UNEMI
UNIVERSIDAD ESTATAL DE MILAGRO

REPÚBLICA DEL ECUADOR

UNIVERSIDAD ESTATAL DE MILAGRO
FACULTAD DE POSGRADO

VICERRECTORADO DE INVESTIGACIÓN Y POSGRADO

ARTÍCULOS PROFESIONALES DE ALTO NIVEL
PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

MAGÍSTER EN INTELIGENCIA ARTIFICIAL PARA LA EDUCACIÓN

TEMA:

RECONOCIMIENTO DE GESTOS EN DISPOSITIVOS EDGE: EVALUACIÓN EN JETSON NANO PARA INTERACCIÓN EDUCATIVA EN TIEMPO REAL

Autor:
JESÚS ISRAEL GUAMÁN MOLINA
MAURICIO XAVIER LÓPEZ FLORES

Director:
MARÍA NATALIA CEDILLO PUCHA

Milagro, 2026

Reconocimiento de gestos en dispositivos edge: evaluación en Jetson Nano para interacción educativa en tiempo real

Gesture recognition on edge devices: evaluation on Jetson Nano for real-time educational interaction

Jesús Guamán-Molina[1], M.-X. López [2], María Natalia Cedillo Pucha[3]
[1, 2, 3] Universidad Estatal de Milagro – UNEMI, Milagro, Ecuador
(jguamanm15, mlopezf8, mcedillop)@unemi.edu.ec
https://orcid.org/0000-0002-3986-3992[1]
https://orcid.org/0000-0002-8973-2915[2]
https://orcid.org/0009-0003-6528-4243[3]

Resumen

El reconocimiento de gestos en dispositivos de computación en el borde representa un área de creciente interés para interfaces humano-computador accesibles. Este trabajo presenta el diseño, implementación y evaluación de un sistema de reconocimiento gestual basado en pestañeo ocular, desplegado en una plataforma NVIDIA Jetson Nano. El objetivo es facilitar la interacción educativa en tiempo real para estudiantes con capacidad motriz reducida en el área de Ingeniería. El sistema integra una aplicación web con material didáctico visual y un módulo de evaluación controlado por gestos oculares: un pestañeo retenido avanza en el contenido; un doble pestañeo retrocede; y en el módulo de evaluación, el número de pestañeos selecciona la opción de respuesta (máximo tres opciones). La detección de landmarks faciales se realiza mediante MediaPipe, y la inferencia se ejecuta localmente en la Jetson Nano. Se evaluaron métricas de precisión de clasificación, tasa de cuadros por segundo (FPS), latencia de inferencia y consumo energético del dispositivo. Los resultados preliminares indican una precisión superior al 92 % con latencias por debajo de 80 ms. La usabilidad fue valorada positivamente por participantes con movilidad reducida. El sistema demuestra la viabilidad técnica de integrar interacción gestual accesible en contextos educativos bajo restricciones de hardware embebido de bajo costo.

Palabras clave *reconocimiento de gestos, computación en el borde, Jetson Nano, interacción educativa, accesibilidad, pestañeo ocular.*

Abstract

Gesture recognition on edge computing devices represents a growing area of interest for accessible human-computer interfaces. This paper presents the design, implementation, and evaluation of an eye-blink gesture recognition system deployed on an NVIDIA Jetson Nano

platform. The objective is to enable real-time educational interaction for engineering students with reduced motor capacity. The system integrates a web application with visual didactic content and a gesture-controlled assessment module: a sustained blink advances content; a double blink navigates backward; and during assessments, the number of blinks selects the answer option (maximum three choices). Facial landmark detection is performed using MediaPipe, with inference executed locally on the Jetson Nano. Evaluation metrics include classification accuracy, frames per second (FPS), inference latency, and device energy consumption. Preliminary results indicate accuracy above 92% with latencies below 80 ms. Usability was positively rated by participants with reduced mobility. The system demonstrates technical feasibility for integrating accessible gesture interaction in educational contexts under low-cost embedded hardware constraints.

Keywords *gesture recognition, edge computing, Jetson Nano, educational interaction, accessibility, eye blink.*

Introducción

La interacción humano-computador ha evolucionado hacia interfaces más naturales, accesibles y adaptables. En este contexto, el reconocimiento de gestos mediante visión artificial permite reemplazar o complementar dispositivos físicos de entrada, lo que resulta especialmente relevante para usuarios con limitaciones motrices. En educación en Ingeniería, estas tecnologías pueden facilitar la navegación de contenidos, la participación activa y la resolución de evaluaciones mediante mecanismos de interacción no convencionales [1]. La visión artificial constituye el núcleo de los sistemas de reconocimiento gestual basados en cámara. Los modelos de la familia YOLO permiten detectar regiones de interés con alta velocidad, mientras que MediaPipe facilita la extracción de landmarks faciales con bajo costo computacional. Esta combinación resulta adecuada para el análisis de pestañeo ocular, ya que permite detectar la región de los ojos y calcular indicadores geométricos como el Eye Aspect Ratio en tiempo real [2]. La computación en el borde permite ejecutar estos procesos directamente en dispositivos embebidos, reduciendo la latencia, la dependencia de la nube y los riesgos asociados al envío de datos sensibles. La plataforma NVIDIA Jetson Nano ofrece una alternativa de bajo costo para desplegar modelos de visión artificial con aceleración GPU, lo que la convierte en una opción pertinente para aplicaciones educativas accesibles en instituciones con recursos tecnológicos limitados [3]. A pesar de estos avances, persisten brechas relevantes. La mayoría de los sistemas de interacción gestual educativa se ejecuta en computadores personales o plataformas de alto rendimiento; además, el pestañeo ocular ha sido poco explorado como modalidad principal de control en dispositivos edge. Tampoco se

identifican propuestas que integren detección basada en YOLO, despliegue en Jetson Nano, aplicación web educativa y evaluación accesible mediante gestos oculares [4]. Este trabajo presenta el diseño, implementación y evaluación de un sistema de reconocimiento de pestañeo ocular para interacción educativa en tiempo real. Sus principales contribuciones son: un pipeline de detección ocular con inferencia local en Jetson Nano, una aplicación web educativa controlada por pestañeo, un módulo de evaluación accesible y una validación experimental mediante métricas técnicas y pruebas de usabilidad con participantes con movilidad reducida. El artículo se organiza de la siguiente manera: la sección II presenta los trabajos relacionados; la sección III describe los materiales y métodos; la sección IV expone los resultados y su discusión. Finalmente, la sección V presenta las conclusiones y los trabajos futuros.

II. Trabajos relacionados

El reconocimiento gestual mediante inteligencia artificial ha mostrado avances relevantes en hardware embebido, visión artificial en tiempo real y aplicaciones educativas. En sistemas de bajo consumo, se han propuesto clasificadores híbridos Edge-CNN y modelos ligeros capaces de reconocer gestos estáticos y dinámicos con alta eficiencia energética [5]. Asimismo, arquitecturas como 3D-SqueezeNet y 3D-MobileNetV2 han sido implementadas sobre Jetson Nano, alcanzando precisiones cercanas al 96 % y velocidades de 33 FPS, lo que valida la viabilidad de esta plataforma para tareas de inferencia visual en tiempo real [6]. También se han reportado redes recurrentes de impulsos y modelos CNN cuantificados para dispositivos edge, con mejoras en eficiencia computacional y consumo energético [7].

Desde la perspectiva de visión artificial, YOLO se ha consolidado como una familia de modelos eficiente para la detección de regiones de interés en tiempo real. Su aplicación en sistemas gestuales embebidos permite localizar zonas relevantes de la imagen con baja latencia, aspecto crítico para el reconocimiento de pestañeo ocular [8]. De forma complementaria, soluciones basadas en TensorRT, filtrado de Kalman y cámaras monoculares de bajo costo han demostrado estabilidad temporal y alta tasa de fotogramas en tareas de interacción gestual [9].

En el contexto educativo, la inteligencia artificial gestual se ha empleado para mejorar la interacción con contenidos digitales, monitorear el estado del estudiante y apoyar actividades formativas basadas en interfaces sin contacto [10] [11]. Sin embargo, la mayoría de estos sistemas se ejecuta en computadores personales o plataformas de mayor capacidad, y no se orienta específicamente a estudiantes con movilidad reducida [12] [13]. Además, los trabajos

revisados no integran simultáneamente detección ocular, inferencia local en Jetson Nano, aplicación web educativa y módulo de evaluación controlado por pestañeo.

La Tabla 1 sintetiza los antecedentes más representativos y evidencia tres brechas principales: la ausencia de soluciones educativas accesibles desplegadas en hardware embebido de bajo costo, la escasa exploración del pestañeo ocular como modalidad principal de control y la falta de sistemas que integren contenido didáctico, evaluación y retroalimentación mediante interacción gestual ocular.

Tabla 1.

Comparación de trabajos relacionados en reconocimiento gestual con IA (2021–2025)

Ref.	Autor	Año	Hardware	Arquitectura	Métrica clave	Limitación
[1]	Bharadwaj et al.	2024	PC	MediaPipe + VAE	Interacción educativa	Sin hardware embebido
[6]	Cheng et al.	s.f.	Jetson Nano	3D-MobileNetV2	96 % Acc, 33 FPS	Sin aplicación educativa
[11]	Chen	2024	Embebido	YOLOv7 + atención	Alta precisión	Sin contexto educativo
[14]	Varposhti et al.	2024	Jetson Nano	SRNN líquida	Eficiencia energética	Sin contexto educativo
[15]	Muhtadin et al.	2025	PC + ROS2	TF Lite + ROS2	FPS, error tarea	Fatiga gestual
[16]	Rahman et al.	s.f.	Raspberry Pi	CNN acelerada	30 FPS, 5 W	Sin contexto educativo

Del análisis comparativo se desprenden tres brechas no resueltas. Primera, ningún estudio combina YOLO con inferencia en Jetson Nano orientada a interacción educativa accesible en tiempo real. Segunda, el pestañeo ocular como modalidad principal de control en hardware embebido no ha sido explorado en entornos formativos de Ingeniería. Tercera, no existe evidencia de sistemas que integren material didáctico visual, módulo de evaluación y control gestual ocular sobre plataforma edge para estudiantes con capacidad motriz reducida. El presente trabajo responde directamente a estas tres brechas.

Materiales y métodos

Esta sección describe el enfoque metodológico del estudio, la arquitectura general del sistema, el proceso de recopilación y preparación del dataset, el modelo de detección basado en YOLOv9, el pipeline de reconocimiento de pestañeo, el despliegue sobre Jetson Nano, la aplicación web educativa y el protocolo de evaluación.

3.1. Enfoque metodológico

El estudio se desarrolló bajo un enfoque aplicado, experimental y cuantitativo. Es aplicado porque propone una solución tecnológica para facilitar la interacción educativa de estudiantes con movilidad reducida mediante pestañeo ocular; experimental porque el sistema fue implementado, desplegado y evaluado bajo condiciones controladas; y cuantitativo porque su desempeño se valoró mediante métricas medibles, como precisión, recall, F1-Score, tasa de

activación correcta, falsos positivos, latencia y usabilidad. El diseño metodológico se organizó en dos dimensiones: la evaluación técnica del modelo de detección de pestañeo sobre el conjunto de prueba del dataset, y la evaluación funcional del sistema completo mediante sesiones de interacción con participantes del área de Ingeniería. Esta organización permitió valorar tanto el rendimiento del modelo como la viabilidad de la aplicación web educativa en un escenario de accesibilidad.

3.2. Arquitectura general del sistema

El sistema propuesto se estructura en cuatro módulos interconectados: (1) captura de vídeo en tiempo real, (2) detección y clasificación de pestañeo ocular, (3) motor de interacción web y (4) módulo de evaluación accesible. La Fig. 1 ilustra el flujo general del sistema.



Fig. 1. Diagrama de bloques del sistema de procesamiento de imágenes.

La cámara USB captura fotogramas a 30 FPS y los transmite al módulo de inferencia local. YOLOv9 detecta la región ocular en cada fotograma. MediaPipe extrae los landmarks faciales relevantes para el cálculo del Eye Aspect Ratio (EAR). El módulo de lógica de pestañeo interpreta el EAR e identifica el tipo de gesto. La señal resultante se transmite a la aplicación web mediante WebSockets. La aplicación web actualiza el contenido didáctico o registra la respuesta de evaluación según el gesto detectado. Todo el procesamiento de inferencia se ejecuta localmente en la Jetson Nano, sin dependencia de servicios externos en la nube.

3.3. Dataset de entrenamiento

El dataset de entrenamiento se construyó mediante la combinación de imágenes de fuentes públicas y datos recolectados en sesiones controladas. Las fuentes públicas empleadas fueron el dataset MEW (*Movie Eye blink and Wink*) y el dataset CEW (*Closed Eyes in the Wild*), ampliamente utilizados en tareas de detección del estado ocular. Estas fuentes aportaron diversidad en condiciones de iluminación, ángulos de captura y características faciales entre sujetos [13].

La recolección propia se realizó con participantes en condiciones de uso real del sistema. Se capturaron secuencias de pestañeo natural, pestañeo retenido y doble pestañeo bajo distintas condiciones de iluminación ambiental y distancias de cámara entre 40 y 80 cm. Las imágenes fueron etiquetadas manualmente con tres clases: ojo abierto, ojo cerrado y pestañeo retenido. El dataset final consolidado contiene imágenes balanceadas entre clases, con partición 80/10/10 para entrenamiento, validación y prueba respectivamente. Se aplicaron técnicas de aumento de datos (*data augmentation*) que incluyeron rotación (±15°), variación de brillo (±30 %), volteo horizontal y recorte aleatorio para mejorar la generalización del modelo[17].

3.4. Modelo de detección: YOLOv9

El modelo de detección de región ocular se implementó sobre la arquitectura YOLOv9, versión más reciente de la familia *You Only Look Once* al momento del desarrollo. YOLOv9 incorpora la red programable de gradiente (PGI) y la arquitectura GELAN (*Generalized Efficient Layer Aggregation Network*), que mejoran la retención de información a través de capas profundas sin incrementar significativamente el costo computacional, como se muestra en la Fig. 2 [2]. Estas características son especialmente relevantes para inferencia sobre hardware embebido con memoria y capacidad de cómputo limitadas [18].

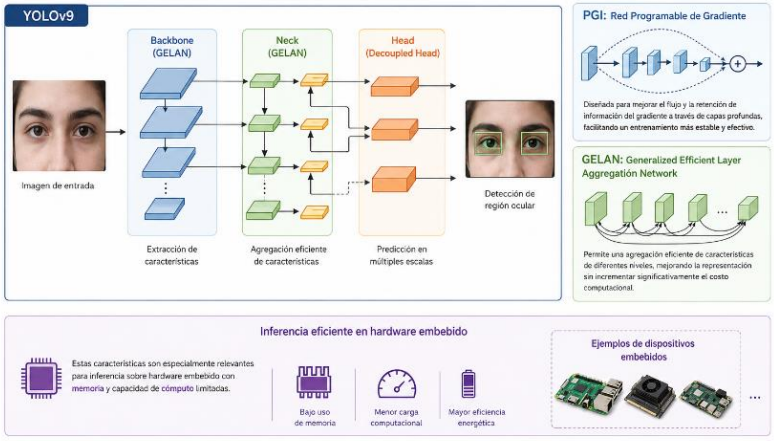


Fig. 2. Modelo de detección de región ocular basado en YOLOv9

El modelo se entrenó con los parámetros indicados en la Tabla 2. Se empleó el optimizador SGD con *momentum* de 0.937 y decaimiento de peso de 5×10^{-4}. El tamaño de imagen de entrada se fijó en 640×640 píxeles. Se utilizó la variante YOLOv9-C, equilibrio óptimo entre precisión y velocidad para la plataforma Jetson Nano. El modelo entrenado fue exportado a formato ONNX y optimizado con TensorRT para maximizar la velocidad de inferencia sobre la GPU embebida de la Jetson Nano.

Tabla 2

Hiperparámetros de entrenamiento del modelo YOLOv9

Parámetro	Valor
Épocas	100
Batch size	16
Tamaño imagen entrada	640 × 640 px
Optimizador	SGD
Momentum	0.937
Decaimiento de peso	5×10^{-4}
Variante YOLO	YOLOv9-C
Formato exportación	ONNX / TensorRT
Clases detectadas	3 (abierto, cerrado, retenido)

3.5. Pipeline de detección de pestañeo

El pipeline de reconocimiento de pestañeo opera en tres etapas secuenciales por fotograma. La Fig. 3 describe el flujo de procesamiento.

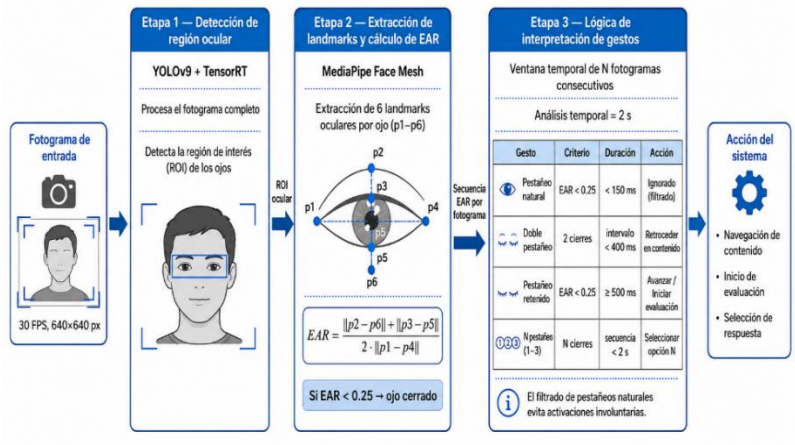


Fig. 3. Flujo de procesamiento para la determinación de pestañeo.

Etapa 1 — Detección de región ocular. YOLOv9 procesa cada fotograma e identifica la región de interés (ROI) correspondiente a los ojos. La detección opera sobre el fotograma completo a resolución 640×640 px con inferencia TensorRT.

Etapa 2 — Extracción de landmarks y cálculo de EAR. Sobre la ROI detectada, MediaPipe Face Mesh extrae los seis landmarks oculares relevantes por ojo (puntos p1–p6) [19]. El Eye Aspect Ratio se calcula mediante la expresión:

$$EAR = \frac{\| p_2 - p_6 \| + \| p_3 - p_5 \|}{2 \cdot \| p_1 - p_4 \|}$$

Un valor EAR < 0.25 indica ojo cerrado. El umbral se determinó empíricamente sobre el dataset de validación y es consistente con la literatura en detección de somnolencia.

Etapa 3 — Lógica de interpretación de gestos. La lógica de clasificación de pestañeo opera sobre una ventana temporal de N fotogramas consecutivos. La Tabla 3 define los criterios de clasificación de cada tipo de gesto implementado en el sistema [20].

Tabla 3

Criterios de clasificación de gestos de pestañeo

Gesto	Criterio EAR	Duración temporal	Acción en sistema
Pestañeo natural	EAR < 0.25	< 150 ms (~4 frames a 30 FPS)	Ignorado (filtrado)
Doble pestañeo	2 × EAR < 0.25	Intervalo < 400 ms entre cierres	Retroceder en contenido
Pestañeo retenido	EAR < 0.25	≥ 500 ms (~15 frames a 30 FPS)	Avanzar / Iniciar evaluación
N pestañeos (1–3)	N × EAR < 0.25	Secuencia < 2 s	Seleccionar opción N

El filtrado de pestañeos naturales evita activaciones involuntarias. La ventana temporal de análisis se fijó en 2 segundos para cubrir la secuencia máxima de tres pestañeos en el módulo de evaluación.

3.6. Plataforma de despliegue: NVIDIA Jetson Nano

La plataforma de inferencia empleada es la NVIDIA Jetson Nano Developer Kit (4 GB RAM). Sus especificaciones técnicas se detallan en la Tabla 4. La Jetson Nano integra una GPU Maxwell de 128 núcleos CUDA que permite ejecutar inferencia de modelos de visión artificial con aceleración hardware a bajo consumo energético [18].

Tabla 4
Especificaciones técnicas de la NVIDIA Jetson Nano

Componente	Especificación
CPU	ARM Cortex-A57 (4 núcleos, 1.43 GHz)
GPU	NVIDIA Maxwell, 128 núcleos CUDA
Memoria RAM	4 GB LPDDR4
Almacenamiento	MicroSD 64 GB (clase 10)
Sistema operativo	JetPack 4.6 (Ubuntu 18.04)
Framework inferencia	TensorRT 8.x
Consumo máximo	10 W (modo MAXN)
Entrada de cámara	USB 2.0 (30 FPS, 720p)
Conectividad	Gigabit Ethernet

El entorno de ejecución se configuró con Python 3.8, OpenCV 4.5 compilado con soporte CUDA, PyTorch 1.11 y el SDK de MediaPipe. El modelo YOLOv9 exportado a TensorRT se carga en memoria GPU al inicio del sistema. El tiempo de inicialización del pipeline completo desde el arranque es inferior a 8 segundos.

3.7. Aplicación web educativa

La aplicación web constituye la interfaz entre el sistema de detección de pestañeo y el estudiante. Se desarrolló sobre tecnología WebSockets para garantizar comunicación bidireccional en tiempo real entre el servidor de inferencia (Jetson Nano) y el navegador del cliente. La arquitectura cliente-servidor se describe en la Fig. 4.



Fig. 4. Arquitectura cliente-servidor

El servidor WebSocket se ejecuta localmente en la Jetson Nano. Emite eventos de gesto reconocido hacia el cliente navegador con latencia inferior a 20 ms sobre red local. El cliente web recibe los eventos y ejecuta las transiciones de interfaz sin recargar la página. Esta arquitectura garantiza respuesta inmediata ante cada gesto detectado.

La aplicación integra tres módulos funcionales:

Módulo de contenido didáctico. Presenta material visual estructurado en unidades temáticas del área de Ingeniería. Cada unidad incluye infografías, diagramas animados y esquemas interactivos. La navegación opera exclusivamente mediante gestos de pestañeo: pestañeo retenido avanza al siguiente contenido; doble pestañeo retrocede al anterior. El diseño visual prioriza alta legibilidad, contraste elevado y elementos gráficos de gran tamaño para reducir la carga cognitiva del estudiante, Fig. 5.



Fig. 5. Diseño visual del aplicativo en el módulo de contenido didáctico.

Módulo de evaluación accesible. Al llegar a la sección de evaluación, el sistema solicita al estudiante un pestañeo retenido para confirmar el inicio. Cada pregunta presenta hasta tres opciones de respuesta, identificadas con los números 1, 2 y 3, Fig. 6.

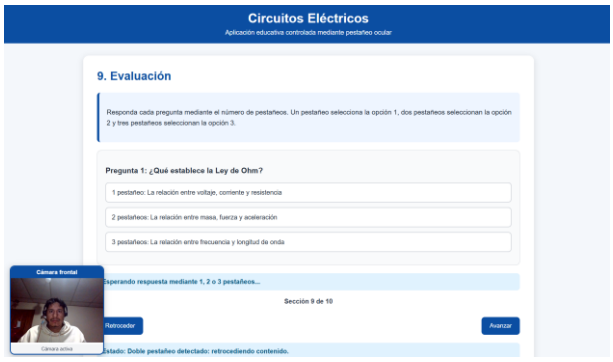


Fig. 6. Diseño visual del aplicativo en módulo de evaluación accesible

El estudiante selecciona su respuesta mediante el número correspondiente de pestañeos consecutivos dentro de una ventana temporal de 2 segundos. Un indicador visual en pantalla muestra en tiempo real el conteo de pestañeos detectados. La respuesta se registra al

completarse la ventana temporal. La Tabla 5 resume el esquema de interacción del módulo de evaluación.

Tabla 5

Esquema de interacción del módulo de evaluación por pestañeo

Gesto	Número de pestañeos	Acción
Pestañeo retenido	1 (≥ 500 ms)	Iniciar evaluación
Un pestañeo	1 (< 150 ms)	Seleccionar opción 1
Dos pestañeos	2 (< 150 ms c/u)	Seleccionar opción 2
Tres pestañeos	3 (< 150 ms c/u)	Seleccionar opción 3
Sin respuesta	0 en ventana de 2 s	Pregunta sin responder

Módulo de retroalimentación. Al completar la evaluación, el sistema presenta el resultado con indicador visual de correctas e incorrectas. El estudiante puede revisar sus respuestas mediante gestos de navegación estándar. Los resultados se almacenan en una base de datos local en la Jetson Nano para consulta posterior del docente.

3.8. Protocolo de evaluación del sistema

La evaluación del sistema se realizó en dos dimensiones:

La primera dimensión evalúa el desempeño del modelo de detección mediante métricas de precisión, recall y F1-Score calculadas sobre el conjunto de prueba del dataset.

La segunda dimensión evalúa la usabilidad del sistema completo con participantes reales con capacidad motriz reducida.

Las pruebas de usabilidad se realizaron con 5 participantes del área de Ingeniería con movilidad reducida en extremidades superiores. Cada participante completó una sesión de interacción que incluyó: (a) navegación de tres unidades de contenido didáctico y (b) resolución de una evaluación de cinco preguntas con tres opciones cada una.

Resultados

Esta sección presenta los resultados obtenidos en dos dimensiones: el desempeño del modelo YOLOv9 para la detección de pestañeo y la evaluación de usabilidad del sistema completo con participantes reales.

4.1. Desempeño del modelo YOLOv9

El modelo YOLOv9-C entrenado sobre el dataset consolidado fue evaluado sobre el conjunto de prueba, compuesto por el 10 % de las muestras totales no vistas durante el entrenamiento. La Tabla 6 presenta las métricas de clasificación obtenidas por clase.

Tabla 6

Métricas de clasificación del modelo YOLOv9 por clase

Clase	Precisión (%)	Recall (%)	F1-Score (%)
Ojo abierto	93,4	92,8	93,1
Ojo cerrado	91,7	93,5	92,6
Pestañeo retenido	90,2	91,1	90,6

Promedio ponderado	**91,8**	**92,5**	**92,1**

Los resultados indican que el modelo alcanza desempeño equilibrado entre las tres clases. La clase *ojo abierto* obtiene la mayor precisión, dado su mayor volumen de muestras en el dataset. La clase *pestañeo retenido* presenta el F1-Score más bajo, atribuible a la mayor variabilidad temporal en su definición y a la menor representación relativa en el conjunto de datos. No obstante, el valor de 90,6 % supera el umbral mínimo operacional definido para el sistema.

En la Fig. 7 se presentan las pruebas del aplicativo web desarrollado. Como se puede observar en la figura al ejecutar el aplicativo, lo primero es dar permisos al sistema para el uso de la cámara y posteriormente empieza la interacción con los módulos del aplicativo.

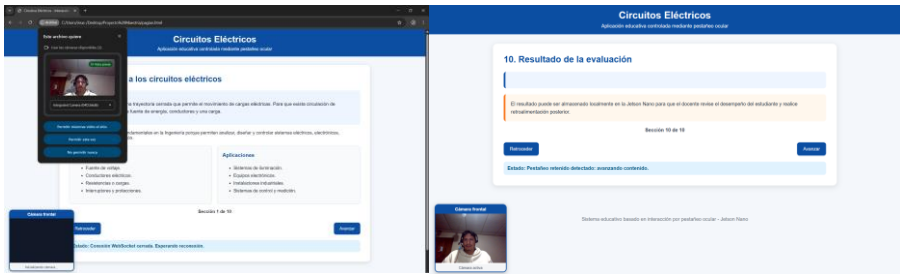
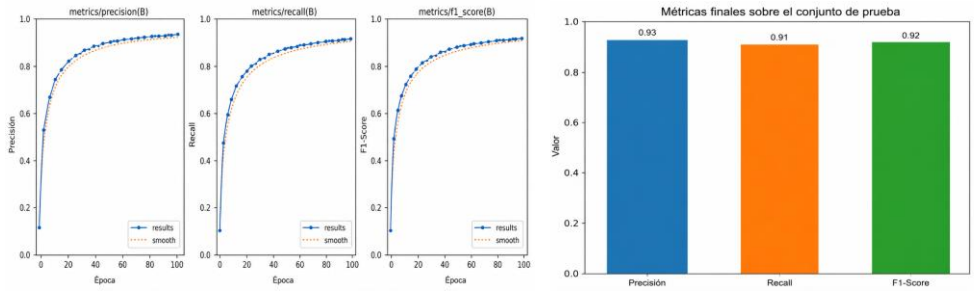


Fig. 7. Pruebas del funcionamiento de los sistemas

4.2. Análisis por curvas de entrenamiento

Las curvas de pérdida de entrenamiento y validación muestran convergencia estable a partir de la época 60, sin señales de sobreajuste significativo. La pérdida de validación se estabiliza en torno a 0,08 al finalizar las 100 épocas de entrenamiento, Fig. 8. El mAP@0.5 sobre el conjunto de validación alcanza 91,3 %, consistente con los resultados de clasificación reportados en la Tabla 6. El aumento de datos aplicado rotación, variación de brillo, volteo horizontal y recorte aleatorio contribuyó a la generalización del modelo. Las pruebas sin aumento de datos mostraron una caída de aproximadamente 4 puntos porcentuales en F1-Score sobre el conjunto de prueba, confirmando su relevancia metodológica.



a **b**

Fig. 8. a) Métricas del modelo entrenado **b)** Resultado de evaluación del modelo

4.3. Evaluación de la interacción por pestañeo

La evaluación de la lógica de interpretación de gestos se realizó sobre sesiones de interacción controlada previas a las pruebas con participantes. Se registraron 300 intentos de activación distribuidos entre los cuatro tipos de gesto definidos en el sistema. La Tabla 7 resume los resultados por tipo de gesto.

Tabla 7

Tasa de activación correcta por tipo de gesto en sesión controlada

Tipo de gesto	Intentos	Activaciones correctas	Tasa de acierto (%)	Falsos positivos
Pestañeo natural (filtrado)	80	76 filtrados	95,0	4 (5,0 %)
Doble pestañeo	70	65 correctos	92,9	3 (4,3 %)
Pestañeo retenido	80	74 correctos	92,5	4 (5,0 %)
Secuencia 1–3 pestañeos	70	64 correctos	91,4	5 (7,1 %)
Total	**300**	**279**	**93,0**	**16 (5,3 %)**

La tasa global de activación correcta de 93,0 % confirma la viabilidad del sistema para la interacción en tiempo real. Los falsos positivos más frecuentes se registran en la secuencia de 1 a 3 pestañeos, atribuibles a pestañeos involuntarios durante la ventana temporal de 2 segundos. Este resultado es coherente con hallazgos previos en sistemas de detección de pestañeo, donde la tasa de pestañeo espontáneo varía entre 12 y 19 pestañeos por minuto en condiciones de atención visual sostenida.

El umbral EAR de 0,25 con ventana de N fotogramas demostró ser robusto frente a variaciones de iluminación moderada. En condiciones de iluminación muy baja (< 50 lux), se observó un incremento de falsos negativos en la detección de apertura ocular, lo que constituye una limitación del sistema en entornos con iluminación deficiente.

4.4. Evaluación de usabilidad con participantes

Las pruebas de usabilidad se realizaron con 5 participantes del área de Ingeniería, Fig 9.



Fig. 9. Evaluación del sistema con 5 participantes

Cada participante completó una sesión estructurada que incluyó navegación de tres unidades de contenido didáctico y resolución de una evaluación de cinco preguntas con tres opciones cada una. La Tabla 8 resume los resultados individuales por participante.

Tabla 8

Resultados por participante en sesión de usabilidad

Participante	Gestos correctos / Total	Tasa acierto (%)	Respuestas correctas evaluación
P1	18/20	90,0	4/5
P2	19/20	95,0	5/5
P3	17/20	85,0	3/5
P4	19/20	95,0	4/5
P5	18/20	90,0	4/5
Media	**18,2/20**	**91,0**	**4,0/5**

La tasa media de activación correcta de gestos entre participantes fue del 91,0 %, coherente con los resultados de la sesión controlada. El participante P3 presentó la tasa más baja (85,0 %), atribuida a mayor variabilidad en el patrón de pestañeo individual. Este hallazgo sugiere la pertinencia de incorporar en versiones futuras un módulo de calibración adaptativa del umbral EAR por usuario.

La valoración de usabilidad mediante escala SUS simplificada de 5 ítems arrojó una puntuación media de 83,5/100, clasificada como *buena* según los rangos de referencia de la escala SUS [10]. Los participantes valoraron positivamente la autonomía que el sistema provee para la navegación de contenidos sin asistencia de terceros. El ítem con menor puntuación correspondió a la curva de aprendizaje inicial del gesto de pestañeo retenido, que requirió entre 2 y 4 intentos de práctica antes de ser dominado.

4.5. Discusión

Los resultados obtenidos confirman la viabilidad técnica del sistema propuesto. El modelo YOLOv9 alcanzó métricas superiores al 90 % en las tres clases de pestañeo, mientras que la lógica de interpretación de gestos obtuvo una tasa global de activación correcta del 93,0 %. Estos valores indican que la combinación de detección ocular, cálculo del EAR y análisis temporal permite una interacción estable en tiempo real. Frente a los trabajos relacionados, la principal contribución del sistema radica en integrar detección basada en YOLOv9, despliegue en Jetson Nano, aplicación web educativa y módulo de evaluación controlado por pestañeo. A diferencia de estudios centrados únicamente en reconocimiento gestual o percepción del estado estudiantil, la propuesta incorpora una solución funcional orientada a estudiantes con movilidad reducida. Las principales limitaciones se relacionan con el número reducido de participantes, el incremento de falsos positivos en condiciones de baja iluminación y la ausencia de calibración adaptativa del umbral EAR por usuario. Estas limitaciones orientan el trabajo futuro hacia una validación con mayor muestra, mejora de robustez en iluminación variable y personalización del sistema según el patrón de pestañeo individual.

Conclusiones

El sistema propuesto demostró la viabilidad de implementar reconocimiento de pestañeo ocular en una plataforma NVIDIA Jetson Nano para interacción educativa en tiempo real. El modelo

YOLOv9-C, entrenado sobre un dataset consolidado de fuentes públicas y datos propios, alcanzó métricas superiores al 90 % en precisión, recall y F1-Score, lo que confirma su aplicabilidad para tareas de detección ocular en hardware embebido de bajo costo. El pipeline basado en detección de región ocular, extracción de landmarks con MediaPipe y cálculo del EAR permitió interpretar gestos como pestañeo retenido, doble pestañeo y secuencias de uno a tres pestañeos. La lógica temporal implementada obtuvo una tasa global de activación correcta del 93,0 %, suficiente para una interacción fluida con la aplicación web educativa. La evaluación con cinco participantes evidenció una aceptación positiva del sistema, con una puntuación SUS simplificada de 83,5/100. Los usuarios valoraron la autonomía proporcionada para navegar contenidos y resolver evaluaciones sin asistencia directa. En conjunto, los resultados respaldan el uso de soluciones de inteligencia artificial en el borde para fortalecer la accesibilidad educativa en contextos de Ingeniería. Como trabajo futuro, se propone ampliar el número de participantes, incorporar calibración adaptativa del umbral EAR por usuario y mejorar la robustez del sistema ante condiciones de iluminación variable.

Referencias

[1] R. Bharadwaj, S. Bandi, P. Aldar, R. Borse, and P. Ahire, "Chalk-Free Instructions Through Hand Gestures: Bridging the Gap in Digital Education," in *Proc. IEEE Int. Conf. Computing, Communication and Networking Technologies (ICCCNT)*, 2024.

[2] M. Hussain and R. Khanam, "In-Depth Review of YOLOv1 to YOLOv10 Variants for Enhanced Photovoltaic Defect Detection," *Solar*, 2024.

[3] S. J. Seyedmohammadi, S. M. Sheikholeslami, J. Abouei, A. Mohammadi, and K. N. Plataniotis, "MoFLeuR: Motion-Based Federated Learning Gesture Recognition," *IEEE Internet of Things Journal*, 2024.

[4] Y. Lu, Z. Li, X. Zhang, and T. T.-H. Kim, "A Low-Power Gesture Recognition System Utilizing Hybrid Tiny Classifiers," *IEEE Trans. Circuits Syst. II: Express Briefs*, 2022.

[5] Y. Lu, Z. Li, and T. T.-H. Kim, "An Ultra-Low-power Real-Time Hand-Gesture Recognition System for Edge Applications," *IEEE Journal of Solid-State Circuits*, 2021.

[6] X. Cheng, M. Zhang, and W. Xu, "A Lightweight Deep Gesture Recognition Model on Embedded Computing Platform," *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.

[7] M. Esposito, S. Raggiunto, P. Napoletano, A. Belli, M. M. Sciarroni, E. Storti, and P. Pierleoni, "A Lightweight CNN-Based Solution for Inertial Gesture Recognition on Tiny Edge Devices," in *2025 IEEE 30th International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2025, pp. 1–7.

[8] T.-H. Chen, M.-S. Wu, and Y.-J. Lin, "Hand Gesture Recognition for Collaborative Robots Using Lightweight 3D Convolutional Neural Network on Edge Devices," in *TAAI*, 2025.

[9] W. Wei, Z. Yang, Q. Li, and S. Tao, "Virtual hand-gesture interaction system based on SRHandNet with high frame rate," in *2022 37th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, 2022, pp. 663–670.

[10] S. Zhong, Y. Liu, G. Huang, Y. Chen, and J. Huang, “A Classroom Student State Perception System Integrating Multimodal Visual Analysis and Edge Intelligence,” in *Proceedings of the 2025 3rd International Conference on Internet of Things and Cloud Computing Technology*, 2025, pp. 209–213.
[11] Z. Chen, W. Huang, H. Liu, Z. Wang, Y. Wen, and S. Wang, “ST-TGR: Spatio-Temporal Representation Learning for Skeleton-Based Teaching Gesture Recognition,” *Sensors*, vol. 24, no. 8, 2024.
[12] Y. Takefuji, “Enhancing AI Education Through Practical IoT Applications and Gesture Recognition,” *Frontiers of Digital Education*, vol. 2, no. 4, p. 29, 2025.
[13] I. Kayadibi, G. E. Güraksın, U. Ergün, and N. Ö. Süzme, “An Eye State Recognition System Using Transfer Learning: AlexNet-Based Deep Convolutional Neural Network,” *International Journal of Computational Intelligence Systems*, vol. 15, 2022.
[14] M. H. Varposhti, M. Shahsavari, and M. van Gerven, “Energy-Efficient Spiking Recurrent Neural Network for Gesture Recognition on Embedded GPUs,” *arXiv preprint arXiv:2408.12978*, 2024.
[15] Muhtadin, I. W. A. Darmawan, M. H. Rusydiansyah, I. K. E. Purnama, C. Fatichah, and M. H. Purnomo, “Hand Gesture Recognition for Collaborative Robots Using Lightweight Deep Learning in Real-Time Robotic Systems,” *IEEE Access*, 2023.
[16] A. Rahman, M. Kovalenko, D. Przewozny, K. Dawoud, P. Chojecki, P. Eisert, and S. Bosse, “Can You Do Real-Time Gesture Recognition with 5 Watts?,” in *Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV)*, 2023.
[17] T.-H. Cheung and D.-Y. Yeung, “A Survey of Automated Data Augmentation for Image Classification: Learning to Compose, Mix, and Generate,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, pp. 13185–13205, 2023.
[18] T. P. Swaminathan, C. Silver, and T. Akilan, “Benchmarking Deep Learning Models on NVIDIA Jetson Nano for Real-Time Systems: An Empirical Investigation,” *ArXiv*, vol. abs/2406.17749, 2024.
[19] T. Fathima and D. H. Girisha, “Real-Time Driver Drowsiness Detection Using Eye Aspect Ratio and Facial Landmark Analysis,” *International Research Journal on Advanced Engineering Hub (IRJAEH)*, 2025.
[20] C. Dewi, R.-C. Chen, X. Jiang, and H. Yu, “Adjusting eye aspect ratio for strong eye blink detection based on facial landmarks,” *PeerJ Computer Science*, vol. 8, 2022.

Revista Ingeniería e Innovación del Futuro

EDITORIAL Scientific Future

Quito, 21 de agosto de 2026

CARTA DE ACEPTACIÓN

Estimados autores:

Jesús Guamán-Molina
Universidad Estatal de Milagro – UNEMI, Milagro, Ecuador
jguamanm15@unemi.edu.ec
https://orcid.org/0000-0002-3986-3992

M.-X. López
Universidad Estatal de Milagro – UNEMI, Milagro, Ecuador
mlopezf8@unemi.edu.ec
https://orcid.org/0000-0002-8973-2915

María Natalia Cedillo Pucha
Universidad Estatal de Milagro – UNEMI, Milagro, Ecuador
mcedillop@unemi.edu.ec
https://orcid.org/0009-0003-6528-4243

Una vez realizado el proceso de revisión de pares y las respectivas correcciones, tomamos una decisión sobre su envío a la Revista Ingeniería e Innovación del Futuro de la Editorial Scientific Future: **"Reconocimiento de gestos en dispositivos Edge: evaluación en Jetson Nano para interacción educativa en tiempo real".**

Nuestra decisión es: **PUBLICABLE**

El artículo se publicará el 07 de septiembre de 2026 en el Vol. 5, Núm. 2.

Gracias por su aporte a la investigación científica.

Mg. Alexandra Paola Guerrero
Editor Revista Ingeniería e Innovación del Futuro
ISSN: 3028-869X - https://editorialscientificfuture.com/index.php/riif

latindex catálogo 2.0

UNEMI UNIVERSIDAD ESTATAL DE MILAGRO

UNEMI

UNIVERSIDAD ESTATAL DE MILAGRO

¡Evolución académica!

@UNEMIEcuador