R-CNN, SSD, YOLO, ViT y ResNet	Clasificación y evaluación para técnicas de detección	Dispositivos móviles
Adam, M., <i>et al.</i>	2025	Faster R-CNN, ResNet, DenseNet-20 y Bi-GRU	Detección, extracción y clasificación de objetos en interiores	Dispositivos de tecnología de asistencia inteligente
Alarood, A., <i>et al.</i>	2025	BViT-CKNN (Bilateral Vision Transformer y Convolution Kernel Neural Network)	Detección de obstáculos mediante datos de visión estéreo	Sistemas de navegación autónoma
Alashjaee, A. M., <i>et al.</i>	2025	YOLOv8, CapsNet y Red de Creencias Profundas (DBN) optimizada con TSA	Detección y localización de objetos en interiores	Sistema de asistencia inteligente
Alohali, M. A., <i>et al.</i>	2025	FFDGCRN-ROD (Faster R-CNN, CapsNet, SqueezeNet, Inceptionv3 y DA-GCRN)	Detección y clasificación de objetos en interiores	Dispositivos en entorno IoT con edge-cloud
Fernando, S., <i>et al.</i>	2025	YOLOv4, SSD, MobileNet y ResNet	reconocimiento de imágenes	Teléfonos inteligentes y gafas inteligentes
Ghanem, A. M., <i>et al.</i>	2025	YOLOv8, YOLOv9, YOLOv10, ResNet-50 y SqueezeNet	Identificación de la moneda egipcia	Dispositivos móviles y embebidos
Kumari, P. y Hammady, R.	2026	YOLOv8	Detección de objetos con retroalimentación de IA y audio	Gafas inteligentes

Fuente: Elaboración propia.

4.1. Limitaciones del estudio

Una limitación inherente a la metodología de esta revisión sistemática radica en la cobertura de las bases de datos consultadas. Si bien se incluyeron repositorios de alto impacto como IEEE Xplore, Biblioteca Digital ACM, Springer Link, Redalyc, y motores basados en IA como Consensus, existe la posibilidad de haber omitido estudios empíricos relevantes indexados exclusivamente en colecciones cerradas como Scopus o Web of Science.

Asimismo, la exclusión de literatura gris, como informes técnicos no publicados comercialmente, y la restricción a documentos en los idiomas inglés y español introduce un sesgo idiomático que podría dejar fuera avances tecnológicos desarrollados en otras regiones asiáticas o europeas, sin embargo, es importante distinguir estas restricciones documentales de las limitaciones tecnológicas y operativas de las soluciones analizadas.

Mientras que el diseño de búsqueda restringe la exhaustividad teórica de la revisión, los desafíos inherentes a los modelos de visión artificial donde la caída de confiabilidad ante oclusiones parciales, las variaciones extremas de iluminación en escenarios dinámicos y la alta demanda computacional en dispositivos móviles de borde pueden representar barreras directas para el despliegue de tecnologías asistivas.

5. Conclusiones

La discapacidad visual es una condición que abarca la baja visión hasta la ceguera total, lo que restringe mayormente la capacidad de las personas al realizar actividades de su vida cotidiana, afectando su autonomía e inclusión social. Frente a esta problemática de salud pública global, la literatura evidencia que la integración de tecnologías asistivas basadas en visión artificial representa una solución transformadora que permite a las personas mitigar barreras al percibir e interpretar su entorno de manera independiente [2].

El análisis de los resultados extraídos de la literatura confirma que la convergencia del aprendizaje profundo y la visión artificial además de demostrar ser computacionalmente eficiente, también poseen un impacto terapéutico y funcional directo en la evaluación clínica de tecnologías asistivas [42]. Además, la implementación de estos modelos reducen significativamente las barreras de autonomía para las personas con discapacidad visual [39]. Y demuestran que los modelos de la familia YOLO poseen una versatilidad y dominio que trasciende el ámbito cotidiano, principalmente debido a que han sido implementados con éxito en marcos sucesivos para la interpretación de tumores cerebrales [41], detección de objetos en fotografías aéreas tomadas por drones [42], e incluso en la detección de fauna peligrosa donde se puede llegar a encontrar serpientes en entornos complejos [43]. Esta capacidad de los modelos para operar bajo condiciones atípicas y detectar objetos diminutos demuestra una madurez algorítmica extrapolable a la identificación de obstáculos con gran precisión en el entorno de una persona con discapacidad visual.

En particular, los resultados evidencian que las variantes de YOLO como la versión 9, YOLO11, YOLO26 con SSD, han demostrado ser los más eficaces al proporcionar un equilibrio óptimo entre alta precisión de detección y velocidad de inferencia en tiempo real [29, 35].

En cuanto a los dispositivos utilizados como apoyo asistivo, el análisis de la literatura clasifica a las principales herramientas físicas en teléfonos inteligentes o smartphones y dispositivos vestibles (wearables en inglés) como gafas inteligentes o bastones ultrasónicos [29]. Donde los teléfonos inteligentes se destacan como dispositivos que operan para ser el medio principal en capturas visuales y retroalimentación auditiva, frecuentemente combinados con sistemas embebidos o módulos de comunicación IoT que permiten una comunicación inalámbrica a largas distancias con un bajo consumo de energía, haciendo posible un funcionamiento inclusive con baterías pequeñas durante varios años, esta convergencia permite un procesamiento eficiente en el borde, también llamado Edge Computing [21, 29].

Paralelamente, para optimizar el entendimiento del espacio, la literatura indica una transición hacia la computación espacial y Realidad Extendida (XR en inglés), ofreciendo interacciones naturales y multimodales con el objetivo de que enriquecer la percepción del entorno [44, 45]. Estos avances en la identificación espacial se apoyan en marcos robustos de estimación de posturas probados en entornos extremos como la navegación submarina, pero comparten las bases de la visión artificial para calcular la posición y orientación del entorno 3D para un usuario con discapacidad visual [46].

Estos hallazgos superan las expectativas de estudios anteriores donde se dependía de detectores de dos etapas que resultaban más lentos, o sensores de hardware voluminosos los cuales limitaban la portabilidad, consolidando así a la inteligencia artificial como una herramienta fundamental para la interpretación semántica y auditiva del entorno en dispositivos móviles [29, 38]. En un contexto más amplio, las implicaciones de estos resultados tienen un impacto profundo en la salud pública e inclusión social [47]. Considerando la alta carga demográfica entre ceguera y baja visión a nivel global, que a menudo conllevan al aislamiento social o inclusive accidentes, la adopción de tecnologías de asistencia integradas en dispositivos de uso cotidiano (como teléfonos inteligentes) fomenta una verdadera inclusión [20].

Las herramientas capaces de interpretar textos mediante OCR o reconocer billetes promueven la independencia financiera y movilidad de los usuarios [3]. Pero además de mejorar notablemente la calidad de vida y la autoconfianza, también disminuye la dependencia hacia los cuidadores, facilitando su desenvolvimiento en entornos complejos [39]. A pesar de los grandes avances, la interpretación de resultados captados en la literatura reciente también revela limitaciones significativas en estudios actuales, los sistemas evidencian que la confiabilidad de los modelos disminuye drásticamente en escenarios dinámicos que no pueden ser controlados en el mundo real, enfrentando problemas de falsos positivos frente a variaciones extremas de iluminación, fondos desordenados u oclusiones de objetos parciales [17].

Con base en las limitaciones observadas, se nota un enfoque en desarrollo de arquitecturas aún más ligeras y algoritmos de optimización que mejoren la eficiencia del procesamiento pero a la vez requiere de dispositivos sin recursos restringidos, al igual que se resalta una necesidad de implementar una validación cruzada con grandes

conjuntos de datos, diversos y anotados por las propias personas con discapacidad visual, con el objetivo para garantizar la generalización de los modelos frente a escenarios no vistos y así lograr reducir el sesgo [17, 48].

Además, es importante explorar a fondo la fusión de datos multimodales, combinando Visión Artificial con sensores ambientales (LiDAR o ultrasonido) y Modelos de Lenguaje Visual para lograr un razonamiento espacial interactivo mediante comandos de voz bidireccionales y adaptables [20, 29].

Finalmente, las próximas tecnologías asistivas deberán garantizar enfoques centrados en el usuario conservando la privacidad y ofreciendo interfaces innovadoras que reduzcan la carga cognitiva [48].

6. Agradecimientos

Este proyecto se puede realizar gracias a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI). Asimismo, se agradece al Instituto Tecnológico Superior de Xalapa por el apoyo académico para el desarrollo de este trabajo.

7. Referencias

- [1] Khadidos, A. O., Yafoz, A. (2025). An intelligent *object detection* and classification framework for assisting visually challenged persons using deep learning and improved crow search optimization. *Scientific Reports*, 15 (1), 1-28. <https://doi.org/10.1038/s41598-025-15793-0>
- [2] Jaramillo Cerezo, A., Torres Yepes, V., Franco Sánchez, I., Llano Naranjo, Y., Arias Uribe, J., Suárez Escudero, J. C. (2021). Etiología y consideraciones en salud de la discapacidad visual en la primera infancia: Revisión del tema. *Revista Mexicana de Oftalmología*, 96 (1S), 27-36. <https://hdl.handle.net/20.500.14330/PER01000452631>
- [3] Bolaños-Fernández, C., Bacca-Cortes, E. B. (2024). Mobile Application for Recognizing Colombian Currency with Audio Feedback for Visually Impaired People. *Ingeniería*, 29 (2), 1-25. <https://doi.org/10.14483/23448393.21408>
- [4] Stahl, A. (2020). The Diagnosis and Treatment of Age-Related Macular Degeneration. *Deutsches Ärzteblatt international*, 117, 513-520. <https://doi.org/10.3238/arztebl.2020.0513>
- [5] Ferrari, A. J., Santomauro, D. F., Aali, A., Abate, Y. H., Abbafati, C., Abbastabar, H., Abd ElHafeez, S., Abdelmasseh, M., Abd-Elsalam, S., Abdollahi, A., Abdullahi, A., Abegaz, K. H., Abeldaño Zuñiga, R. A., Aboagye, R. G., Abolhassani, H., Abreu, L. G., Abualruz, H., Abu-Gharbieh, E., Abu-Rmeileh, N. M., ... Murray, C. J. L. (2024). Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: A systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*, 403 (10440), 2133-2161. [https://doi.org/10.1016/S0140-6736\(24\)00757-8](https://doi.org/10.1016/S0140-6736(24)00757-8)
- [6] Alzahrani, N., Al-Baity, H. H. (2023). *Object Recognition* System for the Visually Impaired: A Deep Learning Approach using Arabic Annotation. *Electronics*, 12 (3), 541-557. <https://doi.org/10.3390/electronics12030541>
- [7] Iqbal, S., N. Qureshi, A., Li, J., Mahmood, T. (2023). On the Analyses of Medical Images Using Traditional *Machine Learning* Techniques and Convolutional Neural Networks. *Archives of Computational Methods in Engineering*, 30 (5), 3173-3233. <https://doi.org/10.1007/s11831-023-09899-9>
- [8] Sharma, P., Takahashi, N., Ninomiya, T., Sato, M., Miya, T., Tsuda, S., Nakazawa, T. (2025). A hybrid multi model *artificial intelligence* approach for glaucoma screening using fundus images. *Npj Digital Medicine*, 8 (1), 1-20. <https://doi.org/10.1038/s41746-025-01473-w>
- [9] Kansal, I., Khullar, V., Sharma, P., Singh, S., Hamid, J. A., Santhosh, A. J. (2025). Multiple model visual feature embedding and selection method for an efficient ocular disease classification. *Scientific Reports*, 15 (1), 1-24. <https://doi.org/10.1038/s41598-024-84922-y>
- [10] Cheng, H., Ning, X., Xu, B., Qin, Y., Li, C., Ling, R., Shen, Y., Jia, W., Zhong, J. (2026). Research on deep learning-based lesion identification in optical coherence tomography. *BMC Ophthalmology*, 26 (1), 1-9. <https://doi.org/10.1186/s12886-025-04579-7>
- [11] Chavan, C., Hembade, S., Jadhav, G., Komalwad, P., Rawat, P. (2023). *Computer Vision* Application Analysis based on *Object Detection*. *International Journal of Scientific Research in Engineering and Management*, 07 (04), 1-6. <https://doi.org/10.55041/IJSREM19015>

- [12]Kayadibi, İ., Güraksin, G. E. (2023). An Explainable Fully Dense Fusion Neural Network with Deep Support Vector Machine for Retinal Disease Determination. *International Journal of Computational Intelligence Systems*, 16 (1), 1-20. <https://doi.org/10.1007/s44196-023-00210-z>
- [13]Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S. (2020). *End-to-End Object Detection with Transformers*. *arXiv*. <https://doi.org/10.48550/arXiv.2005.12872>
- [14]Salih, N., Ksantini, M., Hussein, N., Ben Halima, D., Abdul Razzaq, A., Ahmed, S. (2023). Prediction of ROP Zones Using Deep Learning Algorithms and Voting Classifier Technique. *International Journal of Computational Intelligence Systems*, 16 (1), 1-11. <https://doi.org/10.1007/s44196-023-00268-9>
- [15]Alarood, A., Atoum, M. S., Manaf, A. A., Abubakar, A., Alsmadi, I. (2025). Enhanced obstacle detection using bilateral vision-aided transformer neural network for visually impaired persons. *Cluster Computing*, 28 (15), 1-23. <https://doi.org/10.1007/s10586-025-05740-z>
- [16]Pettersson, T., Riveiro, M., Löfström, T. (2024). Multimodal fine-grained grocery product recognition using image and OCR text. *Machine Vision and Applications*, 35 (4), 1-20. <https://doi.org/10.1007/s00138-024-01549-9>
- [17]Adam, M., Aljehane, N., Alzahrani, M., Al Zanin, S. (2025). Leveraging assistive technology for visually impaired people through optimal deep transfer learning based *object detection* model. *Scientific Reports*, 15 (1), 1-17. <https://doi.org/10.1038/s41598-025-14946-5>
- [18]Gao, Q., Manduchi, R., Ramulu, P. Y., Legge, G. E., Xiong, Y. (2025). VI-OCR: “Visually Impaired” optical character recognition pipeline for text accessibility assessment. *Scientific Reports*, 16 (1), 1-16. <https://doi.org/10.1038/s41598-025-30982-7>
- [19]Fernando, S., Ndukwe, C., Virdee, B., Djemai, R. (2025). Image Recognition Tools for Blind and Visually Impaired Users: An Emphasis on the Design Considerations. *ACM Transactions on Accessible Computing*, 18 (1), 1-21. <https://doi.org/10.1145/3702208>
- [20]Al Duhayyim, M. (2025). Ensemble of deep learning and IoT technologies for improved safety in smart indoor activity monitoring for visually impaired individuals. *Scientific Reports*, 15 (1), 1-15. <https://doi.org/10.1038/s41598-025-09716-2>
- [21]Salinas Buestán, N. R., Miranda Briones, E. T., Torres Quijije, Á. I., Intriago Rodríguez, D. F., Peña Banegas, D. P. (2024). Implementación de un Dispositivo Inteligente para la Asistencia de Personas con Discapacidad Visual en Entornos Universitarios. *Revista Tecnológica - ESPOL*, 36 (E1) 131-145. <https://doi.org/10.37815/RTE.V36NE1.1196>
- [22]Kitchenham, B., Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering* (EBSE-2007-01). Keele University & Durham University Joint Report. https://legacyfileshare.elsevier.com/promis_misc/525444systematicreviewsguide.pdf
- [23]Jiménez Villa, L. A., Alejandro Apolinar, M. S., Lagunes Barradas, V., Hidalgo Reyes, M. Á. (2026). Revisión de la literatura para el conteo de unidades formadoras de colonias en microorganismos mediante visión artificial. *Revista InGenio*, 9 (1), 18-31. <https://doi.org/10.18779/ingenio.v9i1.1117>
- [24]Vieira Da Silva, K., Kafure Munoz, I., Neves Raposo, P. (2023). Terminology applied in the study of user information when aimed at visually impaired people. *Revista General de Información y Documentación*, 33 (1), 203-218. <https://doi.org/10.5209/rgid.83835>
- [25]Ashiq, F., Asif, M., Ahmad, M. B., Zafar, S., Masood, K., Mahmood, T., Mahmood, M. T., Lee, I. H. (2022). CNN-Based *Object Recognition* and Tracking System to Assist Visually Impaired People. *IEEE Access*, 10, 14819-14834. <https://doi.org/10.1109/ACCESS.2022.3148036>
- [26]Qureshi, T. A., Rajbhar, M., Pisat, Y., Bhosale, V. (2021). *AI Based App for Blind People*. 08(03). https://www.academia.edu/51055393/IRJET_AI_Based_App_for_Blind_People
- [27]Alhazmi, S., Kutbi, M., Alhelaly, S., Dawood, U., Felemban, R., Alaslani, E. (2022). Utilizing *Artificial Intelligence* Techniques for Assisting Visually Impaired People: A Personal AI-based Assistive Application. *International Journal of Advanced Computer Science and Applications*, 13 (8), 813-820. <https://doi.org/10.14569/IJACSA.2022.0130894>
- [28]Islam, R. B., Akhter, S., Iqbal, F., Rahman, M. S. U., Khan, R. (2023). Deep learning based *object detection* and surrounding environment description for visually impaired people. *Heliyon*, 9 (6), 1-19. <https://doi.org/10.1016/j.heliyon.2023.e16924>

- [29] Kumari, P., Hammady, R. (2026). Assisting blind people with AI and audio using smart glasses: System design with YOLOv8 variants comparisons. *Multimedia Systems*, 32 (1), 1-26. <https://doi.org/10.1007/s00530-025-02139-z>
- [30] Bala, M. M., Vasundhara, D. N., Haritha, A., Moorthy, C. V. K. N. S. N. (2023). Design, development and performance analysis of cognitive assisting aid with multi sensor fused navigation for visually impaired people. *Journal of Big Data*, 10 (1), 1-21. <https://doi.org/10.1186/s40537-023-00689-5>
- [31] Dragne, C., Todirițe, I., Iliescu, M., Pandealea, M. (2022). Distance Assessment by *Object Detection*—For Visually Impaired Assistive Mechatronic System. *Applied Sciences*, 12 (13), 1-20. <https://doi.org/10.3390/app12136342>
- [32] Alohalı, M. A., Alanazi, F., Alsahafi, Y. A., Yaseen, I. (2025). Intelligent feature fusion with dynamic graph convolutional recurrent network for robust *object detection* to assist individuals with disabilities in a smart Iot edge-cloud environment. *Scientific Reports*, 15 (1), 1-20. <https://doi.org/10.1038/s41598-025-25048-7>
- [33] Harum, N. B., M. S. K, N. I., Emran, N. A., Abdullah, N., Zakaria, N. A., Hamid, E., Anawar, S. (2021). A Development of Multi-Language Interactive Device using *Artificial Intelligence* Technology for *Visual Impairment* Person. *International Journal of Interactive Mobile Technologies (iJIM)*, 15 (19), 79-92. <https://doi.org/10.3991/ijim.v15i19.24139>
- [34] Ghanem, A. M., Youness, H. A., Wahba, M., Abdelaal, H. M. (2025). Recognizing Egyptian currency for people with *visual impairment* using deep learning models. *Scientific Reports*, 15 (1), 1-15. <https://doi.org/10.1038/s41598-025-20646-x>
- [35] Naz, S., Jabeen, F. (2024). Towards Improved *Assistive Technologies*: Classification and Evaluation of *Object Detection* Techniques for Users with *Visual Impairments*. *VAWKUM Transactions on Computer Sciences*, 12 (2), 165-177. <https://doi.org/10.21015/vtcs.v12i2.1911>
- [36] Sapkota, R., Karkee, M. (2026). YOLOE-26: Integrating YOLO26 with YOLOE for Real-Time Open-Vocabulary Instance Segmentation. *arXiv*. <https://doi.org/10.48550/arXiv.2602.00168>
- [37] Nasir, H. M., Brahın, N. M. A., Aminuddin, M. M. M., Mispan, M. S., Zulkifli, M. F. (2021). Android based application for visually impaired using deep learning approach. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 10 (4), 879-888. <https://doi.org/10.11591/ijai.v10.i4.pp879-888>
- [38] Khadidos, A. O., Yafoz, A. (2025). Leveraging retinanet based *object detection* model for assisting visually impaired individuals with metaheuristic optimization algorithm. *Scientific Reports*, 15 (1), 1-19. <https://doi.org/10.1038/s41598-025-99903-y>
- [39] Alashjaee, A. M., Alhashmi, A. A., Darem, A. A. (2025). A smart assistive system for visually challenged people through efficient *object detection* using deep learning with tunicate swarm algorithm. *Scientific Reports*, 15 (1), 1-17. <https://doi.org/10.1038/s41598-025-29947-7>
- [40] Srinivasan, S., Francis, D., Mathivanan, S. K., Rajadurai, H., Shivahare, B. D., Shah, M. A. (2024). A hybrid deep CNN model for brain tumor image multi-classification. *BMC Medical Imaging*, 24 (1), 1-21. <https://doi.org/10.1186/s12880-024-01195-7>
- [41] Priyadharshini, S., Bhoopalan, R., Manikandan, D., Ramaswamy, K. (2025). A successive framework for brain tumor interpretation using Yolo variants. *Scientific Reports*, 15 (1), 1-24. <https://doi.org/10.1038/s41598-025-13155-4>
- [42] Wang, G., Chen, Y., An, P., Hong, H., Hu, J., Huang, T. (2023). UAV-YOLOv8: A Small-Object-Detection Model Based on Improved YOLOv8 for UAV Aerial Photography Scenarios. *Sensors*, 23 (16), 1-27. <https://doi.org/10.3390/s23167190>
- [43] Naresh, E., Babu, J. A., Darshan, S. L. S., Murthy, S. V. N., Srinidhi, N. N. (2023). A Novel Framework for Detection of Harmful Snakes Using YOLO Algorithm. *SN Computer Science*, 5 (1), 1-10. <https://doi.org/10.1007/s42979-023-02366-z>
- [44] Wang, Z.-M., Rao, M.-H., Ye, S.-H., Song, W.-T., Lu, F. (2025). Towards spatial computing: Recent advances in multimodal natural interaction for Extended Reality headsets. *Frontiers of Computer Science*, 19 (12), 1-22. <https://doi.org/10.1007/s11704-025-41123-8>
- [45] Joshi, R. C., Yadav, S., Dutta, M. K., Travieso-Gonzalez, C. M. (2020). Efficient Multi-Object Detection and Smart Navigation Using *Artificial Intelligence* for Visually Impaired People. *Entropy*, 22 (9), 1-17. <https://doi.org/10.3390/e22090941>

- [46]Wu, Y., Li, Y., Shi, B. (2025). A Switching Framework for Robust Underwater Pose Estimation: Integrating IPNet with Vision-based SLAM. *Journal of Intelligent & Robotic Systems*, 111 (3), 1-14.
<https://doi.org/10.1007/s10846-025-02300-w>
- [47]Madueña-Angulo, S. E., Beltran-Ontiveros, S. A., Leal-Leon, E., Contreras-Gutierrez, J. A., Lizarraga-Verdugo, E., Gutierrez-Arzapalo, P. Y., Lizarraga-Velarde, S., Romo-Garcia, E., Montero-Vela, J., Moreno-Ortiz, J. M., Garcia-Magallanes, N., Cuen-Diaz, H. M., Magaña-Gomez, J., Velazquez, D. Z., Hernandez-Carreño, P. E., Jimenez-Trejo, F., Reyes, M., Muñoz, F. P., Diaz, D. (2023). National sex- and age-specific burden of *blindness* and vision impairment by cause in Mexico in 2019: A secondary analysis of the Global Burden of Disease Study 2019. *The Lancet Regional Health - Americas*, 24, 1-12.
<https://doi.org/10.1016/j.lana.2023.100552>
- [48]Han, Z., Liu, X., Hao, J. (2025). Lightweight visual accessibility LLaVA architecture. *Scientific Reports*, 15 (1), 1-12. <https://doi.org/10.1038/s41598-025-23023-w>