ISSN: 2992-7447

Año 5. Vol. 3 Núm. 1 Enero - Junio 2025

IDEAS EN CIENCIAS DE LA INGENIERÍA

Facultad de Ingeniería, Universidad Autónoma del Estado de México

IDEAS EN CIENCIAS DE LA INGENIERÍA, Año 5 , Vol. 3, Núm. 1, 2025, es una publicación semestral editada por la Universidad Autónoma del Estado de México, Instituto Literario 100 Ote., Colonia Centro, Toluca, Estado de México, C.P. 50000, Tels. 7222-14-08-55, https://hemeroteca.uaemex.mx/index.php/ideasingenieria, ideasencienciai_fi@uaemex.mx. Directora: Rosa María Valdovinos Rosas, Facultad de Ingeniería, Ciudad Universitaria, Cerro de Coatepec s/n, Toluca, Estado de México, C.P. 50130, Tel. 7222140855. Reserva de Derechos al Uso Exclusivo no. 04-2023-033108322300-102, ISSN: 2992-7447, ambos otorgados por el Instituto Nacional del Derecho de Autor (INDAUTOR). Responsables de la última actualización de este número, Mireya Salgado Gallegos y Javier Salas García, Facultad de Ingeniería. Diseño de portada y Diseño Editorial, Javier Salas García. Fecha de última modificación: 29 de enero de 2025.

CONTENIDO

Pág.

Carta del Editor

Dr. Vicente García Jiménez

La computación se ha posicionado cada vez más como un impulsor importante en muchas disciplinas científicas, áreas de aplicación, sectores y todo aquello que moldea nuestra sociedad. Los diversos enfoques o revoluciones tecnológicas como la industria 4.0 o la sociedad 5.0, que redefinen no solo el panorama tecnológico y las tendencias, sino también las industrias, modelos de negocio o las soluciones a problemas que afectan a la sociedad, tienen dos pilares importantes que son la tecnología e innovación.

En este número especial, hemos reunido a expertos para ofrecer una visión integral de las tendencias actuales y las perspectivas futuras o emergentes en computación, y que están experimentando un crecimiento que será clave en la evolución futura dentro del campo de la computación, ya sea por la creación de nuevas tecnologías, metodologías o aplicaciones. Cada artículo refleja el dinamismo e impacto de las nuevas tecnologías en diversos campos, desde la visión por computadora hasta la salud, el urbanismo y la seguridad, por lo que, también cada uno de ellos presenta una faceta única de cómo la ciencia y la ingeniería responden a los desafíos contemporáneos.

Comenzamos con un tema crucial que involucra la visión por computadora y el procesamiento digital de imágenes: la necesidad de imágenes estandarizadas que permitan un análisis más preciso y eficiente. En este sentido, Palacios-Ortega et al. proponen en su artículo “Diseño de un prototipo ergonómico para la estandarización de imágenes con aplicaciones en visión artificial”, una cabina de iluminación ergonómica para la normalización de imágenes es un ejemplo de cómo la ingeniería puede facilitar procesos complejos, eliminando interferencias ambientales y mejorando la calidad de los datos desde su captura.

En el ámbito de la salud, García-Avalos et al. exploran en su artículo “Hacia un Modelo Gráfico Causal para el Análisis de la Vaginosis Bacteriana”, el uso del Análisis de Trayectorias para modelar la Vaginosis Bacteriana, una enfermedad que afecta a miles de mujeres, donde el modelo gráfico causal no solo aporta una nueva herramienta de análisis, sino que también ofrece una perspectiva innovadora para la interpretación de datos biomédicos. En esta misma línea de aplicación, Vargas-Solar et al. exploran en su artículo “Evaluación y calibración de la calidad de vida en megaciudades inteligentes: Caso de estudio en La Condesa, Ciudad de México”, la calidad de vida en entornos urbanos en donde destacan cómo la tecnología puede proporcionar índices y métricas que permitan a los gobiernos y organizaciones calibrar las políticas públicas para mejorar el bienestar ciudadano.

No menos relevante es el análisis sobre el reconocimiento de emociones a través de la voz, un área clave en la inteligencia artificial aplicada al bienestar humano. Este artículo de Castorena et al., con el título “Exploring Selective Sampling Bounds for Speaker-Dependent

Speech Emotion Recognition", subrayan la importancia de la personalización de modelos de aprendizaje automático para mejorar la precisión en el reconocimiento emocional, un avance crucial para aplicaciones en salud mental e interacción hombre-máquina.

Asimismo, Reyes-Acosta y Rodríguez-Arce exploran en su artículo "Evaluación del Método DBSCAN para la Identificación de Patrones de Estrés en Estudiantes usando Datos no Etiquetados" el empleo de la técnica de clustering DBSCAN para identificar respuestas fisiológicas al estrés, el cual es un problema que presenta retos y oportunidades, ya que sus datos son ruidosos o complejos. Mientras que la investigación de Villegas-Camacho et al. está enfocada a la detección de microplásticos en cuerpos de agua mediante visión por computadora, en donde, el artículo "Uso de YOLOv8 y aumento de datos para la detección de microplásticos en agua" destaca el potencial de las tecnologías emergentes para enfrentar problemas medioambientales de gran escala.

Por último, la creciente preocupación por la violencia y los actos delictivos, especialmente en el contexto de las instituciones educativas, es abordado por Ballesteros-Reyes et al. en "Detección de Crim usando Deep Learning: un caso de estudio en TecEST", en el que resalta cómo el aprendizaje automático puede ofrecer herramientas poderosas para detectar y prevenir delitos. La aplicación del modelo YOLOv5s es un ejemplo claro de cómo la inteligencia artificial puede ser parte de la solución a problemas sociales urgentes.

En conjunto, estos artículos representan la vanguardia de la investigación científica y tecnológica, abordando cuestiones fundamentales de nuestro tiempo: salud, seguridad, medio ambiente, y la calidad de vida. Nos invitan a reflexionar sobre cómo las innovaciones pueden mejorar nuestra sociedad. Esperamos que la lectura de este número les permita reflexionar sobre cómo las innovaciones pueden mejorar nuestra sociedad y que también sirva de inspiración y motivación para seguir explorando las fronteras del conocimiento y la tecnología.

Atentamente,

Dr. Vicente García Jiménez
Editor Invitado
"Ideas en Ciencias de la Ingeniería"

BIOGRAFIA DEL EDITOR

Vicente García Jiménez es Doctor en Sistemas Informáticos Avanzados por la Universitat Jaume I, España. Es Profesor de Tiempo Completo en la Universidad Autónoma de Ciudad Juárez en el Departamento de Ingeniería Eléctrica y Computación. Asimismo, es miembro del Sistema Nacional de Investigadoras e Investigadores SNII Nivel I y de asociaciones científicas como IEEE, ACM, Sigma Xi, y EUREKAS Community. Dirige/Codirige a diversos estudiantes de licenciatura y posgrado en sus investigaciones en la Universidad Autónoma de Ciudad Juárez, Tecnológico Nacional de México campus Toluca y la Universitat Jaume I de Castellón de la Plana España. Realiza la divulgación del conocimiento y la ciencia en eventos académico-científicos a nivel nacional e internacional. Además, cuenta con constante publicación de artículos científicos en revistas especializadas, capítulos de libro con editoriales de prestigio y artículos de divulgación en revistas de gran alcance. Su área de interés está dentro de las áreas del aprendizaje automático, la ciencia de datos y la inteligencia artificial.

IDEAS EN CIENCIAS DE LA INGENIERÍA

Diseño de un prototipo ergonómico para la estandarización de imágenes con aplicaciones en visión artificial

Design of an ergonomic prototype for the standardization of images with applications in artificial vision

Palacios-Ortega, M.[1]*, Cruz-Florez, L.D.[2], Ortiz-Aguilar, L.[3], Mosiño, J.F.[4], Merino-Torres, A.K.[5]
(*marcela.po@purisima.tecnm.mx)

Recibido: 2024-08-08 Aceptado: 2024-10-29

Resumen

Los investigadores en el campo de la visión artificial y del Procesamiento Digital de Imágenes (PDI) coinciden en la necesidad de obtener imágenes estandarizadas que faciliten el pre-procesamiento y el proceso completo, eliminado la interferencia que genera el medio ambiente. El objetivo del presente proyecto es diseñar una cabina de iluminación que integre elementos ergonómicos, que permita normalizar imágenes en cuanto a tamaño, posición e iluminación para eficientizar las técnicas de procesamiento digital, eliminando el ruido del medio ambiente presente en la imagen desde el momento de la adquisición (captura) de imágenes digitales en formato RGB. La metodología implementada es una variación de DFMA "Diseño para la Fabricación y el Montaje" la cual se enfoca en el diseño de componentes que interactúan entre sí. Se comenzó con la definición de conceptos en los que se orienta el producto a diseñar, funciones y componentes principales, se elige el material en el que será construido el prototipo para poder diseñar las piezas, después se ensamblaron los componentes y se realizaron pruebas de interferencia y movilidad. Los resultados alcanzados son: se diseñó una cabina de iluminación en Solid Works, que permitió establecer las dimensiones del prototipo tomando en cuenta condiciones de ergonomía integradas al modelo; un prototipo físico que integra el funcionamiento correcto de los componentes mecánicos y electrónicos de la cabina para la estandarización

[1]División de Ingeniería Industrial, Tecnológico Nacional de México / ITS de Purísima del Rincón, Blvd. del Valle 2301, Guardarrayas, 36418 Purísima de Bustos, Gto.
[2]División de Ingeniería Industrial, Tecnológico Nacional de México / ITS de Purísima del Rincón, Blvd. del Valle 2301, Guardarrayas, 36418 Purísima de Bustos, Gto.
[3]División de Sistemas Automotrices, Tecnológico Nacional de México / ITS de Purísima del Rincón, Blvd. del Valle 2301, Guardarrayas, 36418 Purísima de Bustos, Gto.
[4]Departamento de Sistemas y Computación, Tecnológico Nacional de México / IT de León, Av. Tecnológico S/N Fracc. Ind. Julián de Obregón, 37290 León, Gto.
[5]División de Ingeniería Industrial, Tecnológico Nacional de México / ITS de Purísima del Rincón, Blvd. del Valle 2301, Guardarrayas, 36418 Purísima de Bustos, Gto.

de imágenes, las pruebas del prototipo incluyen los resultados en el histograma para evaluar si el contraste e iluminación son uniformes y guardan consistencia durante la adquisición de imágenes.

Palabras clave: Visión artificial, Diseño de componentes, Estandarización de imágenes, ergonomía, Procesamiento digital de imágenes.

Abstract

Researchers in artificial vision and Digital Image Processing (DIP) agree on obtaining standardized images that facilitate preprocessing and the entire process, eliminating interference generated by the environment. The objective of this project is to design a lighting booth that integrates ergonomic elements, which allows standardizing images in terms of size, position, and lighting to make digital processing techniques more efficient, eliminating environmental noise present in the image from the moment of acquisition (capture) of digital images in RGB format. The methodology implemented is a variation of DFMA "Design for Manufacturing and Assembly," which focuses on the design of components that interact with each other. It began with the definition of concepts in which the product is oriented to design, functions, and main components; the material in which the prototype will be built is chosen to design parts; then the components are assembled, and interference and mobility tests are performed. The results achieved are: a lighting booth was designed in Solid Works, which allowed establishing the dimensions of the prototype taking into account ergonomic conditions integrated into the model; a physical prototype that incorporates the correct operation of the mechanical and electronic components of the booth for image standardization; the prototype tests include the results in the histogram to evaluate if the contrast and lighting are uniform and consistent during image acquisition.

Keywords: Machine Vision, Component Design, Image Standardization, Ergonomics, Digital Image Processing.

Introducción

La transformación digital, caracterizada por un avance acelerado en tecnologías emergentes, exige la implementación de soluciones innovadoras para resolver problemáticas de eficiencia y precisión en los procesos productivos. En el campo de la visión artificial y el Procesamiento Digital de Imágenes (PDI), una de las necesidades prioritarias es la obtención de imágenes estandarizadas, capaces de facilitar las etapas de preprocesamiento y procesamiento de datos, minimizando así el impacto de variables ambientales (como el fondo de la imagen). De acuerdo con Cuevas (2017) et al [1], se utiliza el histograma para identificar características importantes de una imagen, como el contraste y la dinámica, o bien problemas producidos durante la toma de la imagen y que pueden dificultar el procesamiento. De igual forma los errores de iluminación son reconocidos en el histograma.

En el presente proyecto se tiene el objetivo de diseñar un prototipo ergonómico de estandarización de imágenes, capaz de controlar las variables presentes en el proceso de captura de imágenes digitales como el fondo de la imagen, la iluminación, la posición del objeto de estudio y la posición de la cámara las cuales, dificultan el análisis pixel a pixel de una imagen. Es fundamental mantener las variables controladas en los análisis pixel a pixel, ya que es necesario garantizar condiciones de captura uniformes. La variación en la posición del objeto puede complicar el análisis, ya que el sistema de visión, al captar el mismo

https://doi.org/10.36677/rici.v3i1.24445 Enero-Julio 2025 ISSN: 2992-7447 págs. 4-21

objeto desde diferentes ángulos, podría interpretarlo como uno distinto. Además, la iluminación durante la captura también influye, ya que puede cambiar según la hora del día, generando diferentes contrastes y sombras que complican el procesamiento. También es crucial eliminar el fondo de la imagen, ya que éste introduce pixeles irrelevantes para el estudio, lo que dificulta el análisis. La finalidad del proyecto en cuestión es facilitar el reconocimiento de imágenes en la visión artificial utilizando menor procesamiento al contar con imágenes estandarizadas. El reconocimiento requiere de condiciones estables en la captura de imagen para reducir las variables que puedan alterar o generar ruido innecesario en el proceso. Para esto se buscó normalizar las condiciones lumínicas y de posición tanto de la cámara como la del objeto a analizar. Se utiliza la metodología DFMA (*Design for Manufacturing and Assembly*) [2] con enfoque en la fabricación y montaje, esta metodología implica definir las características básicas que requiere el producto, definir las especificaciones funcionales para realizar diseño de los componentes que integrarán el prototipo, definición de materiales de construcción, diseño de componentes, prototipado, ensamble, pruebas y correcciones.

El artículo se organiza de la siguiente manera, en la sección estado del arte se describen los trabajos relacionados con aplicaciones en visión artificial. En la sección marco teórico, se describen los conceptos y teorías que adquieren relevancia en el desarrollo del proyecto en cuestión, en la sección metodología se ilustran los pasos para solucionar el problema, en desarrollo de la metodología se explican los pasos ejecutados durante proyecto y en la sección resultados se reportan los resultados de esta investigación, por último, se redactan las conclusiones del proyecto.

Estado del Arte

En esta sección se proporciona una visión general de estado del arte en las investigaciones relacionadas con diseño de sistemas para control de iluminación, así como diversas aplicaciones de la inteligencia artificial.

En la investigación realizada por Zhao J. et al (2017) [3] se propone un nuevo método de estimación de iluminación utilizando un modelo de red neuronal profunda llamado DLNet. Este modelo es capaz de predecir mapas de entorno de alto rango dinámico (HDR) a partir de una sola fotografía con un campo de visión (FOV) limitado. La red introduce un enfoque de filtrado dinámico a través de módulos de convolución esférica multi-escala (SMD), lo que permite generar filtros de convolución específicos para cada entrada, ajustándose a las variaciones de iluminación y características de la imagen. Por otra parte, el proyecto de Maza V. et al (2017) [4], propone un sistema prototipo con visión por computadora que realiza la selección automática del grano de cacao orgánico según sus características externas. El procesamiento de imágenes se realiza usando OpenCV aplicado en Raspberry Pi para la clasificación del cacao, ofreciendo una alternativa de evaluación automatizada no invasiva y rentable, que reemplaza los métodos de inspección manual tradicional. De cada imagen digital se pueden extraer propiedades del producto escogido tales como forma, tamaño, color, etc. Bautista en 2019 [5] propone un sistema de visión artificial para análisis de imágenes espectrales útil en la detección de características en los cultivos de brócoli, adquiridas por medio de una cámara Survey 2NDVI red+ NIR, adaptada a un dron y procesada en Matlab. Las imágenes que se obtienen, debidas a la reflectancia del sol, hacen necesaria una calibración de reflectancia por comparación directa, empleando las tarjetas calibradas proporcionadas por una empresa específica. Estas tarjetas incluyen tres objetos de calibración con curvas de reflectancia estanda-

rizadas que permiten normalizar las imágenes tomadas con la cámara SURVEY 2 Red + NIR al ingresar los valores de reflectancia obtenidos de las tarjetas calibradas en el algoritmo de procesamiento. Como resultados presentan la obtención de imágenes claras, sin distorsión y con datos del índice de vegetación de diferencia normalizada (NDVI) en cultivos de brócoli. Con dicho índice, se pudo estimar la calidad, cantidad y desarrollo de los cultivos a través de la codificación en colores del NDVI. Serrano (2020) [6] presenta un proyecto centrado en el desarrollo de un prototipo de aplicación móvil que utiliza inteligencia artificial y análisis de imágenes para identificar mazorcas de cacao enfermas. El autor captura las imágenes con la ayuda de un experto, para posteriormente darle una orientación vertical al cacao para entrenar y calibrar el aprendizaje por medio YOLOv4 (You Only Look Once), utilizando imágenes de mazorcas afectadas por las plagas Fitóftora y Monilia donde se logra una precisión del 60 % en la detección, lo que ofrece una herramienta útil para agricultores e investigadores, facilitando la inspección sin necesidad de un experto en fitosanidad. En la investigación realizada por Ikeuchi, K. et al (2020) [7], titulada “Sistemas de Visualización/AR/VR/MR", aborda las aplicaciones de las técnicas de iluminación activa en la mejora de los sistemas de realidad aumentada (AR), realidad virtual (VR) y realidad mixta (MR). En dicha investigación se destaca cómo el mapeo de proyección y la iluminación activa pueden usarse para manipular la apariencia de objetos, haciéndolos parecer dinámicos incluso cuando están estáticos. Los puntos clave que se describen incluyen: Mapeo de Proyección, Reproducción de BRDF, Interacción Háptica, Integración con Impresión 3D y Metamerismo.

Por otro lado, en la investigación hecha por Ikeuchi K. et al (2020) [8], aborda dos aplicaciones clave en robótica: la Localización y Mapeo Simultáneos (SLAM) y el Aprendizaje por Observación (LfO). Estas tecnologías son fundamentales para mejorar la navegación y manipulación autónoma en robots. En cuanto a Localización y Mapeo Simultáneos (SLAM) los autores definen que el SLAM es una técnica utilizada para que los robots naveguen creando mapas 3D mientras se mueven y estiman su propia posición. Se presentan métodos como el "stop-and-go", donde el robot avanza, recoge datos y los compara con la información previa, y el escaneo continuo de una sola línea, que permite capturar datos en tiempo real sin detenerse. En cuanto al Aprendizaje por Observación (LfO) se explica cómo los robots pueden aprender a realizar tareas observando a humanos. A través de este enfoque, se abstraen los significados de las acciones humanas para luego mapearlos en movimientos robóticos.

Dado el contexto anterior, la planificación de actividades para realizar el diseño adecuado del prototipo que controle diferentes variables y parámetros que faciliten el pre-procesamiento y el procesamiento digital de imágenes, adquiere mayor importancia ya que las aplicaciones de la visión por computadora proporcionan una alternativa eficiente y económica para dar soluciones a problemáticas de diversa índole.

Marco teórico

El DFMA, que se traduce como "Diseño para la Fabricación y el Montaje", es un conjunto de técnicas y metodologías destinadas a optimizar el diseño o rediseño de un producto. Su objetivo principal es mejorar los aspectos relacionados con la fabricación, el ensamblaje y los costos, sin comprometer las funciones fundamentales del producto [2].

https://doi.org/10.36677/rici.v3i1.24445 Enero-Julio 2025 ISSN: 2992-7447 págs. 4-21

Ergonomía

La ergonomía es según [9] la disciplina científica que se ocupa de la comprensión de las interacciones entre los humanos y otros elementos de un sistema, y la profesión que aplica teoría, principios, datos y métodos para diseñar con el fin de optimizar el bienestar humano y el rendimiento general del sistema. La ergonomía con propósitos de diseño se enfoca en que los productos, sistemas o ambientes destinados al uso humano se basen en las características físicas y mentales de los usuarios, que faciliten el trabajo y eviten situaciones poco saludables, incómodas e ineficientes, teniendo en cuenta las capacidades físicas y psicológicas de las personas [10].

Sistemas de control

Un sistema de control es un conjunto de dispositivos automatizados que controlan la energía de diversos dispositivos para hacer su correcto funcionamiento y a su vez realizar tareas en específico [11]. Los elementos de los sistemas de control incluyen:

- Generadores: es un dispositivo encargado de proporcionar la energía al sistema, por ejemplo, las baterías.
- Conductores: son los conectores que permiten el paso de corriente eléctrica como lo son los cables.
- Receptores: es el que recibe una señal eléctrica para hacer alguna acción o efecto. Son los encargados de transformar la corriente eléctrica en otros tipos de energía útil; por ejemplo, las lámparas o los altavoces.
- Elementos de maniobra y control: un ejemplo son los interruptores, los cuales se encargan de manipular y controlar ciertos elementos del sistema.
- Protectores: un ejemplo son los fusibles, estos se encargan de proteger al sistema de descargas y así evitar que colapse el mismo.

Microcontrolador ESP32

El ESP32-WROOM-32 es un módulo MCU que incluye Wi-Fi, Bluetooth, BLE, para tareas como la codificación de voz, la transmisión de música y la decodificación de MP3. El componente principal de este módulo es el chip ESP32-D0WDQ6. Este chip integrado está diseñado para ser escalable y adaptable. Cuenta con dos núcleos de CPU que pueden controlarse de forma independiente, y la frecuencia de reloj de la CPU es ajustable entre 80 MHz y 240 MHz [12].

Obtención de imágenes

La visión artificial es una rama de la inteligencia artificial que permite a los ordenadores y sistemas interpretar información a partir de imágenes digitales, videos y otras entradas visuales, y actuar o hacer recomendaciones basadas en esa información. A diferencia de la visión humana, que se beneficia de toda una vida de contexto para interpretar imágenes, la visión artificial entrena a las máquinas para realizar estas tareas en un tiempo mucho más corto, utilizando cámaras, datos y algoritmos [13].

Componentes de una cámara

De acuerdo con [13], el diafragma es un dispositivo situado en el interior del objetivo, formado por laminillas metálicas de color negro, colocadas de modo que, al superponerse, dejan un orificio central, regulable mediante el movimiento de estas. La función del orificio del diafragma es controlar el paso de la luz hasta la película, si el orificio está muy abierto pasará más luz. El tamaño del orificio influirá en la nitidez obtenida en la fotografía de los elementos que se hallen a distancias diferentes de la cámara. Este fenómeno es conocido como profundidad de campo. El obturador es un mecanismo destinado, como el diafragma, a controlar la cantidad de luz que llega a la película. El tiempo de obturación es directamente proporcional a la cantidad de luz que recibe el sensor de la cámara. En objetivos con movimiento es recomendable utilizar velocidades de obturación rápidas para lograr una congelación de imagen.

Análisis Pixel a Pixel

Una imagen digital se representa comúnmente como una función bidimensional de intensidad de luz, denotada por *I (x, y*), donde el valor de la intensidad (o tono de gris) se expresa mediante una matriz (Figura 1):

$$I(x,y) = \begin{pmatrix} I(0,0) & I(0,1) & \dots & I(0,M-1) \\ I(1,0) & I(1,1) & \dots & I(1,M-1) \\ I(N-1,0) & I(N-1,1) & \dots & I(N-1,M-1) \end{pmatrix}$$

Figura 1: *Función bidimensional de intensidad de luz.*

Una imagen digital se puede expresar por medio de una matriz *M x N* elementos llamados pixeles, el valor de un pixel representa la intensidad o color de la imagen en ese punto [14]. En la Figura 2 se muestra un fragmento de una imagen digital de M x N elementos y su intensidad registrada en cada pixel y sus valores de intensidad registrados en el programa Matlab.

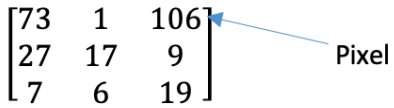


Figura 2: *Fragmento de una imagen digital.*

El análisis pixel a pixel es una técnica utilizada en procesamiento de imágenes y visión artificial que evalúa y compara cada píxel de una imagen o conjunto de imágenes. Esto se hace para identificar y medir características específicas, como cambios de color, patrones, bordes o texturas, esta evaluación se realiza de la siguiente forma: Se comparan los pixeles de la imagen muestra con los pixeles de las imágenes guardadas en la base de datos, se utiliza el Coeficiente de Correlación de Pearson por medio del comando

https://doi.org/10.36677/rici.v3i1.24445 Enero-Julio 2025 ISSN: 2992-7447 págs. 4-21

CoefCor para identificar cuál de las imágenes de la base de datos tiene el mayor índice de correlación con la imagen muestra, esta acción se ejecuta por medio del Algoritmo 1.

Algoritmo 1 Algoritmo para análisis pixel a pixel

```
Require: n ≥ 0, imagen
1: Hacer desde i ← 0 hasta n − 1 hacer
2:     Hacer desde j ← 0 hasta n − 1 hacer
3:         indicei=num2str(i);
4:         indicej = num2str(j);
5:         Archivo = strcat('modelo',indicei,'_',indicej,''.jpg');
6:         imagen2=imread(archivo);
7:         imagen2=rg2gray(imagen2);
8:         coeficientes=correcoef(double(im3),double(imagen2));
9:         Coefcor(i,j)=coeficientes(1,2);
10:    Fin
11:Fin
12: valor_maximo=max(max(CoefCor));
13:[fila columna valor]=find(CoefCor==valor_maximo)
14:indicei=num2str(fila);
15:    indicej=num2str(columna);
16:    archivo= strcat('modelo',indicei,'_',indicej,'.jpg');
17:A=imread(archivo);
18: Regresar A
```

METODOLOGÍA

Para el desarrollo del diseño del prototipo se optó por una metodología basada en DFMA la cual tiene un enfoque de diseño de piezas de forma modular, se pretende que sean fáciles de ensamblar, manufacturar y posteriormente darles un mantenimiento sencillo o ser remplazadas de ser necesario. En la Figura 3 se muestra un esquema de las etapas que se siguieron para el desarrollo del proyecto y cuya descripción son:

I. Concepto. En esta fase se determinan las características básicas que requiere el producto, así como el objetivo del producto, funcionamiento general, interacción entre componentes y consideraciones generales

II. Especificaciones. Se establecen y definen las características y requerimientos establecidos con anterioridad, con el fin de generar un enfoque para el diseño de cada una de las piezas requeridas en el prototipo.

III. Detalles. En esta etapa se incorporan las especificaciones y los bocetos para generar un diseño detallado del producto en CAD, considerando el proceso y material de construcción empleado para cada componente, así como su interacción en los ensambles y subensambles presentes en el producto.

IV. Prototipado. Se fabrican y ensamblan todos los componentes de acuerdo con los planos y archivos de referencia generados con anterioridad, se realizan pruebas y los ajustes necesarios para que se cumpla con los requerimientos de función establecida en la primera etapa.

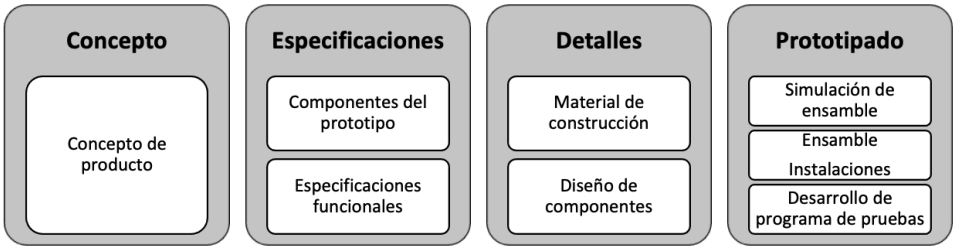


Figura 3: *Diagrama de metodología de diseño.*

DESARROLLO DE LA METODOLOGÍA

Concepto. La función principal de la cabina es capturar imágenes para su posterior procesamiento. Estas imágenes deben mantener una iluminación uniforme y una posición constante que no interfiera con el análisis Pixel a Pixel. El sistema de reconocimiento analiza cada imagen Pixel a Pixel; por lo tanto, es crucial que las imágenes sean lo más similares posible entre cada captura, ver Figura 4.

Figura 4: *Variables a considerar en la captura de imágenes. Fuente: Elaboración propia.*

En la Figura 4 se presentan tres variables principales: iluminación, posición de la cámara y posición del objeto a analizar. Las cuales se consideran importantes de controlar para asegurar que el sistema trabaje de forma adecuada, estas variables serán incluidas en el diseño del concepto del prototipo en conjunto con las especificaciones de ergonomía y los componentes físicos que constituirán el prototipo.

Especificaciones. Los aspectos ergonómicos son indispensables al momento de diseñar ya que permitirán adecuar el modelo al uso del operador de manera fácil y segura. La parte de electrónica permite controlar la iluminación de la cabina y asegurar el centrado del objeto en la misma posición, esto con la finalidad de facilitar el procesamiento, para ello se integran el microcontrolador, el láser y sensores que permiten encender y apagar el sistema para la captura en tiempo real de las imágenes. En cuanto al sistema de reconocimiento se diseña para integrar la cámara en el prototipo la cual se utiliza en la captura

https://doi.org/10.36677/rici.v3i1.24445 Enero-Julio 2025 ISSN: 2992-7447 págs. 4-21

de las imágenes, se adapta la pantalla *touch*, así como el cableado y demás componentes necesarios para la integración del sistema. En la Figura 5 se representa de forma esquemática el funcionamiento de la cabina.

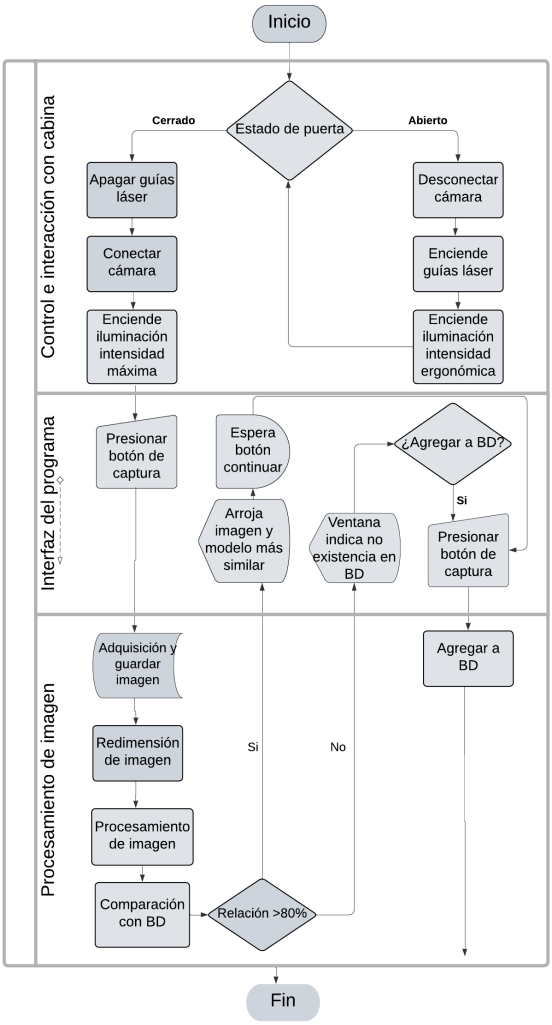


Figura 5: *Funcionamiento del prototipo Fuente: Elaboración propia.*

Detalles (Elección de material y diseño de componentes). Para la selección del material se categorizan las piezas de acuerdo con las necesidades; se utiliza el método de calificación por factores para evaluar y seleccionar el material óptimo para cada grupo de componentes. La elección de ponderaciones se realiza en consenso del equipo de trabajo, para cada criterio se define un rango de 1 a 10 donde un mayor valor corresponde a una característica deseada. Para la estructura los criterios a evaluar son: maquinabilidad, resistencia, estética y costo; cada uno con sus propias ponderaciones de acuerdo con la necesidad establecida.

Prototipado (Ensamble). En este apartado se muestran imágenes de referencia de los planos utilizados, y procesos realizados en la etapa de construcción del prototipo. Para el ensamble de los componentes el programa de diseño permite obtener los planos correspondientes al modelo, en la Figura 6 se muestra un ejemplo del diseño del sistema de elevación, dimensiones de la cabina y demás accesorios que conforman el proyecto. Para complementar el plano ilustrado en la Figura 6, en la Tabla 1 resumen los elementos que conforman el prototipo, el número de pieza, el material con el que será fabricado, así como los requerimientos en cantidades a cortar para el ensamble, además de que se especifica el plano de referencia de ingeniería que se utilizó para el ensamblaje.

Análisis Pixel a Pixel

El análisis Pixel a Pixel es un enfoque fundamental para el procesamiento detallado de imágenes digitales, especialmente en aplicaciones de visión artificial [15]. En este trabajo, se utilizó el análisis Pixel a Pixel para asegurar la precisión en el reconocimiento de objetos dentro del prototipo de cabina de iluminación, cuyo objetivo es estandarizar las imágenes capturadas. Este tipo de análisis evalúa individualmente cada Pixel de una imagen, identificando diferencias sutiles en términos de color e intensidad que pueden afectar el procesamiento y reconocimiento.

La principal dificultad en el análisis Pixel a Pixel radica en las variables externas que interfieren durante la captura de la imagen, tales como la iluminación ambiental y la posición del objeto. Por ello, se integró un sistema ergonómico en la cabina que controla tanto la posición de los objetos de estudio como la iluminación, lo que asegura que las imágenes mantengan condiciones homogéneas. De este modo, se eliminan sombras y otros artefactos visuales que podrían distorsionar el análisis detallado de los Pixeles. Además, para validar el sistema, se evaluaron las imágenes mediante histogramas y coeficientes de correlación de Pearson. Esto permitió determinar que la iluminación y el contraste de las imágenes eran uniformes, lo que facilitó el análisis Pixel a Pixel sin la interferencia de variables externas.

Tabla 1: Especificaciones de materiales y planos de referencia

N°	N° de Pieza	Material	Cantidad
1	CABINA	PTR	1
2	PATAS	PTR	1
3	BASE 3	MDF	8
4	BRAZO C	MDF	16
5	SEPARADOR 1	MDF	16

https://doi.org/10.36677/rici.v3i1.24445 Enero-Julio 2025 ISSN: 2992-7447 págs. 4-21

Tabla 1: Especificaciones de materiales y planos de referencia

N°	N° de Pieza	Material	Cantidad
6	SEPARADOR 3	MDF	4
7	BRAZO E	MDF	4
8	SEPARADOR 4	PLA	8
9	GUIA 2	PLA	2
10	EJE 1B	ALUMINIO	2
11	EJE 1	ALUMINIO	2
12	EJE 3	ALUMINIO	8
13	EJE 5	ALUMINIO	4
14	SIN FIN	VARILLA ROSCADA 3/8 IN	2
15	SEGURIDAD OMEGA	PLA	30
16	SOPORTE BALERO	PLA	4
17	BASE 5	MDF	4
18	BASE 4	MDF	4
19	BRAZO ENGRANE A	MDF	8
20	ENGRANE 1	PLA	4
21	ENGRANE 2	PLA	2
22	SEPARADOR 2	PLA	2
23	SOPORTE MANERAL 1	MDF	4
24	EJE 8	ALUMINIO	2
25	EJE 9	ALUMINIO	2
26	AFBMA 12.1.4.1-0080-228,DE,NC,8_68	RODAMIENTO 8-22-6	6
27	AFBMA 20.1-02-15-10.DE.NC.10_68	RODAMIENTO 15-35-11	4
28	HBOLT 0.3750-24X.25X.25-N	TORNILLO 3/8 IN	32
29	HNUT 0.3750-24-D-N	TUERCA 3/8 IN	32
30	EJE 10	ALUMINIO	2
31	MANGO	PLA	4
32	SEPARADOR 5	PLA	4

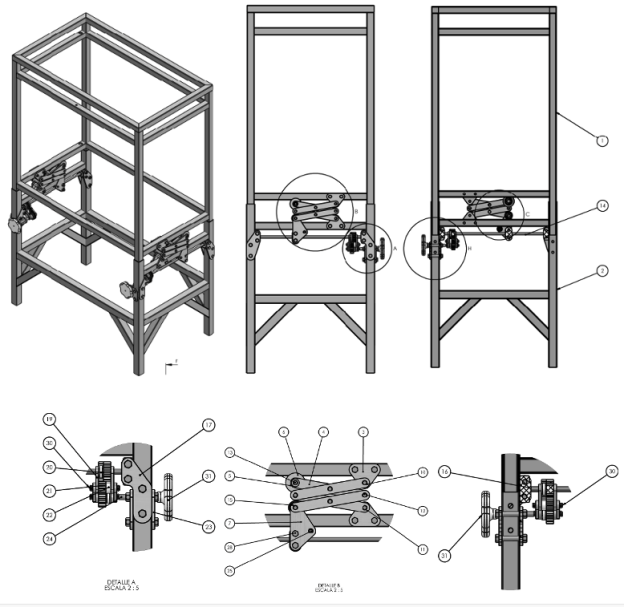


Figura 6: *Plano ING-E01. Fuente: Elaboración propia. En el detalle A se muestran los componentes que conforman la manivela y el engranaje, en el detalle B se muestran los componentes para el sistema de elevación ergonómico.*

RESULTADOS

Una vez desarrollados los planos de ingeniería, cortadas las piezas e impresos los materiales de 3D se ensamblan los componentes en el prototipo funcional, a manera de resumen en la Tabla 2 se explica la función de cada diseño y su aportación, se destacan las funciones de los elementos y se anexa imagen del diseño final en físico.

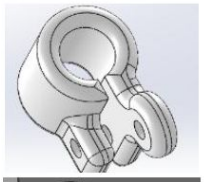
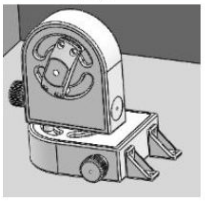
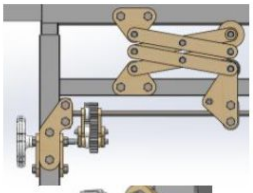
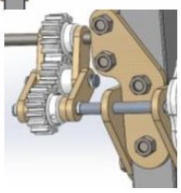
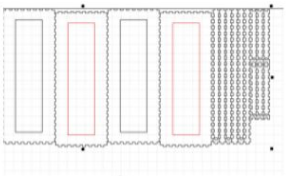
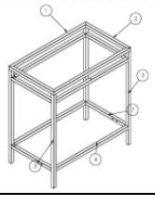
Diseño final. Se presenta el diseño del prototipo en la Figura 7 con todos los elementos integrados en cuanto a ergonomía y la parte electrónica, que permite encender el sistema para la captura de las imágenes. Una vez ensamblado el prototipo, incluyendo los dispositivos de control de iluminación, cámara, láser y demás componentes integrados a la cabina, se procede a evaluar la normalización de la imagen en cuanto a:

1. Normalizar la posición del objeto a estudiar. Para lograr este objetivo se requiere ajustar la posición de la cámara y utilizar el láser como guía para colocar el objeto en la misma posición. Para la primera parte, se realizaron los ajustes pertinentes para lograr eliminar en un alto grado las sombras presentes durante la captura. De igual forma el láser adaptado a la cámara permite colocar el objeto de estudio en la misma posición utilizando el haz de luz como guía para ubicar el objeto. Es importante mencionar que una vez que se cierra la puerta de la cabina, el sensor adaptado también permite que se apague la luz del láser para no interferir en la toma de la imagen.

2. Normalizar la iluminación. Para evaluar los resultados en el sistema de reconocimiento se utiliza el coeficiente de correlación de Pearson.

https://doi.org/10.36677/rici.v3i1.24445 Enero-Julio 2025 ISSN: 2992-7447 págs. 4-21

Tabla 2: Resumen de elementos, diseño y diseño final

Pieza	Función	Diseño	Pieza física
Soporte de diodo láser	Permite colocar el láser que servirá para proyectar las guías láser para posicionar el objeto siempre en la misma posición dentro de a cabina. **Variable que estandariza: Posición del objeto a analizar dentro de la cabina.**		
Porta diodo	El soporte de diodo láser permite proyectar líneas guías para estandarizar el centrado del objeto en una posición fija. El laser se apaga en cuanto se accione el microinterruptor al cerrar las puertas del prototipo. **Variable que estandariza: Posición del objeto a analizar dentro de la cabina.**		
Soporte Cámara	Permite colocar la cámara de captura de imágenes en una posición fija, las roscas diseñadas permiten girar la base para ajustar el ángulo en la posición deseada. **Variable que estandariza: Posición de la cámara y ángulo de inclinación.**		
Sistema de elevación	Permite ajustar la altura de la cabina, a la altura del operador, para evitar posiciones forzadas: **Elemento diferenciador: Provee ergonomía al prototipo**		
Componente del sistema de elevación	Complementa el funcionamiento del sistema de elevación ergonómico, ayuda a elevar la altura, para un manejo más accesible para el operador de la cabina **Elemento diferenciador: Provee ergonomía al prototipo**		
Sistema de iluminación	Permite regular la intensidad: Ajuste mínimo, cuando la cabina esta abierta para colocar el objeto de estudio sin dañar la vista del operador. Intensidad máxima, para permitir una ilumniación óptima dentro de la cabina: **Variable que controla: Imágenes estandarizadas simpre a la misma intensidad.**		
Piezas para ensamble de estructura	Piezas que conformaran la estructura física del prototipo, su función es soportar el peso de los elementos que integraran el modelo. **Elemento diferenciador: Permite la integración de todos los elementos del prototipo, al tiempo que soporta el peso de los elementos.**		

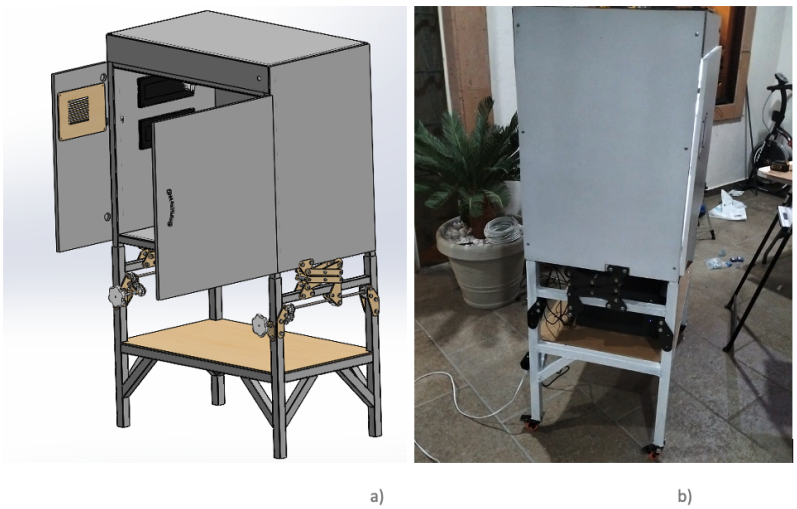


Figura 7: *a) Vista en isométrico del modelo b) Vista lateral izquierda del modelo con enfoque en el sistema ergonómico. Fuente: Elaboración propia.*

En la Tabla 3 se resumen los resultados del reconocimiento utilizando el *scrip* anterior en conjunto con la cabina de iluminación, se contrasta la imagen obtenida sin la cabina, así como los resultados arrojados por el sistema de reconocimiento. El eje de las 'x' representa el valor del pixel que va de cero a 255 (en cuanto a blanco y negro); mientras en que el eje de las 'y' representa la frecuencia de dicho pixel. Los tres objetos presentados en la Tabla 4 se deducen los siguientes hechos:

1. **ZAPATO DE REFERENCIA O CONTROL.** Se estudia un modelo sin la cabina de iluminación, se observa en el histograma una distribución alta de los tonos oscuros, además de que el fondo interfiere con el análisis pixel a pixel.

2. **BOTA.** En la bota se observó lo siguiente:

Bimodalidad: El histograma tiene dos picos principales, uno cerca de la parte baja (alrededor de 0-50) y otro hacia la parte más alta (150-200). Esto sugiere que la imagen contiene principalmente dos grupos de Pixeles: Un grupo de Pixeles oscuros (valores bajos, entre 0 y 50). Un grupo de Pixeles claros (valores alrededor de 150-200).

Valle en el centro: Entre los valores de 50 y 150, hay una disminución significativa en el número de Pixeles. Esto indica que la imagen tiene pocos tonos intermedios, lo que sugiere un alto contraste entre las áreas oscuras y claras.

Mayor concentración en el rango alto: Hacia el final del histograma (entre 150-200), hay un aumento en la cantidad de Pixeles, lo que sugiere que una gran parte de la imagen es relativamente clara o está iluminada.

3. **BOTIN.** Para el segundo zapato de referencia se observó:

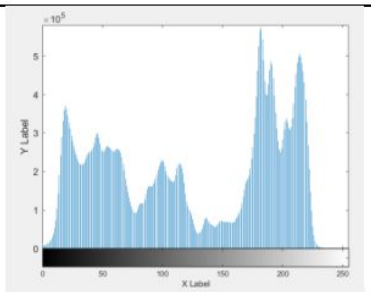
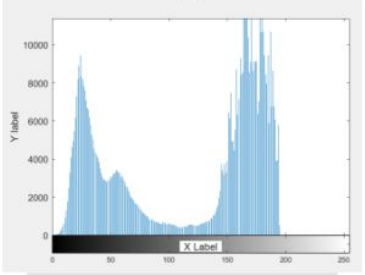
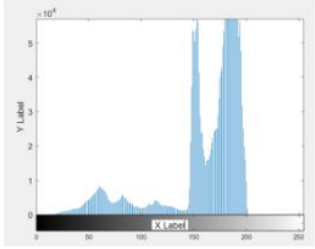
https://doi.org/10.36677/rici.v3i1.24445 Enero-Julio 2025 ISSN: 2992-7447 págs. 4-21

Predominancia de tonos claros (150-200): El histograma tiene dos picos muy marcados en el rango de 150 a 200 en el eje X, lo que indica que una gran parte de los Pixeles en la imagen tienen un brillo alto. Esto sugiere que la imagen está dominada por áreas bien iluminadas o claras. Los dos picos cercanos en este rango podrían representar la presencia de dos áreas prominentes con niveles de iluminación similares, pero no idénticos.

Tonos oscuros y medios (0-150): En la región entre 0 y 50, se observa un número relativamente bajo de Pixeles, lo que indica que solo una pequeña parte de la imagen contiene áreas oscuras o en sombra. Entre 50 y 150 hay una presencia moderada de Pixeles, lo que sugiere que la imagen tiene algunos detalles en tonos medios, pero no son predominantes en comparación con los tonos claros. Estos podrían representar áreas de transición o detalles de menor contraste en la imagen.

Baja variabilidad en tonos intermedios: A pesar de que hay una pequeña cantidad de Pixeles en el rango medio, no hay ningún pico importante en esta zona, lo que indica que los tonos intermedios no son muy destacados. Esto sugiere que la imagen tiene un contraste elevado, con mayor presencia de áreas claras y oscuras en lugar de tonos de transición suaves.

Tabla 3: Resumen de elementos, diseño y diseño final

Objeto a estudiar	Histograma de la imagen	Eficiencia del reconocimiento
		El sistema de reconocimiento no logra identificar adecuadamente al objeto de estudio ya que no se encuentra estandarizado el fondo de la imagen, el tamaño de las imágenes y la iluminación es muy variada. **Resultado: No reconocido**
		Se ingresó el modelo de la derecha al sistema de reconocimiento y encontró una semejanza con el modelo de 0.89. esto representa una coincidencia del **89% respecto a la semejanza** pixel a pixel, del objeto de estudio y el modelo guardado en la Base de Datos.
		Se ingresó el modelo de niño en la base de datos del sistema de reconocimiento, al ser un modelo que se encuentra en la BD se reconoció con **un índice del 92% de semejanza**

Una vez desarrollado el prototipo, se procedió a la realización de las pruebas de funcionamiento. Durante estas pruebas, se obtuvo una imagen nítida, sin sombras ni ruido detectable a simple vista. El

micro interruptor funcionó correctamente, detectando cuando se abría y cerraba la puerta, lo que permitió encender y apagar la guía de luz de manera efectiva. Sin embargo, con el uso continuo, se observaron señales falsas generadas por el micro interruptor, lo que llevó a la necesidad de implementar una cerradura magnética para mejorar la precisión y fiabilidad del sistema. En cuanto a la estandarización de imágenes, se logró colocar los objetos de estudio en una posición similar gracias al uso de un láser. Además, se logró obtener un fondo uniforme y sin elementos extraños en las imágenes adquiridas, lo cual es esencial para el análisis de Pixel a Pixel. El fondo blanco resultante fue adecuado para este tipo de análisis.

Conclusiones

En este trabajo se ha desarrollado un prototipo ergonómico para la estandarización de imágenes, cuya finalidad es mejorar la calidad y consistencia en la captura de imágenes para aplicaciones de procesamiento digital. El prototipo diseñado integra un sistema de iluminación controlada y un ajuste ergonómico que permite normalizar variables clave como la posición del objeto y la cámara, así como la iluminación, eliminando las interferencias causadas por factores externos. Los resultados obtenidos mediante pruebas de histogramas y análisis pixel a pixel han demostrado que el prototipo es eficaz en la estandarización de imágenes, lo que facilita el procesamiento posterior y mejora la precisión en el reconocimiento de objetos. El diseño ergonómico del sistema, que incluye un mecanismo de ajuste de altura y control de la iluminación, aporta una ventaja significativa al reducir la fatiga del operador y asegurar condiciones uniformes durante la captura de imágenes. Estas características convierten al prototipo en una herramienta robusta y flexible para diversas aplicaciones en visión artificial y procesamiento de imágenes, donde la precisión es crítica.

En términos de impacto, este proyecto proporciona una solución práctica para el desafío de obtener imágenes consistentes en entornos industriales o de investigación, donde las condiciones ambientales pueden afectar el rendimiento de los algoritmos de reconocimiento. Cabe mencionar que dicho prototipo es funcional y aplicable en las zapaterías, ya que su costo no es elevado y el diseño no es bromoso, lo que permite poder replicarlo para diferentes sucursales, ya que dichos elementos también llegaron a delimitar el alcance de esta investigación. Además, el prototipo tiene un gran potencial para ser aplicado en campos como la agricultura, la robótica y la inspección automatizada, donde la estandarización de imágenes es fundamental para la precisión de los análisis. Futuras investigaciones podrían enfocarse en la mejora de la versatilidad del prototipo, integrando tecnologías más avanzadas de reconocimiento de imágenes y ampliando su aplicabilidad a otros tipos de sensores. Además, la implementación de otros algoritmos de inteligencia artificial como redes neuronales directamente en el sistema de captura podría optimizar aún más los procesos de reconocimiento en tiempo real.

https://doi.org/10.36677/rici.v3i1.24445 Enero-Julio 2025 ISSN: 2992-7447 págs. 4-21

Agradecimientos

Agradecemos al Tecnológico Nacional de México por su contribución al desarrollo del presente proyecto dentro del marco de la convocatoria 2024: Proyectos de investigación científica, desarrollo tecnológico e innovación.

Agradecemos al Consejo Nacional de humanidades, ciencias y Tecnología (CONAHCYT) en conjunto con el Tecnológico Nacional de México/Instituto Tecnológico Superior de Purísima del Rincón División de ingeniería Industrial y división de ingeniería en Sistemas Automotrices por su tiempo y apoyo para el desarrollo de esta investigación.

Declaración de contribución de autores y colaboradores

Marcela Palacios Ortega: elaboró conceptualización, software, Redacción y borrador original. **Luis Daniel Cruz Florez:** apoyo Metodología, administración del proyecto. **Lucero de Monserrat Ortiz-Aguilar:** apoyo en Investigación, validación, Redacción, revisión y edición. **Juan Francisco Mosiño:** realizó Análisis formal, Supervisión. **Merino-Torres A.K** apoyo en investigación.

Referencias

[1] E. Cuevas, M. D. Cortés, y J. O. C. Méndez. *Tratamiento de imágenes con MATLAB*. Marcombo, 2018.

[2] H. M. Ramirez, E. E. N. Núñez, y J. T. Salcedo. *Diseño para la fabricación y ensamble de productos soldados. Un enfoque metodológico y tecnológico*. Universidad del Norte, 2009.

[3] J. Zhao, A. Chalmers, y T. Rhee. Adaptive Light Estimation using Dynamic Filtering for Diverse Lighting Conditions. *IEEE Trans Vis Comput Graph*, vol. 27, núm. 11, pp. 4097–4106, nov. 2021. doi: 10.1109/TVCG.2021.3106497.

[4] V. Maza y G. Isamar. Procesamiento de imágenes usando OpenCV aplicado en Raspberry Pi para la clasificación del cacao. 2017.

[5] R. A. Bautista, P. Constante, A. Gordon, y D. Mendoza. Diseño e implementación de un sistema de visión artificial para análisis de datos NDVI en imágenes espectrales de cultivos de brócoli obtenidos mediante una aeronave pilotada remotamente. *Infociencia, São Luís*, vol. 12, núm. 1, pp. 30–35, 2019.

[6] S. A. J. Sebastián y T. V. C. Andrés. Prototipo de aplicación móvil para la identificación de mazorcas de cacao enfermas haciendo uso de visión por computadora y aprendizaje de máquina. 2020, [En línea]. Disponible en: http://hdl.handle.net/20.500.12749/13367.

[7] K. Ikeuchi *et al.* Visualization/AR/VR/MR Systems. 2020, pp. 213–239. doi: 10.1007/978-3-030-56577-0_9.

[8] K. Ikeuchi *et al.* Robot Vision, Autonomous Vehicles, and Human Robot Interaction. 2020, pp. 289–303. doi: 10.1007/978-3-030-56577-0_12.

[9] Y. T. Pérez. *Principios teórico-prácticos de ergonomía para el diseño y evaluación de herramientas, puestos de trabajo y máquinas*, vol. 51. Editorial de la Universidad Pedagógica y Tecnológica de Colombia-UPTC, 2022.

[10] L. S. Nogueira. Importancia de la ergonomía en el diseño de productos. *Actas de Diseño*, núm. 11, 2011.

[11] R. P. Borase, D. K. Maghade, S. Y. Sondkar, y S. N. Pawar. A review of PID control, tuning methods and applications. *Int J Dyn Control*, vol. 9, núm. 2, pp. 818–827, jun. 2021. doi: 10.1007/s40435-020-00665-4.

[12] D. Hercog, T. Lerher, M. Truntič, y O. Težak. Design and Implementation of ESP32-Based IoT Devices. *Sensors*, vol. 23, núm. 15, ago. 2023. doi: 10.3390/s23156739.

[13] K. Ikeuchi *et al.* Geometry. 2020, pp. 31–62. doi: 10.1007/978-3-030-56577-0_2.

[14] R. Parekh. *Fundamentals of IMAGE, AUDIO, and VIDEO PROCESSING Using MATLAB®*. First edition, Boca Raton: CRC Press, 2021. doi: 10.1201/9781003019718.

[15] P. Flores-Vidal, J. Castro, y D. Gómez. Pixel-by-pixel classification of edges with machine learning techniques. en *Machine Learning, Multi Agent and Cyber Physical Systems*, WORLD SCIENTIFIC, ene. 2023, pp. 218–227. doi: 10.1142/9789811269264_0026.

IDEAS EN CIENCIAS DE LA INGENIERÍA

Hacia un Modelo Gráfico Causal para el Análisis de la Vaginosis Bacteriana

Towards a Causal Graphical Model for the Analysis of Bacterial Vaginosis

García-Avalos, M.[1], Canul-Reich, J.[1]*, Rodríguez-Henríquez, L.M.[2], De la Cruz-Hernández, E.[1]
(*juana.canul@ujat.mx)

Recibido: 2024-10-30 Aceptado: 2024-11-22

Resumen

El Análisis de Trayectorias (*AT*) es una técnica de Estadística Multivariante que estudia los efectos directos (causales) e indirectos de un conjunto de variables observables a través del análisis de correlaciones. Este estudio aplica *AT* a un conjunto de datos sobre la Vaginosis Bacteriana (*VB*) en mujeres embarazadas, el cual tiene tres tipos de diagnósticos. Aunque se han realizado investigaciones sobre las bacterias asociadas a la *VB* utilizando métodos como selección de variables, agrupaciones y reglas de asociación, el *AT* no ha sido utilizado como herramienta visual para detectar las variables coexistentes que varían de una mujer a otra. Se presenta un Modelo Gráfico Causal (*MGC*) que se basa en un modelo teórico derivado de la matriz de correlación de Pearson y métricas estadísticas. Este modelo muestra las variables que influyen en el diagnóstico de *VB,* identificando las bacterias relevantes: Mycoplasma Hominis (*Mh*), Atopobium Vaginae (*Av*), Gardnerella Vaginalis (*Gv*), Megasphaera Tipo 1 (*MT1*) y Bacteria Asociada a Vaginosis Bacteriana Tipo 2 (*BVBA2*). La validez biológica del modelo fue evaluada por un experto en biología. El análisis se llevó a cabo utilizando el software R.

Palabras clave: Vaginosis Bacteriana, Mujeres Embarazadas, Métricas Estadísticas, Correlación, Análisis de Trayectoria.

Abstract

The Trajectory Analysis (*AT)* is a Multivariate Statistical technique that studies the direct (causal) and indirect effects of a set of observable variables through correlation analysis. This study applies *AT* to a dataset on Bacterial Vaginosis (*VB*) in pregnant women, which includes three types of diagnoses. Although research has been conducted on the bacteria associated with *VB* using methods such as variable selection, clustering, and association rules, *AT* has not been utilized as a visual tool to detect

[1]Universidad Juárez Autónoma de Tabasco, Carr. Cunduacán a Jalpa de Méndez, KM 1, C.P. 86690.
[2]Instituto Nacional Astrofísica Óptica y Electrónica, Santa María Tonantzintla, Puebla, México, C.P. 72840.
[3]Universidad Juárez Autónoma de Tabasco, Ranchería sur 4ta secc., Comalcalco Tabasco, C.P. 86650.

coexisting variables that vary from one woman to another. A Causal Graphical Model (*MGC*) is presented, based on a theoretical model derived from the Pearson correlation matrix and statistical metrics. This model illustrates the variables influencing the *VB* diagnosis, identifying the relevant bacteria: Mycoplasma Hominis (*Mh*), Atopobium Vaginae (*Av*), Gardnerella Vaginalis (*Gv*), Megasphaera Type 1 (*MT1*), and Bacterial Vaginosis Associated Bacteria Type 2 (*BVBA2*). The biological validity of the model was assessed by a biology expert. The analysis was conducted using R software.

Keywords:Bacterial Vaginosis, Pregnant Women, Statistical Metrics, Correlation, Path Analysis.

Introducción

La investigación científica en medicina es importante para mejorar el diagnóstico y tratamiento de enfermedades. Se divide en tres tipos: investigación biomédica, clínica y salud pública [1]. La *VB* es común en mujeres en edad fértil, especialmente en las sexualmente activas, y puede ser asintomática o sintomática [2]. El síntoma más destacado en las mujeres es el flujo vaginal anormal, que puede variar en color (gris o verde) y presenta un olor característico a pescado [3]. En biología, se ha observado que múltiples bacterias son responsables de la aparición de la *VB*, y que la composición de estas bacterias puede variar entre mujeres. Este estudio tiene como objetivo identificar las variables coexistentes que influyen en el diagnóstico de la *VB*. Esta enfermedad en el embarazo, influenciada por cambios hormonales, aumenta el riesgo de complicaciones como parto prematuro y endometritis puerperal, por lo que requiere monitoreo y tratamiento [4]. La *VB* puede aumentar la susceptibilidad a infecciones de transmisión sexual, como el Virus del Papiloma Humano (*VPH*). Diversos estudios, tanto transversales como longitudinales, han investigado la relación; sin embargo, no han encontrado asociaciones significativas en estudios transversales entre la *VB* y *VPH* [3].

En el estudio de la *VB* en mujeres sexualmente activas se han utilizado diversos enfoques metodológicos. Uno de ellos es la selección de atributos [5], que identifica las variables más relevantes para predecir un diagnóstico positivo. Otro enfoque es el análisis de agrupaciones [6], que determina grupos de bacterias asociadas con el diagnóstico, observando variaciones entre estos grupos. Además, se han aplicado reglas de asociación para identificar factores coexistentes que contribuyen al estado positivo de la *VB* [7]. Estos métodos buscan aclarar las variables bacterianas que influyen en el diagnóstico de esta condición. Los autores [8] consideran que las tecnologías pueden mejorar los sistemas de salud al analizar grandes datos clínicos para crear modelos predictivos, de tamizaje y diagnóstico. Los gráficos estadísticos [9] son esenciales en el análisis de datos biomédicos, ya que muestran relaciones entre variables. Por otro lado, los Modelos de Ecuaciones Estructurales (*MEE)* permiten comprender las relaciones entre variables observables y no observables. El *AT*, derivado del MEE, examina los efectos directos (causales) e indirectos de las variables observables y, mediante un *MGC,* analiza las interacciones entre múltiples variables del sistema [10]. En este trabajo se utilizaron datos de mujeres embarazadas de zonas rurales y urbanas que acudieron a una revisión ginecológica y prueba de diagnóstico de *VB.* Se aplicó el *AT* a estos datos para obtener un *MGC* que ayude a visualizar las variables coexistentes que influyen en el diagnóstico de *VB*.

https://doi.org/10.36677/rici.v3i1.25016 Enero-Julio 2025 ISSN: 2992-7447 págs. 22-34

MATERIALES Y MÉTODOS

Matriz de correlación

Una matriz de correlación es una representación tabular que contiene los coeficientes de correlación entre distintas variables que indica la fuerza y dirección de la relación entre pares de variables. La correlación oscila entre -1 y 1, [11] donde:

- Un valor de 1 indica una correlación positiva perfecta, lo que significa que las variables tienden a aumentar juntas.
- Un valor de -1 indica una correlación negativa perfecta lo que significa que una variable tiende a aumentar, mientras que la otra tiende a disminuir.
- Un valor de 0 indica que no existe correlación entre las variables.

Modelo de Ecuaciones Estructurales

Los *MEE* son técnicas estadísticas multivariantes para analizar la relación de dependencia o causalidad entre variables observables y no observables. La Figura 1, presenta la nomenclatura utilizada para representar gráficamente un *MEE* [12].

X	Rectángulo o Cuadrado: significa que la variable X es observable
η	Círculo o Elipse: significa que la variable η es latente
ε, η, Y	La variable Y es un indicador de la variable η que tiene un término de error o perturbación ε
ξ_1, ξ_2	La curva con doble flecha significa asociación entre las dos variables (covariación) ξ_1 y ξ_2
η_1, η_2	La presencia de dos flechas simples significa que las variables tienen una causación recíproca

Figura 1: *Nomenclatura de Modelo de Ecuaciones Estructurales.*

Existen tres categorías de *MEE*: Análisis de trayectoria, análisis factorial y modelo general de ecuaciones estructurales [12, 13]. El *AT* examina los efectos causales e indirectos de un grupo de variables observables sobre otras, analizando sus correlaciones. Este análisis se representa gráficamente mediante el *MGC*, utilizando símbolos y convenciones matemáticas.

Métricas estadísticas de MEE

En estadística, los índices de ajuste (métricas) se utilizan para evaluar el rendimiento de los modelos con datos observados. Los índices miden la calidad del modelo y ayudan a determinar si el modelo describe o predice adecuadamente los datos. La Tabla 1 muestra algunas de las métricas comunes utilizadas en *MEE*, en particular para *AT* [10, 14], lo que permite verificar si el modelo teórico propuesto representa adecuadamente la estructura subyacente en los datos.

Los índices *RMSEA* y *SRMR* se conocen como índices absolutos de ajuste y revelan hasta qué punto el modelo representado en la matriz de covarianzas coincide con el modelo implícito. Cuanto menor sea el resultado, mejor será la calidad del ajuste del modelo. Por otro lado, *CFI*, *NFI* y *GFI*, se denominan índices de ajuste incrementales y evalúan cómo mejora un modelo propuesto en comparación con un modelo base [14].

Tabla 1: Métricas Estadísticas

Abreviaturas	**Significado**	**Criterio**
RMSEA	Error cuadrático medio de aproximación	< 0.05
SRMR	Raíz cuadrática media normalizada residual	< 0.05
GFI	Índice de bondad de ajuste	≥ 0.95
NFI	Índice de ajuste normalizado	≥ 0.95
CFI	Índice de ajuste comparativo	≥ 0.95

Estudio Experimental

En este estudio se utilizó el lenguaje de programación *R*, que es un software gratuito de análisis estadístico [15], que se ha consolidado como la plataforma estadística utilizada para el análisis de datos. Para el desarrollo del estudio, se inició con la exploración de la estructura de los datos, lo que implicó la eliminación de variables no relacionadas con la *VB*. Se abordaron también los valores atípicos, que se encontraban fuera del rango normal de cada variable. A continuación, se realizó la imputación de valores, reemplazando los datos faltantes mediante métodos estadísticos. Posteriormente, se calculó la matriz de correlación, obteniendo los coeficientes de correlación entre las variables. Se diseñó un *MGC* que ilustra las relaciones entre las variables, basado en la matriz de correlación, el modelo teórico y las métricas estadísticas. Finalmente, el *MGC* fue validado por personal del laboratorio con experiencia clínica en el área, asegurando su precisión y relevancia.

Conjunto de datos VB

El conjunto de datos se construyó con información de mujeres embarazadas de comunidades rurales y urbanas del estado de Tabasco, sureste de México. Las pacientes se sometieron a un estudio de diagnóstico de *VB*. Este conjunto de datos incluye información sociodemográfica, bacterias, el diagnóstico de *VB*, entre otros. La recopilación de los datos se realizó durante las campañas de embarazo saludable (agosto 2018-enero 2020), llevadas a cabo por un investigador biólogo de la División Académica Multidisciplinaria de Comalcalco (*DAMC*) de la Universidad Juárez Autónoma de Tabasco (*UJAT*).

Estructura del conjunto de datos

El análisis exploratorio de datos es un enfoque que utiliza estadísticas descriptivas y herramientas gráficas para obtener una comprensión profunda de un conjunto de datos. Su objetivo es identificar las características del conjunto, como la presencia de valores perdidos y atípicos, los cuales deben ser abordados para mejorar la calidad y la validez de los resultados obtenidos en el análisis posterior [16]. En el análisis exploratorio de datos se identificaron 87 atributos y 132 instancias, de los cuales 41 eran atributos categóricos y 46 atributos numéricos. El atributo clase (Diagnóstico de Vaginosis Bacteriana sin Mycoplasma Hominis, *DxVBNoMh*) se dividió en tres categorías: Vaginosis Bacteriana Positivo (*VB+*) con 32 instancias, que indica la presencia de la condición; Indeterminado con 5 instancias, para casos sin un diagnóstico claro; y Microbiota Normal (*VB-*) con 95 instancias, que señala la ausencia de la condición. Como se observa, el conjunto de datos está desbalanceado con VB- como clase mayoritaria. El impacto del desbalanceo de clases no forma parte de este estudio, sin embargo, la investigación está en curso y ello corresponde a una etapa posterior. Se registraron 609 valores faltantes y 143 valores atípicos en el conjunto de datos.

https://doi.org/10.36677/rici.v3i1.25016 Enero-Julio 2025 ISSN: 2992-7447 págs. 22-34

Eliminación de atributos

Durante el análisis se identificaron variables que no están relacionadas con la *VB*, esto se verificó con el área clínico-biológica quien fue creador de los datos. Al tener este conocimiento se decidió eliminar tales variables, en particular se descartó todas las variables relacionadas con el *VPH*. Así, de los 87 atributos originales, se obtuvo 72 variables con 132 observaciones.

Valores atípicos

El diagrama de caja fue desarrollado por Tukey en 1977 como una herramienta para el análisis exploratorio de datos. Se utiliza para resumir y comparar distribuciones basándose en sus valores extremos (máximo y mínimo), así como en la mediana y los cuartiles. Este gráfico también se conoce como diagrama de caja y bigotes [17]. Se implementó la técnica estadística en el conjunto de datos para identificar los valores atípicos, resultando en la detección de 143 valores atípicos. A pesar de que estos datos se encuentran fuera del rango normal en las variables, corresponden a pacientes que presentan otras enfermedades. Tras consulta con un biólogo experto en el área, se determinó que era pertinente mantener estos valores originales en el conjunto de datos. En la Figura 2, se muestran diagramas de cajas de algunos atributos del conjunto de datos. En la gráfica, se observa la mediana de estas variables y se resalta que esta técnica permite identificar los valores atípicos, los cuales se extienden más allá de la longitud del bigote, como en los casos de los atributos: InicioSexual, N.EMBARAZO y N.DEHIJOS.

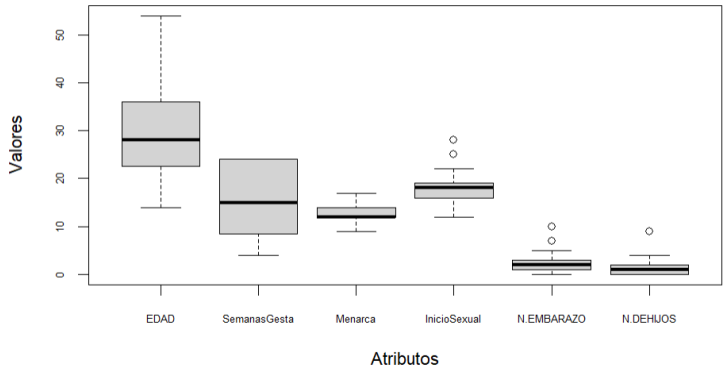


Figura 2: *Diagrama de caja de los primeros atributos del conjunto de datos.*

Valores faltantes

La imputación de datos faltantes es el proceso que utiliza la información disponible en la muestra para asignar un valor a las variables con datos ausentes. Su objetivo es crear un conjunto de datos completo para el análisis que se desea realizar. La imputación por media, que consiste en utilizar el promedio de una variable, es una técnica estadística aplicada para manejar valores faltantes en variables numéricas [18]. Por otro lado, la imputación por moda, que se refiere a la frecuencia más común de una variable, es una técnica estadística empleada para abordar valores ausentes en variables categóricas [19]. En el análisis se detectaron 609 datos perdidos distribuidos en 47 variables, que correspondían a atributos tanto categóricos como numéricos. Para imputar los datos faltantes en las variables categóricas utilizando la moda, primero se identificaron los valores faltantes de cada atributo y se determinó el diagnóstico (*VB*+, Indeterminado o *VB*-) correspondiente. Luego, se agruparon las instancias completas de cada atributo según el diagnóstico, se calculó la moda de cada grupo y, finalmente, se reemplazaron los datos ausentes

por las modas asociadas a su diagnóstico. Por ejemplo, en la Tabla 2 se muestran los datos imputados de las primeras 10 instancias de la variable uso de anticonceptivo (Anticoncepción). La columna *DxVBNoMh* indica el diagnóstico de *VB*, donde 1 representa *VB*+, 2 indica indeterminado y 3 corresponde a *VB*-.

Tabla 2: Datos imputados con la moda del atributo categórico de Anticoncepción y diagnóstico de *VB*

Instancias	**Datos originales**	**Datos imputados**	**DxVBNoMh**
1	NA	2	3
2	NA	2	1
3	NA	2	3
4	NA	2	3
5	2	2	3
6	NA	2	3
7	2	2	3
8	2	2	3
9	2	2	3
10	1	1	3

Para imputar los datos perdidos en las variables numéricas con la media, se siguió un proceso similar a los atributos categóricos, descrito en el párrafo anterior. La Tabla 3 presenta los datos imputados de las primeras 10 instancias del atributo inicio sexual (InicioSexual).

Tabla 3: Datos imputados con media del atributo numérico InicioSexual y diagnóstico de *VB*

Instancias	**Datos originales**	**Datos imputados**	**DxVBNoMh**
1	NA	18.0137	3
2	NA	17.5000	1
3	NA	18.0137	3
4	NA	18.0137	3
5	NA	18.0137	3
6	NA	18.0137	3
7	NA	18.0137	3
8	NA	18.0137	3
9	NA	18.0137	3
10	NA	18.0137	3

Matriz de correlación de Pearson

Para determinar las relaciones entre las variables y el atributo de clase (*DxVBNoMh*), dada la cantidad de atributos en el conjunto de datos y con el fin de visualizar la matriz de correlación con claridad, los datos se agruparon en bloques pequeños atendiendo el orden en que se encuentran en el conjunto de datos de manera secuencial y como criterio único mantener un tamaño equilibrado en cada grupo. Se

https://doi.org/10.36677/rici.v3i1.25016 Enero-Julio 2025 ISSN: 2992-7447 págs. 22-34

calculó la matriz de correlación de Pearson. La Figura 3 presenta las matrices de correlación de dos de los grupos formados. En la matriz de la izquierda se resaltan cuatro variables (*Av*, *Gv*, *MT1* y *BVAB2*) con coeficientes de correlación significativos, mientras que en la matriz de la derecha solo se observa una variable (*Mh*) relevante.

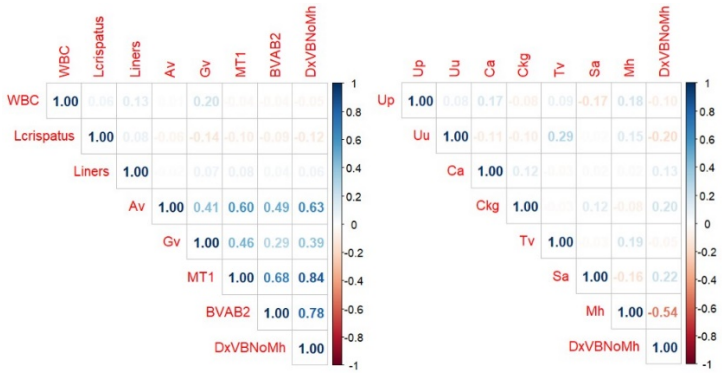


Figura 3: *Conjuntos de matrices de correlación.*

Después de calcular las matrices de correlación para cada grupo, se seleccionaron las variables con Çorrelaciones negativas ≤ -0.54 y positivas ≥ 0.39 respecto al atributo"de clase. Aunque los coeficientes de correlación son bajos, se consideran relevantes debido al contexto clínico y en nuestro caso con datos reales [20, 21] . Los coeficientes identificados fueron 0.39, 0.63, 0.78, 0.84 y -0.54. Se formó un grupo que incluye todas las variables que cumplen con estos criterios de coeficiente de correlación. La Figura 4 muestra la matriz de correlación triangular superior de las variables seleccionadas (*Mh*, *Av*, *Gv*, *MT1*, *BVAB2*) y el *DxVBNoMh*. La diagonal principal de la matriz, que va de la esquina superior izquierda a la esquina inferior derecha, está formada por unos, esto indica una correlación perfecta de una variable consigo misma.

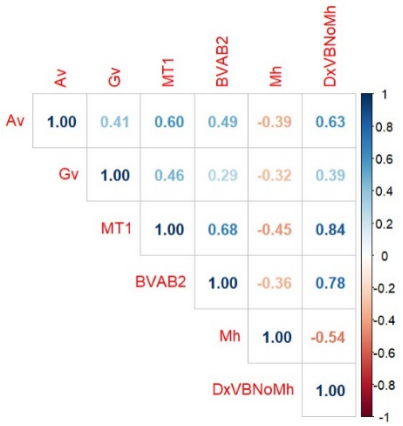


Figura 4: *Matriz de correlación de los atributos relevantes.*

Modelo Gráfico Causal

El desarrollo de un modelo gráfico causal requiere la determinación del modelo teórico basado en la matriz de correlación. Se inicia presentando modelos teóricos simples, que involucran únicamente una variable y el atributo de clase. Los modelos teóricos simples Ec. (1), Ec. (2), Ec. (3), Ec. (4) y Ec. (5) se representan en los diagramas a), b), c), d) y e), de la Figura 5, respectivamente.

$$Mod1 \leftarrow' DxVBNoMh \sim Av' \tag{1}$$

$$Mod2 \leftarrow' DxVBNoMh \sim Gv' \tag{2}$$

$$Mod3 \leftarrow' DxVBNoMh \sim MT1' \tag{3}$$

$$Mod4 \leftarrow' DxVBNoMh \sim BVAB2' \tag{4}$$

$$Mod5 \leftarrow' DxVBNoMh \sim Mh' \tag{5}$$

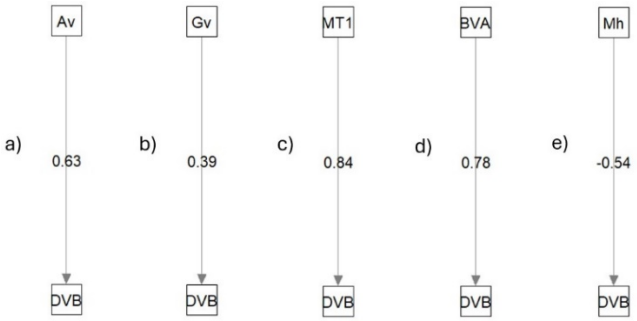


Figura 5: *Modelos gráficos causales simples. Abreviaturas:* BVA *representa* BVAB2 *y* DVB *representa* DxVBNoMh.

Para evaluar el ajuste de los *MGC* simples Ec. (1), Ec. (2), Ec. (3), Ec. (4) y Ec. (5), se calcularon las métricas estadísticas presentadas en la Tabla 1. Los valores obtenidos de las métricas para los cinco modelos teóricos son: $RMSEA = 0.000, SRMR = 0.000, GFI = 1.000, NFI = 1.000$ y $CFI = 1.000,$ todos dentro de los rangos esperados de acuerdo con la Tabla 1. Los coeficientes de trayectoria deben interpretarse de la siguiente manera: los valores cercanos a ±1 indican una fuerte asociación entre las variables o con el atributo, mientras que los valores cercanos a 0 reflejan una asociación débil. La Figura 5 muestra las relaciones causales entre las variables *Av*, *Gv*, *MT1*, *BVA* y *Mh* respecto al *DVB*, con los coeficientes de las trayectorias (0.63, 0.39, 0.84, 0.78 y -0.54) obtenidos a partir de los MGC simples. Estos coeficientes reflejan las relaciones según los modelos teóricos generados, y coinciden con los coeficientes de correlación de Pearson. Los coeficientes de trayectoria positivos indican que, a medida que aumentan las bacterias, también incrementa la probabilidad de obtener el *DVB*. En contraste, el coeficiente de trayectoria negativo indica que la reducción de las bacterias está asociada con un aumento en la probabilidad de tener el *DVB*.

https://doi.org/10.36677/rici.v3i1.25016 Enero-Julio 2025 ISSN: 2992-7447 págs. 22-34

Después de analizar los *MGC* simples, se aplica el mismo *AT* para desarrollar un *MGC* que incluye las variables relevantes junto con el atributo de clase, con el fin de visualizar las relaciones que se determinan al basarnos en la matriz de correlación de la Figura 4. La obtención del modelo teórico Ec. (6), se construyó de la siguiente forma: se comenzó considerando el *DxVBNoMh* en relación con las variables predictoras *Mh*, *MT1*, *BVAB2*, *Av* y *Gv*, es decir la primera relación. Luego, se crearon las demás relaciones basándose en la variable *MT1* (por estar más correlacionada con el diagnóstico de *VB*) y las otras variables restantes, respecto al atributo clase.

$$ModC \leftarrow' DxVBNoMh \sim Mh + Av + Gv + MT1 + BVAB2$$

$$MT1 \sim Mh$$

$$MT1 \sim Av$$

$$MT1 \sim Gv$$

$$MT1 \sim BVAB2' \tag{6}$$

A medida que se añadían relaciones, se calculaban las métricas estadísticas para verificar que se cumplían las condiciones enumeradas en Tabla 1. Los valores obtenidos de las métricas del modelo Ec. (6), que abarca todos los modelos teóricos, son: $RMSEA = 0.000$, $SRMR = 0.000$, $GFI = 1.000$, $NFI = 1.000$ y $CFI = 1.000$, estos valores indican que el modelo teórico está dentro de los rangos esperados de la Tabla 1. El modelo Ec. (6) permitió desarrollar el *MGC* mostrado en la Figura 6, que ilustra las relaciones causales o directas (líneas continuas) entre todas las variables y el *DVB*, así como las covariaciones (líneas discontinuas). Las relaciones causales desde las variables *Mh*, *Av*, *Gv*, *BVA* y *MT1* hacia *DVB* presentan coeficientes de -0.16, 0.14, -0.04, 0.35 y 0.46, respectivamente. Además, se identifican relaciones indirectas, como la conexión entre *BVA* y *DVB* a través de *MT1*, donde *MT1* actúa como variable intermedia, con coeficientes de 0.46.

Validación biológica del MGC

El *MGC* es una herramienta visual que ilustra las relaciones entre las variables y el *DVB*, facilitando la identificación de las bacterias coexistentes *Mh*, *Av*, *Gv*, *BVA* y *MT1* que influyen en el *DVB*, las cuales son frecuentes según la literatura [22, 23]. El *MGC* indica que la bacteria *Gv* tiene la más baja asociación causal con el *DVB*, con un coeficiente de -0.04. No obstante, *Gv* muestra una mejor relación con otras bacterias relacionadas con el *DVB*. Por otro lado, la bacteria *MT1* presenta una fuerte asociación causal con el *DVB*, con un coeficiente de 0.46.

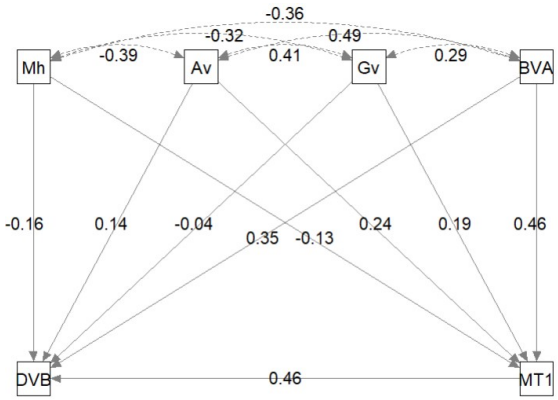


Figura 6: *Modelos gráficos causal de las variables relevantes.*

Resultados y discusión

En este estudio se aplicó el método de *AT* debido a la presencia de variables observables en el conjunto de datos de mujeres embarazadas con *DVB*, que incluye 87 variables y 132 observaciones. El atributo de clase abarca tres categorías: Vaginosis Bacteriana con 32 instancias, Indeterminado con 5 instancias y Microbiota Normal con 95 instancias. Se trabajó con 72 variables y 132 instancias (en la sección de valores faltantes se muestran las primeras 10 instancias). La matriz de correlación del conjunto permitió identificar relaciones significativas entre las variables y el atributo de clase, lo que facilitó la determinación de los modelos Ec. (1) a Ec. (6). Estos modelos fueron evaluados mediante métricas estadísticas, contribuyendo a la construcción de modelos gráficos causales ilustrados en las Figuras 5 y 6. Las variables obtenidas en el análisis de correlación son *MT1*, *BVA*, *Av*, *Gv* y *Mh*, las cuales se visualizan en los *MGC* con respecto a la condición. Además, se muestran las relaciones causales e indirectas entre las variables y el diagnóstico, así como los coeficientes de trayectorias. Los resultados presentados fueron validados biológicamente por el experto por lo que la interpretación del modelo desde el campo biológico permite ofrecer un modelo causal confiable construido con datos reales de pacientes. La investigación está en curso y el modelo aún es preliminar, por lo que el diseño experimental de la siguiente fase permitirá ampliar y dar mayor soporte para utilizar el *MGC* como un modelo causal que identifique variables detonantes de la *VB*.

Conclusiones

En este estudio, se utilizó el método de *AT* debido a que el conjunto de datos de mujeres embarazadas con el *DVB* consta de variables observables. Se trabajó con 72 variables y 132 instancias. El *AT* se aplicó a partir del análisis de correlación, lo que permitió identificar relaciones significativas con el atributo de clase y facilitar la formulación de modelos evaluados mediante métricas estadísticas. Las variables obtenidas en el análisis de correlación *MT1*, *BVA*, *Av*, *Gv* y *Mh* se visualizaron en los *MGC* en relación con el *DVB*. Además, se mostraron las relaciones causales e indirectas entre las variables y *DVB*, así como los coeficientes de trayectoria. A diferencia de los enfoques tradicionales, el *MGC* permite detectar tanto asociaciones fuertes como débiles, revelando interacciones que no son evidentes en otros modelos. Un hallazgo importante es que la literatura ha identificado a *MT1* como una bacteria fuertemente asociada al diagnóstico de *DVB*, lo cual es confirmado con el *MGC*. Además de corroborar lo que dice la literatura,

https://doi.org/10.36677/rici.v3i1.25016 Enero-Julio 2025 ISSN: 2992-7447 págs. 22-34

en el *MGC* se puede observar las correlaciones entre las bacterias mismas y no solo de las bacterias con el *DVB*. El *MGC,* como herramienta visual, podría facilitar en el futuro la identificación de las bacterias más influyentes en el desarrollo de la *VB*, lo que permitiría mejorar su diagnóstico y tratamiento. Adicionalmente, con la interpretación presentada del *MGC*, se sientan las bases para aplicarlo en el análisis causal de la *DVB* y para otras enfermedades.

Declaración de contribución de autores y colaboradores

Maricela García: Ejecución de experimentos y redacción. **Juana Canul**: Diseño experimental y supervisión. **Lil María Rodríguez:** Diseño experimental y revisión. **Erick De la Cruz:** Validación biológica de los hallazgos.

Referencias

[1] D. Mendoza, A. Quintana, y A. Díaz. La importancia de la investigación científica en medicina. *Revista Médica Sinergia, Atención Familiar UNAM*, vol. 24, no. 3, pp. 224–227, 2023. DOI: 10.22201/fm.14058871p.2023.3.85789.

[2] C. Celeste *et al.* Ethnic disparity in diagnosing asymptomatic bacterial vaginosis using machine learning. *npj Digital Medicine*, vol. 6, no. 211, 2023. DOI: 10.1038/s41746-023-00953-1.

[3] E. Sánchez *et al.* Molecular epidemiology of bacterial vaginosis and its association with genital micro-organisms in asymptomatic women. *Journal of Medical Microbiology*, vol. 68, no. 9, pp. 1373–1382, 2019. DOI: 10.1099/jmm.0.001044.

[4] I. Morelli y S. Gamboa. Vaginosis bacteriana en el embarazo: últimos avances hasta la fecha. *Revista Médica Sinergia*, vol. 7, no. 7, pp. e838, 2022. DOI: 10.31434/rms.v7i7.838.

[5] J. Pérez, J. Canul, J. Hernández, y B. Hernández. Selección de predictores para el diagnóstico de la vaginosis bacteriana utilizando árboles de decisión y algoritmos de relief. *Applied Sciences*, vol. 10, no. 9, pp. 3291, 2020. DOI: 10.3390/app10093291.

[6] H. Hernández, J. Canul, B. Hernández, and E. De la Cruz. An agglomerative hierarchical clustering approach to identify coexisting bacteria in groups of bacterial vaginosis patients. *Intelligent Data Analysis*, vol. 27, no. 3, pp. 583–611, 2023. DOI: 10.3233/IDA-216488.

[7] F. De la Cruz, J. Canul, R. Rivera, and E. De la Cruz. Impact of data balancing a multiclass dataset before the creation of association rules to study bacterial vaginosis. *Intelligent Medicine, Elsevier*, vol. 4, no. 3, pp. 188–199, 2023. DOI: 10.1016/j.imed.2023.02.001.

[8] M. Álvarez, L. Quirós, y M. Cortés. Inteligencia artificial y aprendizaje automático en medicina. *Revista Médica Sinergia*, vol. 5, no. 8, pp. 1–12, 2020. DOI: 10.31434/rms.v5i8.557.

[9] N. Sagaró y L. Zamora. Métodos gráficos en la investigación biomédica de causalidad. *Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta*, vol. 44, no. 4, 2019. Disponible en: `http://revzoilomarinello.sld.cu/index.php/zmv/article/view/1846`.

[10] A. Lepera. Introducción a los Modelos de Ecuaciones Estructurales y su Implementación en R mediante un ejemplo. *Revista de Investigación en Modelos Matemáticos Aplicados a la Gestión y la Economía*, vol. 1, no. 8, pp. 15–37, 2021. Disponible en: https://www.economicas.uba.ar/wp-content/uploads/2016/04/Lepera-Andrea-1.pdf.

[11] M. Molinas, C. Ochoa, y E. Ortega. Correlación, Modelos de regresión, Evidencia Pediátrica. vol. 17, no. 2, pp. 17–25, 2021. Disponible en: https://evidenciasenpediatria.es/articulo/7827/correlacion-modelos-de-regresion.

[12] J. Iglesias. Modelo de Ecuaciones Estructurales. Universidad de Almería, pp. 1–72, 2021. Disponible en: https://repositorio.ual.es/bitstream/handle/10835/13177/IGLESIAS%20LABRACA,%20JUAN%20MIGUEL.pdf?sequence=1.

[13] K. Barbeau, K. Boileau, F. Sarr, and K. Smith. Path analysis in Mplus: A tutorial using a conceptual model of psychological and behavioral antecedents of bulimic symptoms in young adults. *The Quantitative Methods for Psychology*, vol. 15, no. 1, pp. 38–53, 2019. DOI: 10.20982/tqmp.15.1.p038.

[14] F. Jordan. Valor de corte de los índices de ajuste en el análisis factorial confirmatorio. *Revista de Investigación en Psicología Social, Redalyc*, vol. 7, no. 1, pp. 66–71, 2021. Disponible en: https://www.redalyc.org/articulo.oa?id=672371335005.

[15] F. Jahuey, J. Herrera, y F. Paredes. El programa R: una estrategia inicial para su entendimiento y aprendizaje. *Revista Digital Universitaria*, vol. 23, no. 4, 2022. DOI: 10.22201/cuaieed.16076079e.2022.23.4.4.

[16] C. Bouza. Análisis exploratorio de datos univariados para la ciencia de datos. *Red Iberoamericana de Estudios Cuantitativos Aplicados*, 2023. DOI: 10.13140/RG.2.2.26038.06722.

[17] A. Madrid, S. Valenzuela, C. Batanero, y J. Garzón. Interpretación del diagrama de caja por estudiantes universitarios de ciencias de la actividad física y deporte. *Educación Matemática*, vol. 34, no. 3, 2022. DOI: 10.24844/EM3403.10.

[18] A. Contreras, N. Luna, P. Macías, y N. González. Imputación de datos faltantes en encuestas en hogares. *Memorias de la Décima Cuarta Conferencia Iberoamericana de Complejidad, Informática y Cibernética*, pp. 310–316, 2024. DOI: 10.54808/CICIC2024.01.310.

[19] E. Mejía, J. Aguilar, P. Bolaños, y J. López. Métodos de imputación por regresión, imputación por moda, imputación múltiple y árboles de decisión para variables categóricas en perspectiva comparada. *Revista De La Universidad Del Zulia*, vol. 15, no. 43, pp. 541–555, 2024. DOI: 10.46925//rdluz.43.29.

[20] L. Poaquiza y G. Gavilanes. El estrés y su relación con la inteligencia emocional en docentes universitarios. *Ciencia Latina Revista Científica Multidisciplinar*, vol. 6, no. 3, pp. 1357–1372, 2022. DOI: 10.37811/cl_rcm.v6i3.2301.

https://doi.org/10.36677/rici.v3i1.25016 Enero-Julio 2025 ISSN: 2992-7447 págs. 22-34

[21] R. Ponce y M. Cáceres. Dependencia emocional y su relación con el riesgo suicida en adultos jóvenes. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, vol. 4, no. 1, pp. 329–339, 2023. DOI: 10.56712/latam.v4i1.247.

[22] R. Drew *et al.* An interpretation algorithm for molecular diagnosis of bacterial vaginosis in a maternity hospital using machine learning: proof-of-concept study. *Diagnostic Microbiology and Infectious Disease*, vol. 96, no. 2, pp. 114950, 2020. DOI: 10.1016/j.diagmicrobio.2019.114950.

[23] J. Salas, L. Angulo, y E. Garita. Vaginosis Bacteriana-Actualización y novedad terapéutica. *Revista Ciencia y Salud, UCIMED*, vol. 5, no. 6, pp. 85–93, 2022. DOI: 10.34192/cienciaysalud.v5i6.387.

IDEAS EN CIENCIAS DE LA INGENIERÍA

Evaluación y calibración de la calidad de vida en megaciudades inteligentes: Caso de estudio en La Condesa, Ciudad de México

Evaluation and Calibration of Quality of Life in Smart Megacities: A Case Study in La Condesa, Mexico City

Vargas-Solar, G.[1]*, Castillo-Camporro, S.[2], Arias-Guzmán, E.J.[3],
(*genoveva.vargas-solar@cnrs.fr)

Recibido: 2024-11-06 Aceptado: 2024-11-27

Resumen

El artículo explora cómo el urbanismo contemporáneo vincula las características de los entornos urbanos con el rendimiento social, económico, ambiental y el bienestar ciudadano. Analiza la calidad de vida (QoL) en áreas urbanas considerando factores como el costo de vida, acceso a servicios de salud y educación, y la define como un índice multidimensional que mide el bienestar general. El trabajo propone una plataforma para calibrar la calidad de vida (QoL) a través de un índice que combina variables de naturaleza social, económica, cultural, psicológica y de salud, proporcionando una visión integral del bienestar. Además, el índice se utiliza para introducir una medida innovadora denominada «elasticidad de la calidad de vida» (E-QoL). La E-QoL mide hasta qué punto dichas variables pueden ajustarse sin comprometer una calidad de vida aceptable. Esta medida ofrece una herramienta útil para identificar límites críticos en la adaptación de políticas públicas que afectan al bienestar de las personas. Para evaluar esta metodología, se estudia el caso de la colonia La Condesa en Ciudad de México, analizando el impacto de políticas públicas en la QoL de los habitantes, con énfasis en problemas como la segregación, la sobrepoblación y la desertificación.

Palabras clave: cómputo urbano, ciencia de datos, calidad de vida, políticas públicas guiadas por datos.

[1]CNRS, Univ. Lyon, INSA Lyon, CPE, UCBL, LIRIS, UMR 5205, Lyon France.
[2]Universidad Nacional Autónoma de México, Facultad de Estudios Superiores, Estudios de Posgrados, Acatlán, México.
[3]Universidad Nacional Autónoma de México, Facultad de Estudios Superiores, Secretaría General Académica, Acatlán, México.

https://doi.org/10.36677/rici.v3i1.25065 Enero-Julio 2025 ISSN: 2992-7447 págs. 35-47

Abstract

The article examines how contemporary urban planning connects urban characteristics to social, economic, environmental performance, and citizen well-being. It defines quality of life (QoL) as a multidimensional index incorporating social, economic, cultural, psychological, and health variables. The work introduces the .[el]asticity of quality of life"(E-QoL), a measure evaluating how adjustments to QoL variables affect the overall index without compromising well-being. This tool helps identify critical policy limits. Using the La Condesa neighborhood in Mexico City as a case study, the article analyzes public policy impacts on QoL, focusing on segregation, overpopulation, and desertification.

Keywords: urban computing, data science, quality of life, data-driven public policies.

Introducción

Las megalópolis presentan desafíos significativos para los gobiernos, como la segregación social y la sobrepoblación. Diversas organizaciones internacionales como la ONU (Organización de las Naciones Unidas), la *UNESCO* (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura) y el BID (Banco Interamericano de Desarrollo) planifican proyectos urbanos para abordar estos problemas, impulsando la economía del territorio para integrar a los residentes en el desarrollo urbano. También existen políticas de transferencia [12] que permiten aprender de procesos exitosos en otras áreas, aunque en los últimos años, las propuestas urbanas con apoyo estatal han mostrado consecuencias negativas para los residentes [28][29][30][31]. Los gobiernos deben adoptar mecanismos que incorporen proyectos exitosos de otras regiones o que enfrenten problemas como la segregación y la sobrepoblación mediante proyectos que impulsen la economía local. Estas herramientas deben evaluar la vida diaria y el bienestar de los residentes para mejorar su desarrollo y el entorno urbano. El índice de Calidad de Vida (QoL – *Quality of Life*) proporciona una evaluación actualizada que facilita estrategias para áreas que requieren intervención. La Organización Mundial de la Salud define la QoL como “La percepción de un individuo de su posición en la vida dentro del contexto cultural y de valores en que vive y respecto a sus objetivos, expectativas, estándares y preocupaciones” [35]. Cuantitativamente, se evalúa a través de aspectos subjetivos como salud física, bienestar psicológico y contacto social, y aspectos objetivos como acceso al trabajo, costo por m² y calidad del aire.

Este artículo propone una plataforma de bienestar para evaluar cómo ajustar variables de QoL sin afectar significativamente el bienestar a los habitantes. Analiza la QoL en áreas urbanas considerando factores como costo de vida y acceso a servicios, y la define como un índice multidimensional. La metodología propone el concepto de «elasticidad de la calidad de vida» (E-QoL – *Elasticity of Quality of Life*) [35], una medida que evalúa la relación entre las modificaciones de las variables que contribuyen a la calidad de vida y el impacto de estas modificaciones en el índice de QoL[30]. En esencia, mide la capacidad de ajustar estas variables sin comprometer una calidad de vida aceptable, proporcionando un marco para analizar los límites críticos en los que los cambios en políticas o condiciones sociales pueden mantener el bienestar general de las personas. Para evaluar la metodología, se estudia la colonia de La Condesa en Ciudad de México, analizando el impacto de políticas públicas en la QoL, con foco en segregación, sobrepoblación y desertificación.

El artículo se organiza de la siguiente forma. Primero, se revisan estudios sobre enfoques para definir y medir la QoL en entornos urbanos. Se presenta la plataforma propuesta con un enfoque holístico para calibrar la QoL. Introducimos la elasticidad de la QoL (E-QoL) como método para ajustar métricas, evaluando impactos de políticas públicas en la QoL. La plataforma apoya a la toma de decisiones para evaluar impactos de políticas y efectos en la QoL de la E-QoL. El estudio es apoyado por un estudio de caso en La Condesa. Finalmente, se discuten resultados iniciales y líneas de investigación futuras.

TRABAJOS RELACIONADOS

Sőrés y K. Pető [7][8] destacan dos modelos de investigación sobre la QoL: el escandinavo, que prioriza factores objetivos relacionados con recursos y posesiones, y el estadounidense con un enfoque subjetivo. Erik Allardt [13][14], basado en Maslow [15][16], combinó ambos enfoques y estableció tres niveles: "Tener, amar, ser"(necesidades materiales, sociales y de desarrollo personal). Los indicadores objetivos incluyen ingresos, situación laboral, bienes de consumo y otros similares, considerando también el Producto Interno Bruto (PIB) per cápita y otros índices económicos. El bienestar subjetivo (SWB - *Subjective Well Being*) fue propuesto por Kahneman [17][18] como una medida de satisfacción con la vida, que se considera estable y abarca la vida familiar, laboral y el ocio [20][24]. Cada vez se estudia más cómo el SWB depende de factores contextuales como consumo, calidad escolar o menor estrés en desplazamientos, según Diener y Seligman [19], Diener y Suh y Guzmán Pozo [20][24]. QoL y SWB no son equivalentes, pues corresponden a diferentes conceptos teóricos, según Stewart-Brown [23]. Algunos estudios han relacionado la QoL subjetiva y el SWB en el modelo de Calidad de Vida y Bienestar (LQW – *Life Quality and Wellbeing*) [25], con una aparente dependencia lineal. Este modelo incluye 14 facetas de la QoL subjetiva basadas en seis dominios internacionales de QoL y, según sus autores, ofrece un enfoque integral validado en un estudio con 11 culturas de "las regiones más habitadas del mundo"[24].

La medición de la QoL en áreas urbanas ha sido estudiada por comisiones internacionales. El informe "¿Cómo va la vida?"[25] sugiere 11 grupos de indicadores cuantitativos y cualitativos (vivienda, ingresos, empleos, comunidad, educación, medio ambiente, compromiso cívico, salud, vida, seguridad, y equilibrio laboral-personal) que definen el bienestar y pueden medir la QoL. El Índice de Desarrollo Humano (IDH) [26] se basa en salud, educación y nivel de vida. La Organización para la Cooperación y el Desarrollo Económicos (OCDE - *Organisation for Economic Co-operation and Development*) y la ONU evalúan la QoL en países miembros para (i) determinar si las políticas públicas reducen desigualdades y (ii) medir el progreso general. El Parlamento Europeo, la OCDE y el Fondo Mundial para la Naturaleza (WWF - *World Wide Fund for Nature)* han creado índices que modelan la vida diaria, la pobreza y la desigualdad en zonas específicas. Un estudio del Instituto Nacional de Estadística de España propuso un índice de QoL basado en investigaciones como Eurostat, que incorpora medidas cualitativas. El Instituto Nacional de Estadística (ISTAT) definió 12 indicadores para evaluar los avances que incluían aspectos económicos, sociales y ambientales[1]. La Comisión para la Medición del Desarrollo Económico y el Progreso Social (CMPEPS), creada en 2008 en Francia, definió nuevas medidas para evaluar el progreso social en Stiglitz-Sen-Fitoussi [3]. Esta guía es la base para índices de QoL en el mundo e incluyó un análisis multidimensional que sumó componentes de bienestar. La medida del Grupo de Trabajo de Expertos en Calidad de Vida abarca 9 indicadores: condiciones materiales de vida, actividad principal, educación,

[1]El Instituto Nacional de Estadística (ISTAT) definió 12 variables para evaluar el progreso que incluyeron aspectos económicos, sociales y ambientales, 5ª Edición, 2017.

https://doi.org/10.36677/rici.v3i1.25065 Enero-Julio 2025 ISSN: 2992-7447 págs. 35-47

salud, ocio, seguridad económica y física, gobernanza y derechos, entorno natural y experiencia de vida. Los datos provienen de encuestas a individuos y de fuentes[2] como la Encuesta sobre la Calidad de Vida (ECV - Enquête sur les Conditions de Vie) y la Encuesta sobre la Población Activa (EPA – Enquête sur la Population Active). Los índices de Hong Kong integran medidas personales, sociales, políticas, culturales, económicas y ambientales, con 21 indicadores divididos en grupos: social, económico y ambiental, añadiendo variables como libertad de prensa y estrés.

En América Latina, Chile, realiza un estudio de Calidad de Vida por parte del Ministerio de Salud (MISAL). La última encuesta nacional de calidad de vida y salud 2015-2016 (ENCAVI),[3] realizada por la Dirección de Estudios Sociales de Chile (DESUC), se realizó para determinar el estado de la sociedad para el diseño, desarrollo y evaluación de políticas públicas[4]. El proyecto Çalidad de vida en Argentina"[5] propone un enfoque de ranking del bienestar por departamentos. Identifica diferentes estrategias para medir la pobreza y la QoL. La pobreza es una medida de privación que incluye a quienes no alcanzan un umbral mínimo. La QoL se define en relación con un nivel económico óptimo, mientras la pobreza se mide con un valor mínimo, la QoL se mide respecto a un valor máximo. Para calcular el índice de QoL, el estudio emplea indicadores cuantitativos y cualitativos para modelar la satisfacción personal. Este índice combina 5 categorías y 20 indicadores definidos por variables socioeconómicas y ambientales. Los datos para estas medidas provienen de censos, fuentes estadísticas, imágenes satelitales y encuestas.

La Universidad Presbiteriana Mackenzie, a través de su Núcleo de Investigación en Calidad de Vida (NPQV) está elaborando el Índice Económico de Calidad de Vida (IEQV). Este índice utiliza otros indicadores además de los propuestos por el Índice de Desarrollo Humano (IDH) como el transporte, la contaminación visual y el ruido. Estas variables se ponderan con diferentes pesos cuando se combinan para definir el índice de QoL. Los datos utilizados para el cálculo de estas variables son recolectados por el Instituto Brasileño de Geografía y Estadística y la Encuesta Nacional de Muestreo de Viviendas. El IEQV tiene como objetivo lograr un análisis más profundo de los diversos elementos que afectan la evolución del desarrollo humano.

En México, el índice de bienestar denominado Índice Nacional de Calidad de Vida (INCAVI) propuesto por la Universidad de Monterrey 2011 utiliza siete medidas cada una dividida en diferentes variables de valores cualitativos y cuantitativos[6]: salud, educación, bienestar personal, buen gobierno, economía, vida colectiva y seguridad. El Instituto Nacional de Estadística y Geografía (INEGI) propuso el índice

[2]https://www.insee.fr/fr/information/4230346 visitado el 24 de Noviembre 2024

[3]Ministerio de Salud de Chile (2017), Encuesta de Calidad de Vida y Salud (ENCAVI) https://blog.desuc.cl/posts/2023-10-18-encavi-2023/ visitado el 04 de octubre 2024

[4]Ibídem.

[5]https://ciudadaniametropolitana.org.ar/2019/10/el-mapa-de-la-calidad-de-vida-donde-se-vive-mejor-y-peor-en-la-argentina visitado el 2 de noviembre 2024

[6]INCAVI (García Vega, J. de J 2011b) define las siguientes subcategorías. Para la salud (salubridad, número de visitas al médico, servicio médico justo); economía (¿los ingresos son suficientes para acceder a otras necesidades básicas además de la alimentación, el acceso a una vivienda justa, el acceso a un buen trabajo); educación (nivel académico de las escuelas, accesibilidad a una buena educación, accesibilidad a eventos culturales, deportivos y de ocio); seguridad (seguridad en la comunidad, víctima de inseguridad, capacidad de la policía para enfrentar la inseguridad); buen gobierno (honestidad, eficiencia, calidad de los servicios públicos); la vida colectiva (clima, calidad del medio ambiente, calidad de los servicios no gubernamentales, movilidad en la ciudad); bienestar (disponibilidad de tiempo libre, percepción de calidad de vida, disposición a pasar toda su vida en la comunidad, facilidad para establecer interacción social con familiares y amigos).

de bienestar subjetivo BIARE (autoinforme de bienestar[7]) para medir la forma en que las personas experimentan su propia QoL. Se basa en la medición de los indicadores subjetivos de bienestar y define las directrices de la OCDE.

El Paraíso de Michalos [27] reconoce que las personas en la misma zona pueden tener diferentes percepciones sobre las condiciones de vida. Michalos [27] sugiere una matriz que clasifica (i) el paraíso de los tontos, (ii) el paraíso real, (iii) el infierno real y (iv) el infierno de los tontos, donde la percepción de la vida depende de los residentes. Resalta la importancia de las condiciones en las que se aplican las encuestas para captar la percepción de QoL. La secuencia de preguntas y el área de trabajo son factores que influyen en las respuestas.

Discusión

Basado en la literatura existente, el estudio concluye que la QoL es un constructo medible con dos dimensiones, como lo proponen Sores y Peto [7][8]. La primera faceta se relaciona con el bienestar medido por indicadores objetivos como ingresos, salud, infraestructura y seguridad pública. La segunda aborda el bienestar subjetivo, incluyendo felicidad, satisfacción y realización emocional. Se ha estudiado la relación entre ingresos y salud, mostrando que ingresos más altos suelen mejorar la salud, y mejor salud puede elevar los ingresos por mayor productividad [7]. Esto explica por qué las iniciativas urbanas priorizan el crecimiento económico. También existe una relación bidireccional entre educación e ingresos, siendo la educación clave en la participación social [5][6].

Este artículo resalta la importancia del estatus socioeconómico en la percepción de la QoL y las expectativas de proyectos urbanos. Medidas cualitativas como felicidad y estado de ánimo capturan percepciones que pueden calibrarse por niveles socioeconómicos, demostrado en proyectos de Quercia, Schifanella y Aiello [4]. Es vital identificar fuentes adecuadas para definir el índice de QoL. En muchos países, esto depende de censos y encuestas hechas cada 5-10 años a segmentos específicos. Sin embargo, los datos pueden no ser representativos y los análisis a menudo se hacen mucho tiempo después de la recolección. Es crucial asegurar la calidad de los datos, considerando su procedencia, criterios de selección y fiabilidad. Las Tecnologías de la Información y la Comunicación (TIC) y la ciencia de datos presentan oportunidades para nuevos métodos de recolección de datos que implican participación ciudadana a través de redes sociales, juegos en línea, sensores y cámaras. Estas técnicas modernas complementan las encuestas tradicionales. Los datos pueden ser almacenados y actualizados con parámetros de calidad claros, permitiendo análisis automatizados y una interpretación más completa y representativa.

PLATAFORMA DE BIENESTAR: UNA PERSPECTIVA HOLÍSTICA PARA CALIBRAR LA CALIDAD DE VIDA

La percepción de los territorios urbanos está influenciada por variables sociales, económicas y psicológicas. Este artículo propone una plataforma digital (Figura 1) que permite seleccionar variables y personalizar su estructura matemática, ayudando a analistas y tomadores de decisiones a evaluar los efectos en la elasticidad al implementar programas y cambios urbanos. Esta plataforma identifica variables que definen la QoL conforme al marco de capacidades de Amartya Sen e incluye el utilitarismo para diseñar la elasticidad de la QoL (E-QoL), permitiendo observar cómo los individuos ajustan su comportamiento

[7]https://www.inegi.org.mx/investigacion/bienestar/basico/default.html

https://doi.org/10.36677/rici.v3i1.25065 Enero-Julio 2025 ISSN: 2992-7447 págs. 35-47

en busca de felicidad, apoyando programas gubernamentales efectivos. La plataforma permite obtener una visión integral de la QoL como un estado dinámico que refleja las percepciones de los residentes en un entorno cambiante a través de la E-QoL. La investigación explora cómo las estrategias urbanas y económicas impactan los estándares de vida y el índice de QoL, si han mejorado el bienestar y si la segregación, exclusión, cambios de uso de suelo, escasez de agua y sobrepoblación son sostenibles a largo plazo. Se propone un método cuantitativo basado en datos para medir la QoL en La Condesa, incluyendo la recopilación de datos de indicadores de bienestar y percepciones de QoL. Se utilizan datos del Instituto Nacional de Estadística y Geografía de México para calcular la E-QoL.

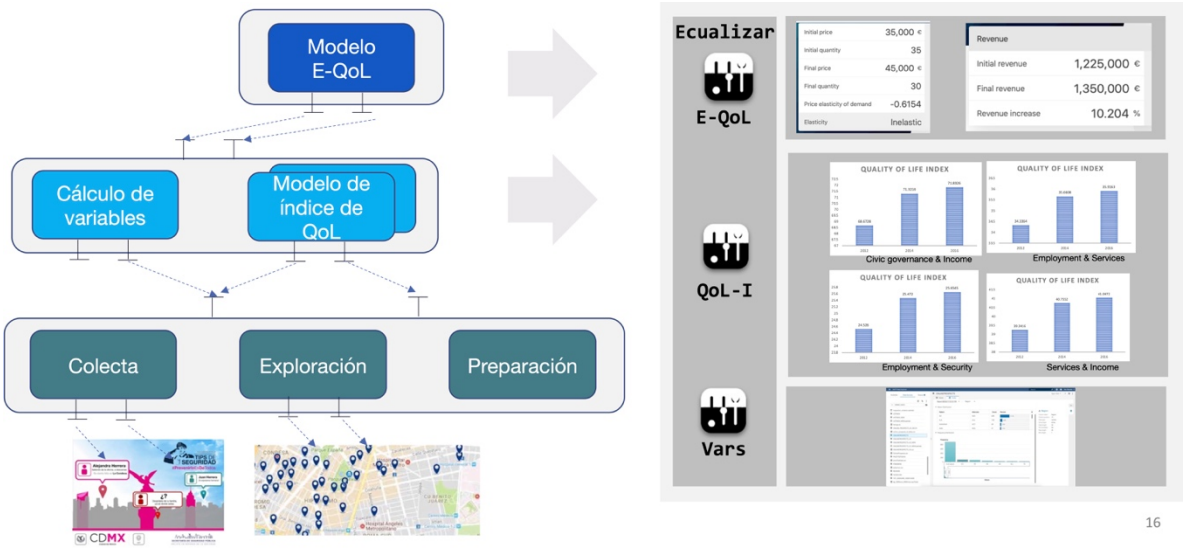


Figura 1: *Plataforma de bienestar para determinar E-QoL.*

Modelo cuantitativo para la calidad de vida: metodología

La calidad de vida puede representarse como un índice matemático que integra varias variables. La selección y combinación de estas variables se ven influenciadas por tendencias económicas y políticas. Nuestro estudio utiliza el modelo matemático de Stasys Puskoruis [8] que usa el modelo propuesto en [9]. Este modelo es la base del concepto de elasticidad urbana, ya que incluye tanto factores económicos como indicadores de calidad de vida humana. El índice de calidad de vida se expresa mediante la fórmula en la Ec. (1):

$$I = \sum_{i=1}^{10} a_i b_i \tag{1}$$

El resultado del índice es la suma de 10 indicadores ponderados. Los pesos utilizados para ponderar los indicadores están representados por los términos a_i.

b_1 - salud

b_3 - calidad del tiempo en el trabajo

b_4 - estatus de ingresos

b_5 - consumo
b_6 - medio ambiente y alojamiento
b_7 - educación de la población
b_8 - ley, seguridad, orden y niveles de corrupción de la población
b_9 - ética-moralidad, espiritualidad, valor de la cultura y ocio de la población
b_{10} - Indicador de igualdad de género de la población

El índice es el resultado de la suma de 10 indicadores ponderados. Los pesos utilizados para ponderar los indicadores están representados por los términos a_i. Por ejemplo, b_1 y el coeficiente de peso de ese indicador a_1. La clave en esta fórmula es cómo distribuir los pesos entre los indicadores para calibrar el índice.

Elasticidad de la calidad de vida (e-QoL)

La E-QoL mide hasta qué punto dichas variables pueden ajustarse sin comprometer una calidad de vida aceptable. Este estudio considera que la combinación de variables cualitativas y cuantitativas, recopiladas bajo diversas condiciones y protocolos, permite una evaluación más integral de la calidad de vida (QoL) al ajustar el peso de cada variable. Además, los modelos matemáticos existentes para medir el índice de QoL a menudo ignoran el "punto de no retorno", lo que implica que no se evalúa la viabilidad de alcanzar un índice de QoL satisfactorio al implementar políticas públicas en un territorio. El concepto de .[el]asticidad de la demanda.[en] economía puede incorporarse en los modelos de índice de QoL. En economía, la elasticidad se define como la relación entre el cambio porcentual en la cantidad demandada o suministrada y el cambio porcentual en el precio, expresada en la Ec. (2).

$$E_p = \frac{\Delta Quantity}{\Delta Price} = \frac{\frac{(P_1)+P_2}{2}}{\frac{(Q_1)+Q_2}{I_{12}}} \tag{2}$$

Este trabajo redefine el concepto de elasticidad en urbanismo para mostrar la relación entre la QoL de los habitantes y las variables que influyen en el bienestar. Originalmente de la microeconomía por Alfred R. Marshall, la elasticidad cuantifica cómo los cambios en una variable afectan a otra [33]. Mide la capacidad de respuesta de una variable dependiente Y (por ejemplo, Cantidad) ante cambios en una independiente X (por ejemplo, Precio). Se propone aplicar este concepto al índice de QoL, donde Y es el índice de QoL y X son variables que influyen en el comportamiento urbano. El objetivo es calcular valores que indiquen bienestar según cambios en estas variables por políticas públicas, mostrando impactos en la QoL, como la relación entre actividades nocturnas y seguridad, o el nivel de ruido que provoca estrés.

La E-QoL puede mostrar cómo varía la elasticidad tras implementar planes urbanos y sociales, modelando variables que toman valores específicos al aplicarse y varían según el contexto. Una herramienta de monitoreo basada en ciencia de datos podría destacar impactos previstos y no previstos en territorios urbanos, mostrando discrepancias entre los efectos esperados y reales en la QoL. Es clave entender las condiciones del cálculo del índice de QoL, las variables involucradas y los métodos de recopilación de datos usados para asignar estos valores.

https://doi.org/10.36677/rici.v3i1.25065 Enero-Julio 2025 ISSN: 2992-7447 págs. 35-47

Caso Experimental: La Condesa

Este estudio se centra en la colonia histórica de La Condesa, ubicado a 4 km del Centro Histórico de la Ciudad de México. Esta área ha sido objeto de programas gubernamentales (por ejemplo, el Bando 2) destinados a promover la construcción de vivienda social para atraer residentes, actividades de ocio y vida comercial a los espacios urbanos centrales. Sin embargo, el crecimiento económico resultante y el aumento de la población no parecen alinearse con las percepciones de calidad de vida y bienestar de los ciudadanos.

Colección de datos. En México, el INEGI proporciona estadísticas de los censos nacionales de diferentes años. El experimento usa doce conjuntos de datos para calcular el índice de calidad de vida (QoL), cada uno de los cuales contiene datos agregados usados por organizaciones internacionales para medir diversos indicadores. Estos conjuntos de datos se filtraron para enfocarse específicamente en el área de La Condesa de la Ciudad de México. La base de datos de indicadores de bienestar consta de 35 indicadores definidos por la OCDE, los cuales se utilizan para calcular el Índice de Vida Mejor basado en conceptos de bienestar y progreso. Estos indicadores se agrupan en 12 dimensiones: accesibilidad a servicios, comunidad (relaciones sociales), educación, equilibrio vida-trabajo, ingresos, medio ambiente, compromiso cívico y gobernanza, salud, satisfacción, seguridad, empleo y vivienda. Los datos recopilados abarcan los años 2010 a 2018. Para el análisis, se utilizan datos de indicadores agregados, sino datos detallados necesarios para calcular el índice de QoL presentado en la sección anterior.

Cálculo de Medidas Cuantitativas y Cualitativas. Los conjuntos de datos de indicadores de QoL exportados por INEGI se derivan de censos bianuales realizados desde 2010. Se observa que no todos los indicadores estaban presentes en cada conjunto de datos. El experimento se centró en calcular el índice de QoL en México para los años 2012, 2014 y 2016. Al analizar la distribución de los valores de los indicadores, se observa que algunos se presentan como porcentajes de la población, mientras que otros se mostraban como valores de intervalo o medidas ad hoc (por ejemplo, calidad del aire). Debido a la falta de datos sin procesar para algunos de estos indicadores, se excluyeron de los cálculos. Se optó por usar las medidas más consistentes para asegurar la precisión en los cálculos.

El experimento incluyó seis dimensiones: accesibilidad a los servicios, equilibrio vida-trabajo, ingresos, compromiso cívico y gobernanza, seguridad y empleo. Para cada dimensión, se seleccionaron subdimensiones expresadas como porcentajes del total de participantes del censo. Dado que estos censos son promovidos por el gobierno y considerados un deber cívico, una parte significativa de la población participa. Los datos proporcionan información sobre las percepciones de los ciudadanos respecto a los indicadores tras la implementación del programa Bando 2, observadas de 5 a 10 años después de su implementación. De acuerdo con la fórmula de QoL adoptada, se ponderaron los indicadores basándose en los conocimientos de colegas expertos en urbanismo. Este enfoque está alineado con el objetivo de nuestro estudio, que busca capturar la percepción de la población sobre su calidad de vida en la colonia de La Condesa. Se incorporaron N=12 medidas seleccionadas de los grupos de indicadores adoptados que integran aspectos de bienestar no económico como seguridad o accesibilidad a servicios con bienestar económico como ingresos. La accesibilidad a servicios e ingresos (a_1, a_2, a_4, a_5) se ponderaron con un valor de 0.08 considerando que el bien estar incluye el factor económico con la protección social; el equilibrio entre vida y trabajo (a_3) y el empleo (a_9 - a_{12}) con un valor de 0.04; y el compromiso cívico, la

gobernanza y la seguridad (a_6 - a_8) con un valor de 0.16. Estas ponderaciones se calibraron con el aporte de expertos en urbanismo, asignando un mayor valor a los aspectos derivados de políticas públicas, como los servicios, y un menor peso a los elementos de carácter intangible. Esta distribución refleja un modelo de calidad de vida (QoL) donde el factor económico tiene un papel predominante en la definición del bienestar. El proceso de calibración, realizado mediante programas computacionales, produjo varios resultados. Los resultados presentados aquí reflejan un cálculo representativo alineado con observaciones empíricas (no matemáticas) y evaluaciones de los impactos en La Condesa tras la implementación del programa Bando 2.

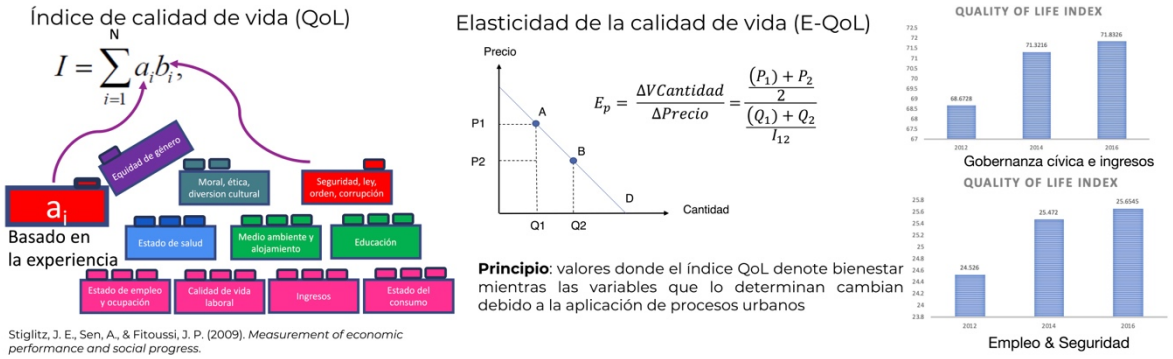


Figura 2: *Índices de QoL 2012 - 2016 en La Condesa.*

El índice de QoL indica que la percepción de la calidad de vida de los ciudadanos no es muy alta (ver Figura 2), lo que sugiere que, si bien la vida se considera aceptable, aún se necesitan mejoras en servicios, seguridad e ingresos a través de políticas públicas. Como se ilustra, la QoL mejoró entre 2012 y 2016, pero el cambio de 2014 a 2016 fue marginal. El análisis inicial del índice de QoL respalda la hipótesis de que se necesitan estrategias adicionales para su medición. La integración de indicadores tanto cuantitativos como cualitativos es un paso preliminar hacia una visión más completa de la QoL.

Alcances y lecciones aprendidas

La calibración de las ponderaciones de las variables produjo diversas evaluaciones del índice de QoL que fomentaron una comprensión holística del índice, aunque no se relacionan directamente con acciones específicas del programa de políticas públicas, Bando 2. Los esfuerzos actuales se centran en estudiar esta correlación mediante la recopilación de datos de redes sociales, encuestas dedicadas que involucren a residentes de largo plazo y temporales, cámaras recién instaladas en el vecindario y startups de transporte alternativo. Estas fuentes contribuirán con datos subjetivos utilizados en la literatura para medir el bienestar y la QoL. También se busca utilizar información de un proyecto del INEGI que mide el estado de ánimo de los usuarios de Twitter en la Ciudad de México. En colaboración con Twitter, el INEGI está innovando en soluciones basadas en datos para problemas urbanos. Considerar el estado de ánimo de los ciudadanos activos en las redes sociales—principalmente aquellos de clases económicas media y alta en La Condesa—puede proporcionar una medida relevante para nuestro estudio. Para comprender mejor la relación entre los cambios en las variables y las variaciones en el índice de QoL, se propuso analizar sus posibles dependencias lineales a través de un enfoque matemático, introduciendo el concepto de E-QoL.

https://doi.org/10.36677/rici.v3i1.25065 Enero-Julio 2025 ISSN: 2992-7447 págs. 35-47

Comprender el alcance teórico de la calidad de vida (QoL) y el bienestar, junto con las variables que los influyen en un área urbana específica, es crucial ya que cada espacio urbano puede caracterizarse por sus propias variables y definiciones de QoL.

Conclusión y Trabajo Futuro

Este artículo propuso una plataforma para estudiar y calibrar el bienestar ciudadano mediante el cálculo del índice de QoL con indicadores cuantitativos y cualitativos, junto con un modelo de calibrado E-QoL. El objetivo fue diseñar programas gubernamentales que revitalicen áreas urbanas y fomenten el crecimiento económico sin comprometer el bienestar ciudadano. En la colonia La Condesa, Ciudad de México, los datos entre 2012 y 2016 mostraron un aumento en la percepción de QoL, aunque afectaron de manera distinta a poblaciones transitorias, que experimentaron un bienestar temporal, y a residentes permanentes, quienes señalaron problemas como ruido, falta de estacionamiento y crimen.

Declaración de contribución de autores y colaboradores

Sagrario Castillo Camporro: Investigación, Conceptualización, Metodología, Experimentación. **Genoveva Vargas-Solar**: Redacción - Preparación del borrador original. **Ericka Arias Guzmán**: Supervisión. **Sagrario Castillo Camporro y Genoveva Vargas-Solar**: Redacción - Revisión y Edición.

Referencias

[1] R. Palomba. Calidad de Vida: Conceptos y medidas Concepto de calidad de vida. 2002, pp. 1–12.

[2] S. Puskorius. The Methodology of Calculation the Quality of Life Index. *International Journal of Information and Education Technology*, vol. 5, no. 2, pp. 156, 2015. DOI: 10.7763/IJIET.2015.V5.494.

[3] J. E. Stiglitz, A. Sen, y J.-P. Fitoussi. Measurement of economic performance and social progress. 2009. [Online]. Available: `http://bit.ly/JTwmG`. Accessed: June 26, 2012.

[4] D. Quercia, R. Schifanella, y L. M. Aiello. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. in *Proceedings of the 25th ACM conference on Hypertext and social media*, 2014, pp. 116–125. DOI: 10.1145/2631775.2631799.

[5] R. Putnam. The Prosperous Community: Social capital and public life. *The American Prospect*, vol. 4, no. 13, pp. 21, 1993.

[6] V. Pukeliene y V. Starkauskiene. Quality of life: Factors determining its measurement complexity. *Engineering Economics*, vol. 22, no. 2, pp. 147–156, 2011. DOI: 10.5755/j01.ee.22.2.311.

[7] A. Sőrés y K. Pető. Measuring of subjective quality of life. *Procedia Economics and Finance*, vol. 32, pp. 809–816, 2015. DOI: 10.1016/S2212-5671(15)01466-5.

[8] K. Ruggeri *et al.* Well-being is more than happiness and life satisfaction: A multidimensional analysis of 21 countries. *Health and Quality of Life Outcomes*, vol. 18, pp. 1–16, 2020. DOI: 10.1186/s12955-020-01423-y.

[9] V. J. J. García. Hacia un nuevo sistema de indicadores de bienestar. *Revista Internacional de Estadística y Geografía*, vol. 2, no. 1, pp. 78–95, 2011.

[10] R. Loveridge *et al.* Measuring human wellbeing: A protocol for selecting local indicators. *Environmental Science & Policy*, vol. 114, pp. 461–469, 2020. DOI: 10.1016/j.envsci.2020.09.002.

[11] IMCO. ¿Cómo se mide la felicidad? [Online]. Available: `https://imco.org.mx/articulo_es/como_se_mide_la_felicidad/`.

[12] C. Isaza Espinosa y A. Navarro Burgos. Transferencia de políticas de gobierno abierto. Análisis a tres documentos de política pública en Colombia. *Estudios Políticos (Universidad de Antioquia)*, vol. 70, 2024. DOI: 10.17533/udea.espo.n70a07.

[13] E. Allardt. Having, loving, being: An alternative to the Swedish model of welfare research. *The quality of life*, vol. 8, pp. 88–95, 1993. DOI: 10.1093/0198287976.003.0008.

[14] C. Proctor. Subjective well-being (SWB). in *Encyclopedia of Quality of Life and Well-Being Research*, Cham: Springer International Publishing, 2024, pp. 6952–6956. DOI: 10.1007/978-3-031-17299-1_2905.

[15] A. Maslow y A. Turóczi. A lét pszichológiája felé. Ursus Libris, 2003.

[16] P. Desmet y S. Fokkinga. Beyond Maslow's pyramid: Introducing a typology of thirteen fundamental needs for human-centered design. *Multimodal Technologies and Interaction*, vol. 4, no. 3, p. 38, 2020. DOI: 10.3390/mti4030038.

[17] D. Kahneman. Objective happiness. in *Wellbeing: The Foundations of Hedonic Psychology*, D. Kahneman, E. Diener, and N. Schwarz, Eds. Russell Sage Foundation, New York, 1999, pp. 25.

[18] M. S. Kimball and R. J. Willis. Utility and Happiness. Working Paper No. w31707, National Bureau of Economic Research, 2023. DOI: 10.3386/w31707.

[19] E. Diener y M. E. P. Seligman. Beyond money: toward an economy of wellbeing. *Psychological Science in the Public Interest*, vol. 5, no. 1, pp. 1–31, 2004. DOI: 10.1111/j.0963-7214.2004.00501001.x.

[20] E. Diener y E. Suh. Measuring quality of life: economic, social, and subjective indicators. *Social Indicators Research*, vol. 40, pp. 189–216, 1997. DOI: 10.1023/A:1006859511756.

[21] C. Guzmán-Pozo *et al.* Impacto de la inteligencia emocional y del clima escolar sobre el bienestar subjetivo y los síntomas emocionales en la adolescencia. *Estudios sobre Educación*, 2024. DOI: 10.15581/004.49.003.

https://doi.org/10.36677/rici.v3i1.25065 Enero-Julio 2025 ISSN: 2992-7447 págs. 35-47

[22] G. Tejeda Parra y B. Burgos Flores. Influencia del empleo en el bienestar subjetivo de las personas en México. *Relaciones. Estudios de Historia y Sociedad*, vol. 41, no. 163, pp. 58–81, 2020. DOI: 10.24901/rehs.v41i163.689.

[23] S. L. Stewart-Brown. Public mental health: an interdisciplinary subject? *British Journal of Psychiatry*, vol. 207, pp. 192–194, 2015. DOI: 10.1192/bjp.bp.114.157966.

[24] S. M. Skevington y J. R. Böhnke. How is subjective wellbeing related to quality of life? Do we need two concepts and both measures? *Social Science & Medicine*, vol. 206, pp. 22–30, 2018. DOI: 10.1016/j.socscimed.2018.03.044.

[25] M. Durand. The OECD better life initiative: How's life? and the measurement of well-being. *Review of Income and Wealth*, vol. 61, no. 1, pp. 4–17, 2015. [Online]. Available: https://www.oecd.org/statistics/better-life-initiative-country-note-mexico-in-espagnol.pdf. DOI: 10.1111/roiw.12156.

[26] A. D. Sagar y A. Najam. The human development index: a critical review. *Ecological Economics*, vol. 25, no. 3, pp. 249–264, 1998. DOI: 10.1016/S0921-8009(97)00168-7.

[27] A. C. Michalos. *Essays on the Quality of Life*, vol. 19. Springer Science & Business Media, 2013. DOI: 10.1007/978-94-007-6133-6.

[28] A. M. F. Poncela. Una revisión del programa Pueblos Mágicos. *CULTUR: Revista de Cultura e Turismo*, vol. 10, no. 1, pp. 3–34, 2016.

[29] L. A. Castillo, C. A. Hiriart Pardo, y D. Barrera Fernández. Resiliencia y conservación en Pueblos Mágicos de México. Los casos de Pátzcuaro y Mexcaltitán. *Bitácora Urbano Territorial*, vol. 31, no. 1, pp. 195–210, 2021. DOI: 10.15446/bitacora.v31n1.86340.

[30] M. G. Lúgigo. El Programa Pueblos Mágicos y algunos de sus efectos locales. *Dimensiones Turísticas*, vol. 5, no. 8, pp. 131–141, 2021. DOI: 10.47557/MJCU3028.

[31] A. Pinassi y J. D. Albarrán Periáñez. Entre discursos patrimoniales y turísticos: Análisis comparado de los programas 'Los pueblos más bonitos de España' y 'Pueblos Auténticos' de Argentina. *Revista Investigaciones Turísticas*, no. 24, pp. 1–22, 2022. DOI: 10.14198/INTURI2022.24.1.

[32] J. M. Colomer. *El utilitarismo: una teoría de la elección racional*, vol. 49. Barcelona, Spain: Editorial Montesinos, 1987.

[33] J. Mill. An Essay on Government, reprinted in Lively and Rees 1978, 53–95. M. Bronfenbrenner, Notes on the elasticity of derived demand. *Oxford Economic Papers*, vol. 13, no. 3, pp. 254–261, 1961. DOI: 10.1093/oxfordjournals.oep.a040871.

[34] G. Vargas-Solar y A.-S. Castillo-Camporro. A Holistic Approach for Measuring Quality of Life in 'La Condesa' Neighbourhood in Mexico City. *ILCEA. Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, vol. 39, 2020. DOI: 10.4000/ilcea.10063.

[35] WHO Quality of Life Assessment Group. ¿Qué calidad de vida? *Foro Mundial de la Salud*, vol. 17, no. 4, pp. 385–387, 1996. [Online]. Available: https://iris.who.int/handle/10665/55264.

IDEAS EN CIENCIAS DE LA INGENIERÍA

Exploración de los Límites del Muestreo Selectivo para el Reconocimiento de Emociones del Habla Dependientes del Hablante

Exploring Selective Sampling Bounds for Speaker-Dependent Speech Emotion Recognition

Castorena, C.[1]*, Lopez-Ballester, J.[1], Ferri, F.J.[1], Cobos, M.[1],
(*carlos.castorena@uv.es)

Recibido: 2024-11-12 Aceptado: 2024-11-27

Resumen

El reconocimiento de emociones en el habla es crucial para diversas aplicaciones, desde el ámbito de la salud mental hasta la interacción con máquinas. Aunque los modelos basados en aprendizaje profundo han demostrado ser efectivos en este tipo de tareas, su rendimiento puede verse afectado cuando se enfrentan a nuevos individuos, incluso si se entrenaron con grandes y diversos conjuntos de datos. En este estudio, se analizó la relevancia de ajustar y personalizar estos modelos para mejorar la precisión en el reconocimiento emocional. A pesar de que los modelos generales ofrecen buenos resultados iniciales, se percibe que su capacidad para predecir emociones disminuye cuando se aplican a nuevos sujetos. A través de este trabajo, se muestra cómo personalizar un modelo con una pequeña cantidad de muestras de un nuevo individuo puede aumentar considerablemente su desempeño, utilizando técnicas como el ajuste fino y la transferencia de aprendizaje. Finalmente, se discuten las implicaciones de estos resultados en aplicaciones reales y se. sugieren posibles direcciones para futuras investigaciones en la detección de emociones mediante inteligencia artificial.

Palabras clave: reconocimiento de emociones, deep learning, adaptación al dominio.

Abstract

Emotion recognition in speech is crucial for various applications, ranging from mental health to human-machine interaction. Although deep learning-based models have proven effective in these tasks, their performance can be affected when faced with new individuals, even if they were trained with large and diverse datasets. In this study analyze the importance of fine-tuning and personalizing these models to improve emotional recognition accuracy. Despite the good initial results of general models, we

[1]Departament d'Informàtica, Universitat de València, Burjassot, Spain. 46100.

observe a decline in their predictive ability when applied to new subjects. Through this work, it is demonstrated how personalizing a model with a small number of samples from a new individual can significantly improve its performance, using techniques such as fine-tuning and transfer learning. Finally, the implications of these findings in real-world applications are discussed, and possible directions for future research in emotion detection through artificial intelligence are suggested.

Keywords: emotion recognition, deep learning, domain adaptation.

INTRODUCTION

Speech Emotion Recognition (SER) is a significant challenge in natural language processing and human-machine interaction, as it allows the identification of emotions based on acoustic signals. This technology has practical applications in various fields: in customer service, where it monitors the user's emotional state to improve interaction quality by detecting frustration or dissatisfaction [1]; in healthcare, where SER facilitates mental health monitoring by providing essential information for diagnosing and treating emotions such as sadness or fear [2]; and in the automotive industry, where it can be used to assess the driver's emotional state, enhancing both safety and comfort.

Although current systems efficiently process speech for tasks such as transcription or automatic translation [3], SER faces significant limitations. One key challenge lies in the subjective perception of emotions by annotators [4], a complex issue as different people may interpret varying emotional content from the same audio. Furthermore, the diversity of speakers, contexts, and environments impacts model robustness. Additionally, SER is influenced not only by the acoustic characteristics of the voice but also by speech content; while these semantic associations can improve performance, they may also introduce challenges as they are not reliably transferable to new contexts [5].

Emotions in SER can be represented through either categorical or dimensional labels, each suited to different aspects of emotion recognition. Categorical representations label emotions as discrete classes (such as happiness, anger, sadness, and neutrality) making classification straightforward and widely interpretable. This approach is often advantageous, as it simplifies both model design and performance evaluation by aligning with commonly understood emotional states. Moreover, categorical labels can be more robust and consistent across different speakers, languages, and cultural contexts, which is especially beneficial in practical applications requiring clear emotion identification.

Dimensional representations, in contrast, define emotions by continuous variables, typically arousal (intensity), valence (positivity/negativity), and dominance (control), allowing a nuanced analysis of emotional states. While these continuous labels offer flexibility to capture subtle variations within each emotion, they add complexity to model training and interpretation. The dimensional approach also requires meticulous labeling, as subjective interpretations of arousal, valence, and dominance may vary significantly among annotators. As a result, categorical representations generally offer a practical balance of simplicity and clarity, making them well-suited for many SER tasks, whereas dimensional representations may be preferred for specialized applications that demand fine-grained emotional assessment.

In traditional SER, when a model is trained and evaluated with the same speaker, it is known as speaker-dependent SER [6]. These models tend to achieve high levels of accuracy [7, 8]. However, their applicability is limited due to low generalization capacity. By contrast, speaker-independent SER

https://doi.org/10.36677/rici.v3i1.25132 Enero-Julio 2025 ISSN: 2992-7447 págs. 48-58

[6] presents a greater challenge, as it requires the model to recognize emotions in a generalized way across individuals, each with unique emotional expressions and acoustic characteristics. This demands that models adapt to effectively handle variations not fully represented in the training phase.

To address these limitations, speaker-independent SER studies often rely on domain adaptation techniques, where the source data representation (training speakers) is adjusted to resemble the characteristics of new speakers. For example, Lu et al. [9] focused their research on measuring domain shift between source and target data to reduce this discrepancy. AbdelWaha et al. [10, 6, 4] have made significant contributions to this field: in 2015, they presented a method to adapt the target domain using different percentages of target domain samples; in 2017, they introduced an active learning-based method to incorporate these samples into source domain training, combined with domain adaptation strategies; and in 2018, they proposed the use of adversarial multitask learning to align representations between training and test domains using a gradient reversal layer. These studies aim to develop models capable of generalizing with a relatively low number of new samples.

In this work, we explore how, through the use of a small set of labeled samples, it is possible to fine-tune a pre-trained model to improve its accuracy in recognizing emotions in a new speaker. We first analyze the model's performance using a random selection of samples, then compare its performance when using "key samples," which have a significant impact on classification compared to others. Although there are previous studies along this line [10, 11], most do not fully explore classifier capacities when working with a small number of samples, limiting themselves to variations in training set size without examining optimal sample selection. The structure of this work is organized as follows: in the Methodology section, we describe the data, the model used, and the strategy applied for adapting to new speakers. In the Results and Discussion section, we analyze and discuss the main findings obtained from our experiments. Finally, in the Conclusions and Future Work section, we summarize key insights and propose directions for future research.

Methodology

Dataset

The landscape of available datasets for emotion detection in speech is diverse, with each resource distinguished by unique attributes, such as the number of participants, represented languages, types and quantities of emotions, and sample count. For this study, we drew on multiple datasets that are commonly applied in SER research, each contributing specific advantages to the robustness of our model. All datasets consist of audio samples with a sampling rate of 16,000 Hz and durations ranging from 15 to 60 seconds. Each audio recording features isolated speech from a single speaker without background noise, and every audio sample is assigned a single, unique emotion label.

The MSP-Podcast dataset [12] offers a substantial foundation with its 83 hours of English podcast recordings, annotated for valence, activation, and dominance dimensions. This corpus includes contributions from 60 speakers (32 male and 28 female), with sample counts per speaker ranging widely from 42 to 912, resulting in a total of over 50,000 samples. This dataset's dimensional labeling provides a nuanced approach to capturing emotional expression in naturalistic settings. Similarly, the IEMOCAP dataset [13] offers 12 hours of dialog interactions, including both scripted and improvised segments from 10 speakers (balanced with 5 men and 5 women). The dataset provides 10,039 labeled samples, allowing

for a comprehensive view of emotional dynamics in conversational speech.

The Berlin Emotional Speech Database (EMO-DB) [14] contributes a contrasting structured approach with recordings from 10 professional actors (5 men and 5 women) enacting seven discrete emotions, yielding a dataset of 800 samples. Each sample is short but precisely labeled, allowing a focused analysis of fundamental emotional expressions. From a Spanish-language perspective, the Mexican Emotional Speech Database (MESD) [15] broadens our linguistic scope, featuring recordings from three speakers (a man, a woman, and a child) across six emotions with 48 samples per emotion and speaker, totaling 864 samples.

For additional variation in emotional expression, the RAVDESS dataset [16] includes 7,356 recordings from 24 native English-speaking actors, who express a range of emotions, including calm, happy, sad, angry, fearful, surprise, and disgust. Finally, the Emotional Speech Dataset (ESD) [17] stands out for providing the highest number of samples per emotion per individual, with 350 dialogues for each emotional category. This abundance of samples per speaker makes it ideal for experiments involving new speakers, offering a solid base for adaptation tasks. The dataset includes both native English and Mandarin speakers, with 10 speakers from each language (5 men and 5 women) expressing emotions across five categories: angry, happy, neutral, sad, and surprise.

To maintain consistency in the analysis, we focus on five core emotions (Happiness, Anger, Sadness, Surprise, and Neutral) and work with categorical labels across these datasets, discarding any additional labels. This standardization ensures comparability across datasets.

Base Model

Our approach utilizes a deep classification model pre-trained on data from diverse speakers, as shown in Figure 1. We begin with a modified version of Wav2Vec [5], initially trained on the MSP-Podcast dataset and validated with IEMOCAP, to obtain a deep audio representation. Wav2Vec, a transformer-based architecture, was originally designed to extract both acoustic and semantic features from speech and is primarily used for tasks like automatic speech recognition. Unlike strategies based solely on speech formants [18] or other handcrafted features such as pitch [19], energy [20], and spectral descriptors [21], which focus on specific phonetic or prosodic characteristics, Wav2Vec seeks to capture a broader range of audio features, encompassing both phonetic and semantic information. This comprehensive feature extraction is particularly advantageous for capturing the nuanced and multidimensional nature of emotional expression in speech. Given these capabilities, Wagner et al. adapted Wav2Vec to SER by retraining it specifically for emotion recognition, allowing the model to leverage both acoustic and semantic content for emotion prediction. Notably, their adaptation focuses on dimensional labels, representing emotions within a three-dimensional space of Arousal, Valence, and Dominance.

In our approach, we utilize only the internal representation generated by Wav2Vec, substituting their regression module with our own classification module, which consists of three dense layers with 64, 32, and 5 neurons. The first two layers use ReLU activation functions, while the final layer applies Softmax to produce categorical predictions. To train this classifier module, we use the EMO-DB, MESD, and RAVDESS training partition datasets (70% of the samples), covering a total of 37 speakers. Training is conducted for 200 epochs, with a batch size of 64, using Categorical Cross-Entropy as the loss function. Throughout training, the feature extraction module remains untrainable, focusing all training on the classification module.

https://doi.org/10.36677/rici.v3i1.25132 Enero-Julio 2025 ISSN: 2992-7447 págs. 48-58

This model, referred to as the base model, serves as the foundation for subsequent experiments. Using the final state of this base model, we will apply adaptation techniques to fine-tune the classifier for effective emotion recognition in new speakers.

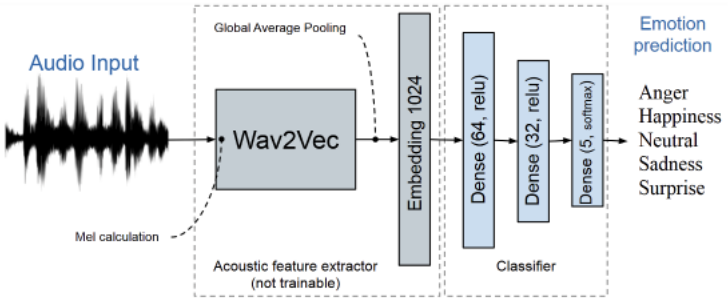


Figure 1: *A pre-trained base model for the classification of 5 emotions.*

Strategies for New Speakers

Our strategy aims to fine-tune the base model's weights to better recognize emotions in new speakers from the ESD dataset. We selected 2 native English speakers and 2 native Mandarin speakers, dividing each speaker's data into separate Training and Testing sets with a 70-30 split. The Training set includes all available samples for potential selection (although not all are necessarily used in each experiment), while the Testing set remains fixed for each speaker throughout all trials.

It is essential to note that each new speaker is processed independently; our approach is designed to improve performance specifically for one new speaker at a time.

The first approach, called Random, incrementally selects 5 random samples (one for each emotion) from the Training set at each step. These samples, combined with those gathered from previous steps, are used to fine-tune the base model over 30 epochs, with only the weights in the classifier layers updated. This incremental process is repeated up to 5 times, resulting in sample sizes ranging from 5 to 25 samples, while using the same loss function configuration as the base model to retrain it. At each step, the model's performance is evaluated on the Testing set using accuracy scores (ACC, percentage of correctly labeled samples), with class-balanced representation across all 5 emotion categories. Each experiment is repeated 5 times per speaker to ensure reliability.

On the other hand, the strategy called Best, selects the most representative samples at each step. At each increment, 200 random sets of 5 samples are evaluated, and the set achieving the highest ACC scores on the Training set is selected. These top-performing samples are accumulated across steps to further refine the model. The base model parameters in the Best strategy remain consistent with those used in the Random approach.

RESULTS AND DISCUSSION

After training, the base model achieved ACC values close to 90% on the test partitions of EMO-DB, MESD, and RAVDESS, as shown in the initial step of Figure 4. It is important to note that these dataset partitions do not ensure speaker independence, meaning the same speaker may be present in both the training and test sets. This result underscores the model's effectiveness in emotion recognition within a speaker-dependent setup, aligning with similar results in recent state-of-the-art studies, and serves as a promising baseline for extending its application to a speaker-independent context.

On the other hand, when the model is tested with new speakers, its performance declines significantly. Figure 2 illustrates the ACC values achieved by two new English-speaking individuals. First, we observe the dashed line marked as "Reference," which indicates the ACC obtained by the base model for the new speaker without any adjustment, yielding scores far below the 90% achieved on the source domain test partition (0.41 and 0.54, respectively). One might assume that this drop in performance could stem not only from speaker domain differences but from other speaker-specific factors. To investigate this further, we calculated the ACC that could be achieved if we were to use all available samples in the training partition for each new speaker; this is marked in the figure by a dashed line labeled "Maximum." In both cases, the results show a significant gap between the Maximum and Reference values, suggesting that the model does indeed have the capacity to classify emotions for these speakers effectively. However, for the first speaker, the ACC value is considerably lower, reaching only 61%, while the second speaker achieves an ACC closer to speaker-dependent results, with 84%. This disparity underscores the challenge models face in generalizing effectively when encountering new speakers, highlighting the complexities of speaker-independent emotion recognition.

The solid black line in Figure 2 shows the average ACC across the five repetitions of the Random selection strategy. In both experiments, we observe that as the model is retrained with a greater number of samples, its performance tends to improve. Notably, after the first step (using only 5 samples), ACC increases significantly compared to the initial state, though subsequent increments yield less pronounced gains. On the other hand, the blue line (Best mean) and violin plots represent the averages and distributions for the Best selection strategy. Initially, when only 5 samples are used, the difference between Random and Best is negligible since both start with a random sample set, reflected in their approximately equal averages. However, with the next step—where the model is retrained with 10 samples (including the 5 most representative ones identified in the previous step and indicated by the dashed blue line and the label Best max)—a marked improvement emerges compared to the Random selection. This finding suggests that not only is it essential to retrain the model with new samples, but careful selection of these samples can lead to significantly better results, approaching those obtained when the entire training partition for the new speaker is utilized. From subsequent steps onward, the ACC values seem to stabilize, indicating a plateau in the gains achieved by adding more samples.

Strictly speaking, the base model was not trained with Mandarin speakers. Although Wav2Vec may contain some information relevant to Mandarin from its pre-training, the model's behavior with new Mandarin speakers aligns with that observed for English speakers, a language highly represented in the datasets. As shown in Figure 3, the classification problem can be solved when the entire training set for the new speaker is used. With the Random selection approach, performance improves as more sam-

https://doi.org/10.36677/rici.v3i1.25132 Enero-Julio 2025 ISSN: 2992-7447 págs. 48-58

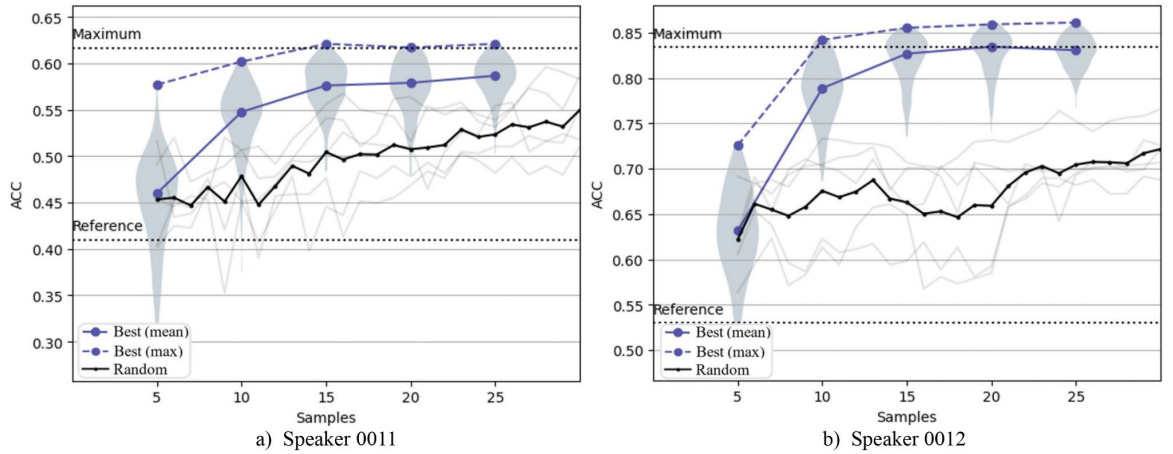


Figure 2: *ACC results for new English speakers with identification codes 0011 and 0012 from the ESD dataset, using a maximum of 25 samples.*

ples are used to retrain the base model; however, it remains consistently below the best strategy, which achieves results equal to or near the Maximum with as few as 10 samples. The model's consistent performance across languages could stem from two factors: first, that the acoustic content and, consequently, the patterns learned by the model may not be language-dependent; or second, that the pre-trained Wav2Vec model retains relevant language information within its deep representations from its unsupervised training phase. In either case, this suggests that language may not be a decisive factor for effective emotion classification.

To analyze the effectiveness of the Random and Best methods, we used the non-parametric Wilcoxon rank-sum test, which is appropriate for comparing the central tendencies of two groups when the samples do not follow a normal distribution or when the sample sizes are unequal, as in this study, where the Random group contains only 5 measures.

The Wilcoxon rank-sum test was applied independently for each speaker at each incremental step to evaluate whether the best method outperforms the Random method. The p-values from these tests can be observed in Table 1. In the first step, when only 5 samples are used, no significant difference is found, as both methods start with a random selection. However, as more samples are added, the p-value tends to decrease, suggesting that the best method becomes increasingly superior to the Random method as the number of processed samples grows.

The accuracy of the same sequence of models in Figure 3, but measured on the test partition of the source dataset is presented in Figure 4. Initially, the model starts with a high ACC of around 90%, as previously discussed. However, as the model is incrementally adjusted with samples from the new speaker, a significant drop in ACC is observed, indicating that the model increasingly prioritizes alignment with the new speaker's characteristics, even at the cost of losing previously learned relationships. After the first few steps, this decline stabilizes, suggesting a local minimum for this particular partition.

The experimental design of this work uncovers several critical points for future contributions in the field. Firstly, while the best strategy demonstrates the effectiveness of intelligently selecting samples,

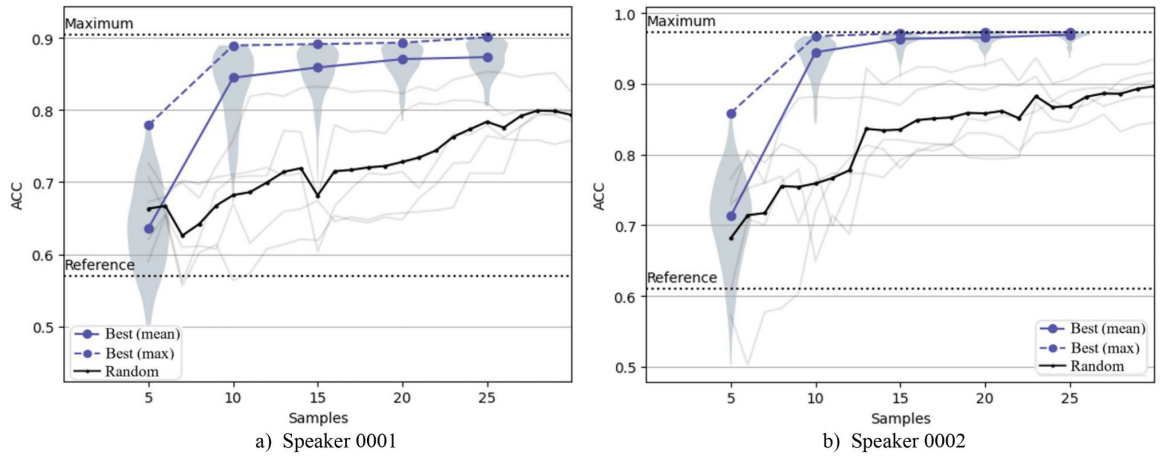


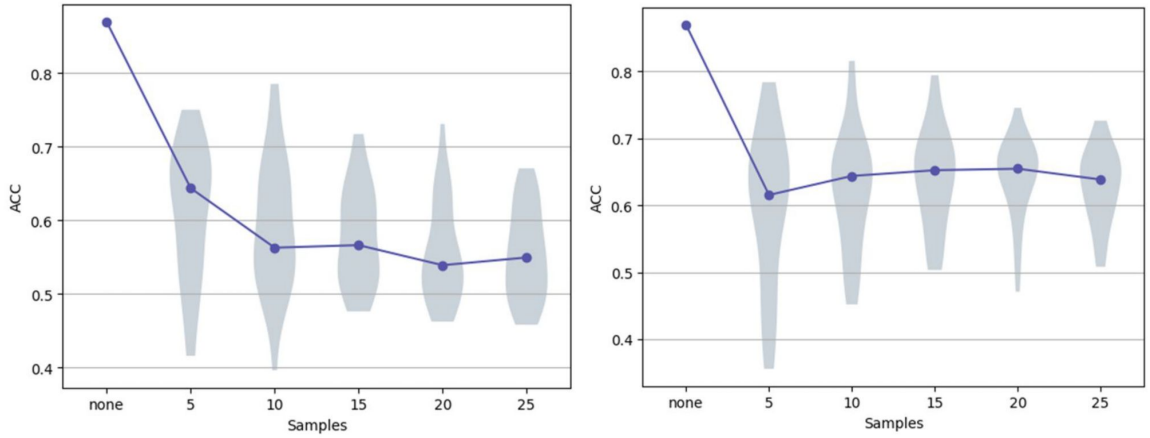
Figure 3: *ACC results for new Mandarin speakers with identification codes 0001 and 0002 from the ESD dataset, using a maximum of 25 samples.*

Table 1: p-values from the Wilcoxon rank-sum tests comparing the performance of the Random and Best methods across incremental steps for each speaker. An asterisk (*) indicates that the null hypothesis (H0) is rejected, and there is a statistically significant difference between the groups at a significance level of α=0.05.

Speaker	5 samples	10 samples	15 samples	20 samples	25 samples
0001	0.0902	0.0001*	0.0001*	0.0001*	0.0030*
0002	0.1700	0.0013*	0.0001*	0.0003*	0.0009*
0011	0.1470	0.0007*	0.0001*	0.0001*	0.0001*
0012	0.1751	0.0021*	0.0002*	0.0001*	0.0001*

https://doi.org/10.36677/rici.v3i1.25132 Enero-Julio 2025 ISSN: 2992-7447 págs. 48-58

it would be impractical in a real-world scenario, as it requires prior knowledge of labels for the new speaker. Nonetheless, it highlights the potential of well-chosen samples to drive significantly better results. Additionally, the evident loss of effectiveness on the original data following retraining underlines the importance of adapting the feature space of the new speaker to be compatible with that of previous speakers. This adjustment will be essential to achieving balanced performance across both the new and original data, paving the way for more robust speaker-independent SER models.



a) Speaker 0001 b) Speaker 0002

Figure 4: *ACC over the source test dataset as the model is retrained using data from new speakers.*

Conclusions

This study has tackled the issue of Speech Emotion Recognition (SER) by customizing a pre-trained model to accommodate new speakers using a limited data set. By employing our approach, which includes both a random and an optimized sample selection strategy, we have shown that the base model's accuracy can be significantly enhanced. Moreover, our results indicate that, even in cases where the optimal samples cannot be precisely identified, a random sampling approach can still lead to meaningful improvements. This insight is especially relevant for real-world applications where comprehensive labeling is impractical. Our findings reinforce the viability of adapting SER models for new speakers with minimal data and underscore the effectiveness of strategic sample selection. This paves the way for further advancements in SER, where models can be tailored more efficiently, requiring fewer resources, thus boosting the potential for enhanced human-machine interaction across diverse applications.

Acknowledgments

We would like to thank the Spanish Research Agency (AEI) and the European Regional Development Fund (ERDF) for partially funding this research through the projects TED2021-131003B-C21 and PID2022-137048OB-C41, funded by MCIN/AEI/10.13039 /501100011033 and the “EU Union NextGenerationEU/PRTR.” Additionally, we thank the Generalitat Valenciana for funding this work through the Santiago Grisolia program (GRISOLIAP/2021/060, CPI-21-232).

DECLARATION OF CONTRIBUTION OF AUTHORS AND COLLABORATORS

Carlos Castorena: Formal analysis, Investigation, Experimentation & Writing. **Maximo Cobos**: Conceptualization, Supervision, Review & Editing. **Jesus Lopez-Ballester**: Review & Editing. **Francesc J. Ferri**: Conceptualization, Supervision, Review & Editing.

REFERENCES

[1] X. Li & R. Lin. Speech emotion recognition for power customer service. *7th International Conference on Computer and Communications (ICCC)*, pp. 514–518, 2021. DOI: 10.1109/ICCC54389.2021.9674619.

[2] N. Elsayed, Z. ElSayed, N. Asadizanjani, M. Ozer, A. Abdelgawad, & M. Bayoumi. Speech emotion recognition using supervised deep recurrent system for mental health monitoring. *IEEE 8th World Forum on Internet of Things (WF-IoT)*, pp. 1–6, 2022. DOI: 10.48550/arXiv.2208.12812.

[3] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. Mcleavey, & I. Sutskever. Robust speech recognition via large-scale weak supervision. *Proceedings of the 40th International Conference on Machine Learning*, vol. 202, pp. 28492–28518, 2023. DOI: 10.48550/arXiv.2212.04356.

[4] M. AbdelWahab & C. Busso. Domain Adversarial for Acoustic Emotion Recognition. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 26, no. 12, pp. 2423–2435, 2018. DOI: 10.1109/TASLP.2018.2867099.

[5] J. Wagner, A. Triantafyllopoulos, H. Wierstorf, M. Schmitt, F. Burkhardt, F. Eyben, & B. W. Schuller. Dawn of the transformer era in speech emotion recognition: Closing the valence gap. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, pp. 10745–10759, 2023. DOI: 10.48550/arXiv.2203.07378.

[6] M. Abdelwahab & C. Busso. Incremental adaptation using active learning for acoustic emotion recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5160–5164, 2017. DOI: 10.1109/ICASSP.2017.7953140.

[7] S. Madanian, T. Chen, O. Adeleye, J. M. Templeton, C. Poellabauer, D. Parry, & S. L. Schneider. Speech emotion recognition using machine learning — A systematic review. *Intelligent Systems with Applications*, vol. 20, 2023. DOI: 10.1016/j.iswa.2023.200266.

[8] A. Hashem, M. Arif, & M. Alghamdi. Speech emotion recognition approaches: A systematic review. *Speech Communication*, vol. 154, 2023. DOI: 10.1016/j.specom.2023.102974.

[9] C. Lu, Y. Zong, W. Zheng, Y. Li, C. Tang, & B. W. Schuller. Domain Invariant Feature Learning for Speaker-Independent Speech Emotion Recognition. *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 30, pp. 2217–2230, 2022. DOI: 10.1109/TASLP.2022.3178232.

https://doi.org/10.36677/rici.v3i1.25132 Enero-Julio 2025 ISSN: 2992-7447 págs. 48-58

[10] M. Abdelwahab & C. Busso. Supervised domain adaptation for emotion recognition from speech. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5058–5062, 2015. DOI: 10.1109/ICASSP.2015.7178934.

[11] C. Castorena, J. Lopez-Ballester, M. Cobos & F. J. Ferri. Explorando la personalización de modelos en el reconocimiento de emociones en el habla. *Congreso Ibérico de Acústica – 55º Congreso Español de Acústica – Tecniacústica*, 2024.

[12] C. Busso, M. Bulut, C. C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, & S. S. Narayanan. Iemocap: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, vol. 42, pp. 335–359, 2008. DOI: 10.1007/s10579-008-9076-6.

[13] R. Lotfian & C. Busso. Building naturalistic emotionally balanced speech corpus by retrieving emotional speech from existing podcast recordings. *IEEE Transactions on Affective Computing*, vol. 10, pp. 471–483, 2019. DOI: 10.1109/TAFFC.2017.2736999.

[14] F. Burkhardt, W. F. Sendlmeier, A. Paeschke, M. Rolfes, & B. Weiss. A database of German emotional speech. *Proc. Interspeech*, pp. 1517–1520, 2005. DOI: 10.21437/Interspeech.2005-446.

[15] M. M. Duville, L. M. Alonso-Valerdi, & D. I. Ibarra-Zarate. Mexican emotional speech database based on semantic, frequency, familiarity, concreteness, and cultural shaping of affective prosody. *Data*, vol. 6, p. 130, 2021. DOI: 10.3390/data6120130.

[16] S. R. Livingstone & F. A. Russo. The Ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLOS ONE*, vol. 13, p. e0196391, 2018. DOI: 10.1371/journal.pone.0196391.

[17] K. Zhou, B. Sisman, R. Liu, & H. Li. Seen and unseen emotional style transfer for voice conversion with a new emotional speech dataset. *ICASSP*, pp. 920–924, 2021. DOI: 10.48550/arXiv.2010.14794.

[18] Z. T. Liu, A. Rehman, M. Wu, W. H. Cao, & M. Hao. Speech emotion recognition based on formant characteristics feature extraction and phoneme type convergence. *Information Sciences*, vol. 563, 2021. DOI: 10.1016/j.ins.2021.02.016.

[19] S. Vaijayanthi & J. Arunnehru. Synthesis Approach for Emotion Recognition from Cepstral and Pitch Coefficients Using Machine Learning. *Lecture Notes in Electrical Engineering*, vol. 733, 2021. DOI: 10.1007/978-981-33-4909-4_39.

[20] R. Thirumuru, K. Gurugubelli, & A. K. Vuppala. Novel feature representation using single frequency filtering and nonlinear energy operator for speech emotion recognition. *Digital Signal Processing: A Review Journal*, vol. 120, 2022. DOI: 10.1016/j.dsp.2021.103293.

[21] J. Ancilin & A. Milton. Improved speech emotion recognition with Mel frequency magnitude coefficient. *Applied Acoustics*, vol. 179, 2021. DOI: 10.1016/j.apacoust.2021.108046.

IDEAS EN CIENCIAS DE LA INGENIERÍA

Evaluación del método DBSCAN para identificar patrones de estrés en estudiantes usando datos no etiquetados

Evaluation of the DBSCAN method to identify student stress patterns using unlabeled data

Reyes-Acosta, J.[1]*, Rodríguez-Arce, J.[2] (*jreyesa838@alumno.uaemex.mx)

Recibido: 2024-11-12 Aceptado: 2024-11-29

Resumen

Un desafío clave en la clasificación del estrés es la obtención de datos etiquetados, que requiere la intervención de expertos, como psicólogos, para garantizar precisión, dada la variabilidad individual en la respuesta al estrés. Este estudio evalúa el método DBSCAN (Density-Based Spatial Clustering of Applications with Noise) para identificar respuestas fisiológicas al estrés en datos no etiquetados, destacando su capacidad para detectar patrones en datos ruidosos o complejos. Se analizaron señales de fotopletismografía (PPG) y actividad electrodérmica (EDA) registradas en estudiantes universitarios durante 51 pruebas de inducción de estrés en entornos académicos de bajo, medio y alto estrés. Las señales fueron segmentadas y procesadas para reducir artefactos y asegurar un número homogéneo de muestras por sujeto y escenario. Se extrajeron características de las señales, normalizadas con z-score, y se aplicó Análisis de Componentes Principales (PCA) para reducir la dimensionalidad y facilitar la visualización, generando conjuntos separados para PPG y EDA. Finalmente, se aplicó DBSCAN al conjunto unificado de datos no etiquetados. Los resultados muestran que DBSCAN no logró formar clústeres claramente definidos con los parámetros seleccionados, evidenciando una limitación del método en este contexto, a pesar de su capacidad para manejar datos complejos. Los hallazgos sugieren la necesidad de explorar métodos más avanzados de procesamiento y clustering para datos no etiquetados en estudios futuros.

Palabras clave: DBSCAN, clustering, aprendizaje no supervisado, identificación de estrés, análisis de señales.

[1]Facultad de Ingeniería, Universidad Autónoma del Estado de México, Paseo Universidad S/N, Toluca, Estado de México, México. CP. 50130.

[2]Facultad de Ingeniería, Universidad Autónoma del Estado de México, Paseo Universidad S/N, Toluca, Estado de México, México. CP. 50130.

https://doi.org/10.36677/rici.v3i1.25133 Enero-Julio 2025 ISSN: 2992-7447 págs. 59-71

Abstract

A critical challenge in stress classification is obtaining labeled data, which requires the intervention of experts, such as psychologists, to ensure accuracy due to the individual variability of stress responses. This study evaluates the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) method to identify physiological stress responses in unlabeled datasets. This method is characterized by its ability to detect patterns in noisy or complex data. Photoplethysmography (PPG) and electrodermal activity (EDA) signals were analyzed from university students during 51 stress induction tests in educational environments representing low, medium, and high-stress levels. The signals were segmented and preprocessed to reduce artifacts, ensuring consistent samples per subject and scenario. Signal features were extracted and standardized using z-score normalization. Principal Component Analysis (PCA) was used for dimensionality reduction to facilitate visualization, resulting in separate datasets for PPG and EDA. Finally, DBSCAN was applied to the unified set of unlabeled features. The results suggest that DBSCAN failed to form well-defined clusters with the selected parameters, highlighting a limitation of the method in this context despite its capability to handle complex data structures. These findings suggest the need to explore more advanced processing and clustering methods for unlabeled data in future research.

Keywords: DBSCAN, clustering, unsupervised learning, stress identification, signal analysis.

Introducción

La salud mental, particularmente el estrés, ha ganado relevancia en el ámbito tecnológico. Su impacto en la sociedad es evidente: al menos el 50 % de los adultos experimenta niveles moderados de estrés, mientras que un 25 % presenta niveles altos, según la Asociación Americana de Psicología (APA) [1]. Estas clasificaciones dependen de la escala psicológica empleada, la interpretación de las respuestas y el análisis del experto. En general, puntuaciones altas indican mayor estrés, mientras que las bajas reflejan niveles menores [A].

Este contexto ha motivado múltiples estudios para explorar situaciones, factores y estresores, así como sus efectos en el cuerpo humano. Entre estos, destaca la encuesta anual “Stress in America”, reconocida por evaluar las actitudes y percepciones relacionadas con el estrés en la población de los Estados Unidos [1]. Asimismo, se ha promovido el desarrollo de sistemas tecnológicos para evaluar los niveles de estrés en sujetos, integrando herramientas como informes psicológicos, sistemas de adquisición de señales fisiológicas y algoritmos de aprendizaje automático (Machine Learning, ML) [2].

El estrés es la respuesta psicofisiológica del cuerpo ante estímulos percibidos como amenazas o desafíos, con efectos notables en la salud, la productividad y la calidad de vida. Según la teoría del Síndrome General de Adaptación (GAS) de Selye, el estrés se desarrolla en tres etapas: alarma, resistencia y agotamiento, describiendo cómo el cuerpo enfrenta y se adapta al estrés, hasta llegar al agotamiento si la exposición es prolongada [3].

Una forma efectiva de medir la respuesta somática o fisiológica al estrés es mediante bioseñales, que

son medidas objetivas evaluadas con, que son medidas objetivas evaluadas con precisión. Estas señales proporcionan información crucial sobre cambios metabólicos, morfológicos y funcionales en el cuerpo, permitiendo realizar evaluaciones objetivas de la salud mental de los sujetos [2].

La adquisición, procesamiento y análisis de bioseñales como el electrocardiograma (ECG) la electromiografía (EMG), la actividad electrodérmica (EDA), la electroencefalografía (EEG) y la fotopletismografía (PPG) permiten crear biomarcadores útiles para detectar el estrés. Estas señales, relacionadas con el sistema nervioso simpático (SNS), aportan información clave para identificar respuestas de estrés en el organismo [2].

En entornos controlados, el monitoreo continuo de señales fisiológicas, como la variabilidad cardíaca y la conductancia de la piel, permite detectar la activación del sistema nervioso simpático ante situaciones de estrés [2]. Integrar un análisis de señales de ECG, EMG y EDA con cuestionarios y entrevistas psicológicas minimiza el sesgo asociado a la subjetividad en la percepción individual [4].

Adicionalmente, la integración de herramientas como el aprendizaje automático, una rama de la inteligencia artificial que utiliza algoritmos para analizar datos, identificar patrones complejos y automatizar decisiones, ha transformado el diagnóstico de enfermedades y la predicción de resultados clínicos. Combinado con técnicas de procesamiento de señales, ha mejorado notablemente la atención médica [5].

En el ámbito de la psicología, el aprendizaje automático se emplea para facilitar el análisis de señales fisiológicas, como PPG, que refleja los cambios en el volumen sanguíneo asociados con la actividad cardíaca, y EDA, que mide la conductancia de la piel como un indicador de la respuesta autonómica al estrés. Algoritmos como SVM (Support Vector Machine), MLP (Multilayer Perceptron), KNN (K-Nearest Neighbors) y DT (Decision Tree) permiten identificar estados y niveles de estrés (bajo, medio, alto). Estas herramientas mejoran la evaluación precisa del estrés, potenciando las intervenciones en salud mental y la calidad de vida [2].

En la literatura predominan las técnicas de aprendizaje supervisado para identificar estados de estrés, las cuales dependen de datos etiquetados para entrenar modelos, evaluar su desempeño y ajustar parámetros [2]. Sin embargo, la obtención de etiquetas precisas durante la evaluación del estrés es un proceso lento, costoso y que requiere la intervención de expertos, como psicólogos, especialmente en contextos reales donde debe considerarse la variabilidad individual de las respuestas somáticas al estrés. Generalmente, estas etiquetas se asignan en función de la actividad y los estresores utilizados, y, en ocasiones, se complementan con autoinformes, como la Escala de Estrés Percibido (PSS), lo que permite reflejar estados binarios (estrés/no estrés) o niveles discretos (bajo, medio, alto) [6]. Si bien estos enfoques son efectivos en escenarios controlados, su dependencia de etiquetas limita su escalabilidad y aplicabilidad en contextos reales con datos fisiológicos complejos y dinámicos.

En contraste, los algoritmos de aprendizaje no supervisado, al no requerir etiquetas, los convierte en una alternativa factible para identificar patrones en las respuestas somáticas al estrés. Ofreciendo una forma autónoma de detectar estados de estrés en datos fisiológicos no etiquetados, lo que sugiere una mejor adaptación a problemas como la variabilidad individual y la identificación de respuestas fisiológicas de estrés en entornos reales.

En este contexto, el algoritmo DBSCAN puede ser un algoritmo auxiliar debido a su utilidad para identificar patrones en conjuntos de datos que, como en el caso del estrés, pueden contener ruido o estructuras complejas. A diferencia de métodos como K-means, DBSCAN no requiere definir previamente el número de clústers, ya que los determina automáticamente según la densidad de los puntos [7]. Los

https://doi.org/10.36677/rici.v3i1.25133 Enero-Julio 2025 ISSN: 2992-7447 págs. 59-71

puntos se clasifican en tres categorías: Core (núcleo), Frontera y Ruido [8, 9]. Las ventajas de DBSCAN son notables, destacándose su capacidad para identificar clústers de formas arbitrarias, manejar eficazmente el ruido y operar sin la necesidad de un número predefinido de clústers [9]. Estas características resultan especialmente útiles en la clasificación del estrés, donde las relaciones entre los datos suelen ser diversas y complejas [2].

De este modo, este estudio evalúa la eficacia del método DBSCAN en la identificación de respuestas fisiológicas al estrés en estudiantes universitarios, utilizando datos no etiquetados. El conjunto de datos incluye señales de PPG y EDA obtenidas durante pruebas controladas de inducción de estrés. El objetivo es identificar patrones fisiológicos que permitan clasificar tres niveles de estrés (bajo, medio y alto), asociados a los diferentes escenarios de la prueba de inducción. Estas pruebas están diseñadas para generar un incremento gradual del estrés en los participantes. A pesar de la ausencia de etiquetas, se espera que DBSCAN sea capaz de detectar y diferenciar patrones en las señales fisiológicas, asociando estos niveles de estrés a los escenarios de la prueba.

Metodología

Las pruebas de inducción de estrés se utilizan para generar respuestas psicológicas y fisiológicas al recrear situaciones estresantes, siendo clave para desarrollar tecnologías de detección de estrés mediante señales fisiológicas. Sin embargo, su capacidad para diferenciar de manera eficaz entre estados de estrés y no estrés sigue siendo incierta. Múltiples investigaciones suponen que estas pruebas provocan cambios fisiológicos significativos y, bajo esta premisa, han desarrollado clasificadores de estrés con alta precisión en condiciones controladas [2]. Sin embargo, su aplicabilidad en entornos reales es limitada, destacando la necesidad de un análisis más detallado de señales fisiológicas, especialmente en contextos de alta demanda cognitiva. Además, la respuesta al estrés varía según el entorno o contexto, incluso al realizar actividades idénticas, como se propone en la prueba de inducción de estrés empleada en este estudio.

Para llevar a cabo esta evaluación, se analizaron señales de PPG y EDA de 51 estudiantes licenciatura del área de ingeniería de la Universidad Autónoma del Estado de México, incluyendo 16 mujeres y 35 hombres, con edades de 19 a 30 años (edad promedio de 22 años). Cada participante realizó una única prueba de inducción de estrés, durante la cual se registraron las señales fisiológicas.

La prueba consistió en el desarrollo de tres exámenes en tres escenarios distintos, cada participante completó la prueba una única vez, resolviendo exámenes de comprensión lectora o de razonamiento lógico-matemático. Los escenarios se diferenciaron por los estresores aplicados, aunque actividad era la misma (ver Figura 1):

- Primer escenario: Sin estresores, se utilizó para establecer la línea base de las respuestas fisiológicas.
- Segundo escenario: Incluyó dos estresores psicológicos: restricción de tiempo (15 minutos) y ruido ambiental de conversaciones, reproducidas en una computadora portátil. Estos fueron seleccionados para inducir un aumento en la respuesta fisiológica al estrés. [10].
- Tercer escenario: Se añadió un estresor físico a los estresores psicológicos del segundo escenario: pequeñas descargas eléctricas inferiores a 5 mA y con una duración menor a 5 segundos al final de la prueba. Antes de este escenario, el dispositivo empleado fue calibrado para determinar el

umbral de estimulación de cada participante, aunque las descargas nunca fueron aplicadas. Estas se utilizaron únicamente como amenaza para intensificar la respuesta fisiológica, sin causar daño a los sujetos [10, 11].

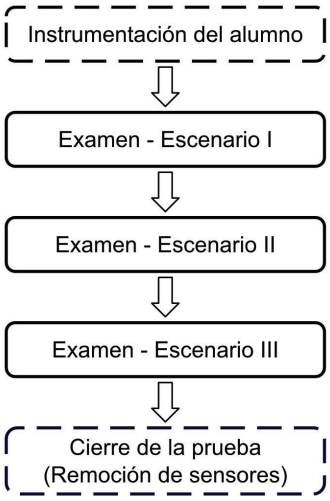


Figura 1: *Distribución general de la prueba de inducción de estrés académico.*

Durante la prueba, cada participante utilizó una banda con sensores de PPG y EDA en la frente, conectada a una computadora para registrar las señales de forma continua. Esta ubicación evitó limitar el movimiento de las manos, donde típicamente se realiza la instrumentación para adquirir estas señales, y redujo la susceptibilidad a artefactos de movimiento [2].

En cada escenario, los participantes realizaron un examen de comprensión lectora o razonamiento lógico-matemático, con un registro continuo de las señales a una frecuencia de 100 Hz para monitorear la respuesta fisiológica. Al finalizar los tres exámenes, se retiró la banda, concluyendo el registro de datos.

Tras la adquisición, las señales se analizaron en una segunda etapa que incluyó el preprocesamiento, la extracción y procesamiento de características, y la aplicación del método DBSCAN. La metodología utilizada se presenta en la Figura 2.

Durante el preprocesamiento, las señales se segmentaron y clasificaron según el escenario de adquisición, asociándose a los tres niveles de estrés esperados:

- Estrés bajo: Primer escenario sin estresores.
- Estrés medio: Segundo escenario con dos estresores psicológicos.
- Estrés alto: Tercer escenario con estresores psicológicos y físicos.

Debido a la variabilidad en la longitud y al ruido por artefactos de movimiento en las señales crudas, se aplicó un filtrado específico según el espectro de cada señal: pasa-bajas a 15 Hz para PPG [12] y a 0.15 Hz para EDA [13], eliminando componentes no deseados que afectaran la extracción de características. Las

https://doi.org/10.36677/rici.v3i1.25133 Enero-Julio 2025 ISSN: 2992-7447 págs. 59-71

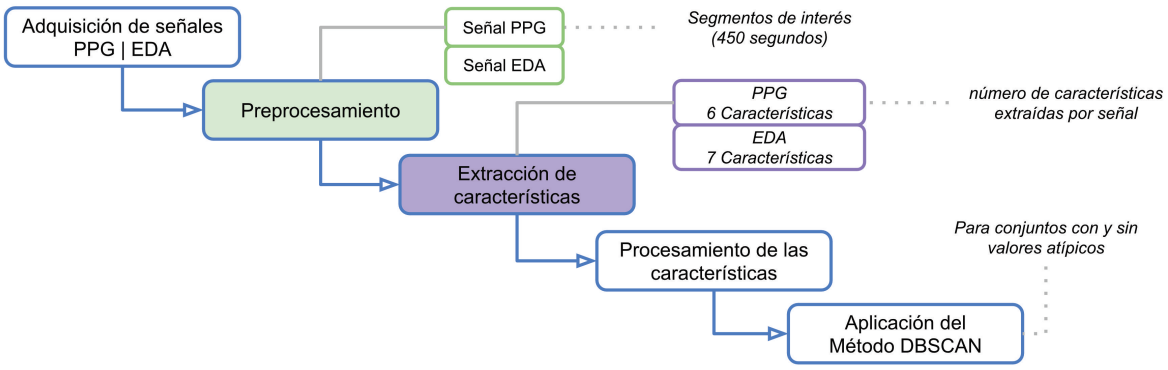


Figura 2: *Diagrama a bloques general para la aplicación del método DBSCAN en señales de PPG y EDA.*

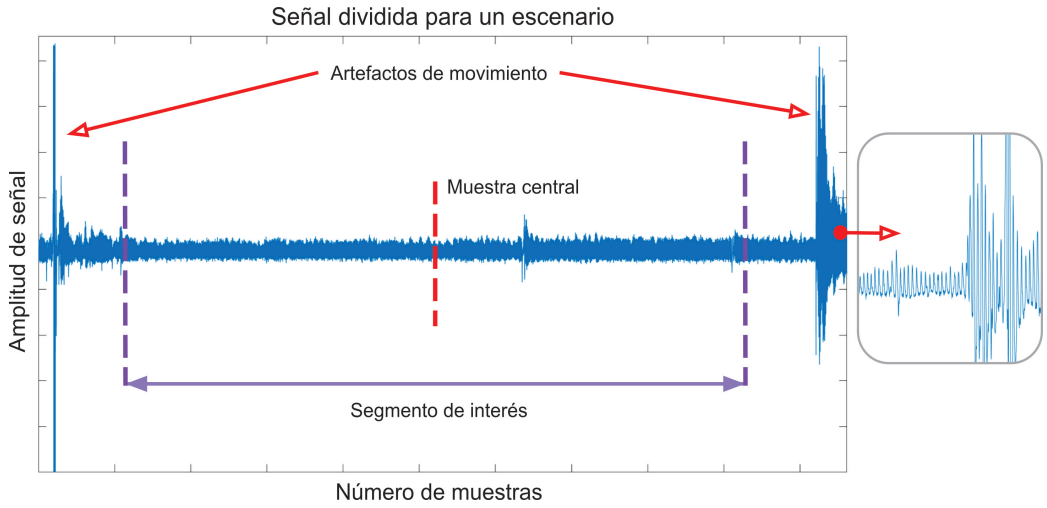


Figura 3: *Selección de los intervalos de interés en las señales para reducir los artefactos de movimiento inicial y final en cada escenario de la prueba.*

señales se segmentaron eliminando muestras iniciales y finales afectadas por ruido, obteniendo intervalos centrales con una duración uniforme de 450 segundos (ver Figura 3). Finalmente, las señales filtradas se agruparon en dos conjuntos, PPG y EDA, para los tres escenarios.

Para cada conjunto de señales, se realizó una extracción de características usando ventanas móviles ajustadas al tipo de señal. Para PPG, se emplearon ventanas de 90 segundos con un traslape de 45 segundos por segmento de interés en los tres escenarios [12]. Para EDA, las ventanas fueron de 60 segundos con un traslape de 30 segundos por segmento en cada escenario. [13].

En las señales PPG, se extrajeron seis características por ventana: promedio (C1), desviación estándar (C2) y valor cuadrático medio (RMS, por sus siglas en inglés) de los intervalos pulso-pulso (C3) así como el valor promedio (C4), desviación estándar (C5) y RMS de la frecuencia cardíaca (C6) [12]. Para las señales de EDA, se calcularon siete características por ventana: pendiente de la regresión lineal (C1), área entre la señal y su regresión lineal (C2), valor promedio (C3), desviación estándar (C4), RMS (C5), promedio de variaciones (C6) y cambio máximo (C7) [14]. Este proceso se aplicó a cada señal filtrada y en los tres escenarios estudiados. Este proceso se aplicó a cada señal filtrada y en los tres escenarios estudiados, seleccionando las características más utilizadas en la literatura.

Realizado el procesamiento de las señales y la extracción de características, se estructuraron dos conjuntos de datos que abarcan los tres escenarios analizados (ver Figura 4). El primer conjunto incluye las características extraídas de la señal de PPG, asociadas a los niveles de estrés en cada escenario, mientras que el segundo contiene las características derivadas de la señal de EDA, también vinculadas a estos niveles de estrés. Los datos fueron normalizados mediante z-score (media 0, desviación estándar 1) para garantizar una contribución uniforme de las características al análisis (ver Figura 5).

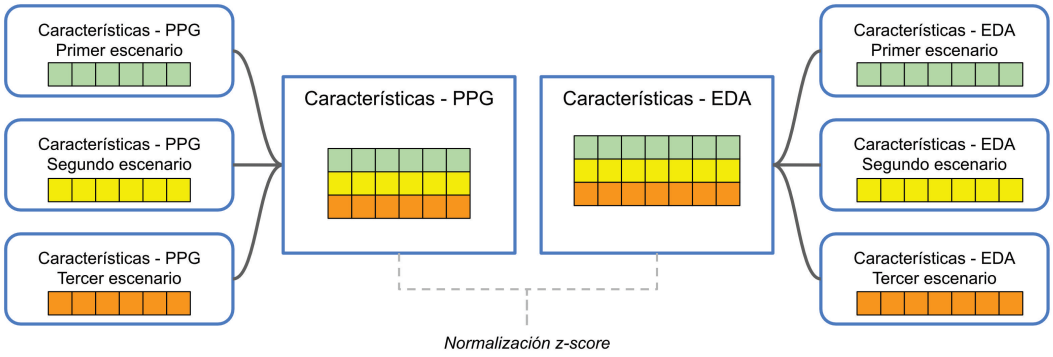


Figura 4: *Estructuración de las características de las señales de PPG y EDA para los tres escenarios de la prueba de inducción de estrés.*

Después, se utilizaron gráficos de caja para identificar datos atípicos en las características de EDA y PPG. En EDA, se detectaron valores atípicos en las características C1, C2, C4, C6 y C7, mientras que en PPG solo se observaron en C2 y C5.

Dada la dispersión de los datos, se realizaron dos pruebas con DBSCAN. En la primera, se incluyeron todas las características, abarcando tanto datos típicos como atípicos, para obtener una visión general de los patrones en las características de las señales. En la segunda, se excluyeron los valores atípicos, enfo-

https://doi.org/10.36677/rici.v3i1.25133 Enero-Julio 2025 ISSN: 2992-7447 págs. 59-71

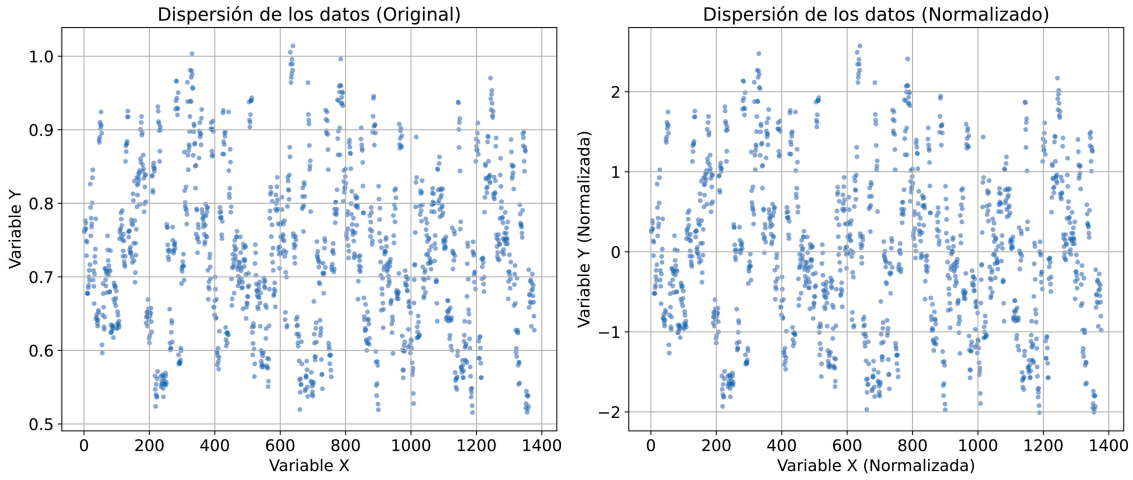


Figura 5: *Dispersión representativa de los datos antes y después de la normalización (media 0, desviación estándar 1).*

cándose en patrones más consistentes y representativos. Esto permitió comparar los clústeres generados en ambos conjuntos de datos.

Para facilitar la visualización de los resultados, se aplicó la técnica de PCA para reducir la dimensionalidad a dos componentes principales en ambos conjuntos de datos (seis características para PPG y siete para EDA). Aunque efectiva, esta técnica puede conllevar cierta pérdida de información. Se incorporó un identificador visual para los niveles de estrés esperados: 1 (bajo), 2 (medio) y 3 (alto), como se ilustra en la Figura 6. Es importante señalar que este etiquetado se empleó exclusivamente con fines visuales y no influyó en la aplicación del método DBSCAN, permitiendo observar la distribución inicial de las características según el escenario.

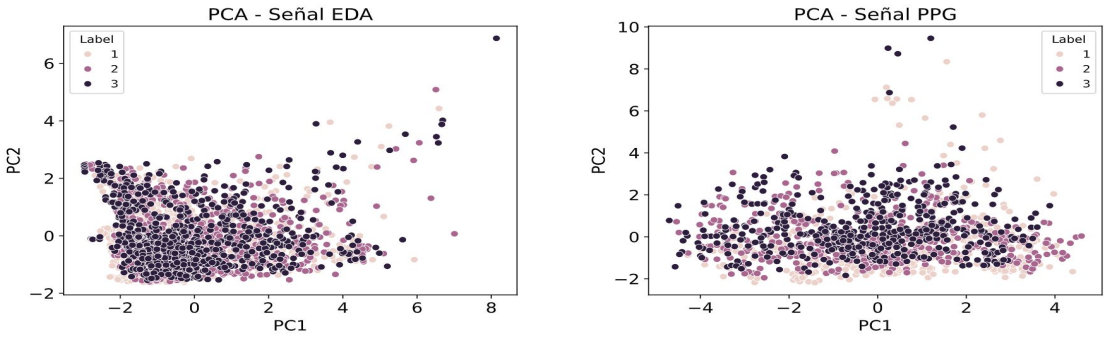


Figura 6: *Visualización de la aplicación de PCA para visualizar la dispersión de los niveles de estrés en el conjunto de características extraídas de las señales de EDA y PPG.*

Posteriormente, se implementó el método DBSCAN para identificar patrones de respuestas fisiológi-

cas asociados con los niveles de estrés en cada escenario. Para ello, fue necesario determinar los valores de los hiper-parámetros del modelo, como la distancia máxima entre dos muestras (eps) para formar un clúster o región densa, y el número mínimo de muestras requeridas para considerar un dato como punto central [7, 8].

Para determinar la distancia máxima entre dos puntos vecinos, se empleó el método de los k-ésimos vecinos más cercanos (KNN) con un valor típico de k=5, que representa menos del 0.5 % de los datos. Esto permitió calcular iterativamente la distancia máxima entre cada punto y sus cinco vecinos más cercanos. Las distancias obtenidas generaron una curva característica de crecimiento exponencial, cuyo punto de inflexión se utilizó para determinar el valor óptimo de eps, diferenciando regiones densas de puntos dispersos o ruido. De esta forma, para las características de la señal EDA, el valor eps fue 4 con todas las características y 0.2 sin datos atípicos. Para la señal PPG, los valores fueron 4.5 y 0.2 en las mismas condiciones.

Resultados y Discusión

En la Figura 7 se observa cómo el método DBSCAN aplicado al conjunto de datos completo (incluidos los valores atípicos) identificó solo una región densa, usando un mínimo de 5 vecinos y valores de eps de 4 y 4.5 para cada señal, respectivamente. Este resultado indica una diferenciación limitada en los datos, con la identificación de un único grupo en ambos casos. Esto contrasta con los grupos inicialmente identificados según los escenarios (ver Figura 6), donde se esperaba una mejor segmentación de los datos.

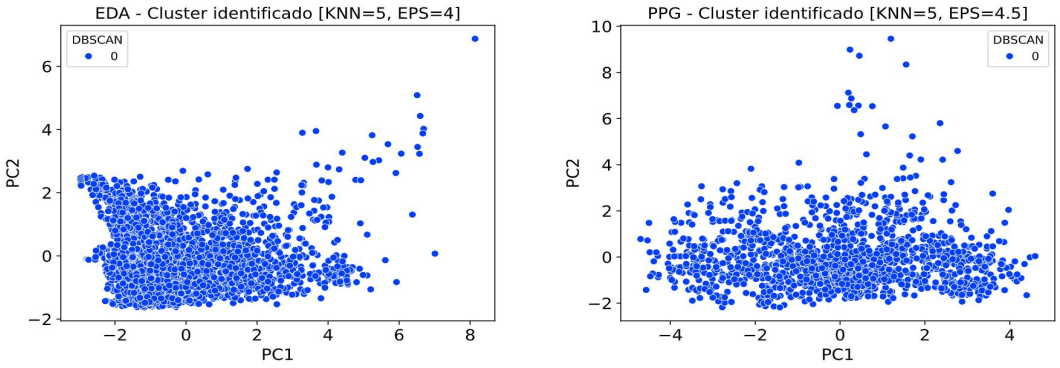


Figura 7: *Visualización de los clústeres obtenidos mediante DBSCAN aplicado a las características de las señales de EDA y PPG, considerando valores atípicos.*

Al eliminar los valores atípicos y aplicar nuevamente DBSCAN, esta vez con un valor de eps de 0.2 para ambos conjuntos y manteniendo el número de vecinos, se identificaron cuatro grupos en las características de EDA y tres en las de PPG. Sin embargo, como muestra la Figura 8 las regiones adicionales contienen pocos elementos y ruido, mientras que el volumen principal de datos permanece asociado al clúster inicial. Esto sugiere que DBSCAN, con estos hiper-parámetros, no es eficaz para formar grupos en datos asociados a respuestas fisiológicas de estrés, en comparación con otros métodos como ANOVA o técnicas de aprendizaje automático supervisado.

https://doi.org/10.36677/rici.v3i1.25133 Enero-Julio 2025 ISSN: 2992-7447 págs. 59-71

Sin embargo, estos resultados no implican que algoritmos de aprendizaje no supervisado no sean capaces de identificar grupos en este tipo de datos Un ejemplo es el método Gaussian Mixture, aplicado al conjunto de datos sin valores atípicos, que mostró mayor flexibilidad al detectar tres grupos. Los clústeres generados (ver Figura 9) evidencian parcialmente la segmentación esperada de los datos fisiológicos, aunque no coinciden completamente con los datos originales, destacando la capacidad del método para adaptarse a grupos dispersos o concentrados [15]. Así, este modelo se presenta como una alternativa prometedora para detectar respuestas de estrés en las señales de PPG y EDA.

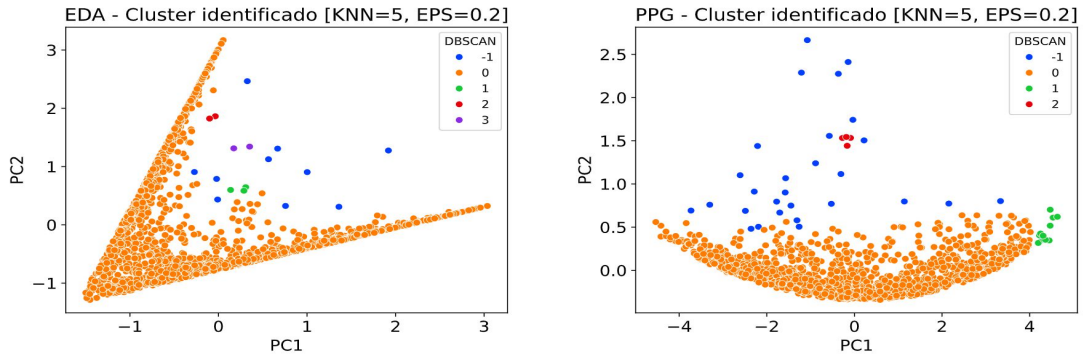


Figura 8: *Visualización de los clústeres obtenidos mediante DBSCAN aplicado a las características de las señales de EDA y PPG, no considerando valores atípicos.*

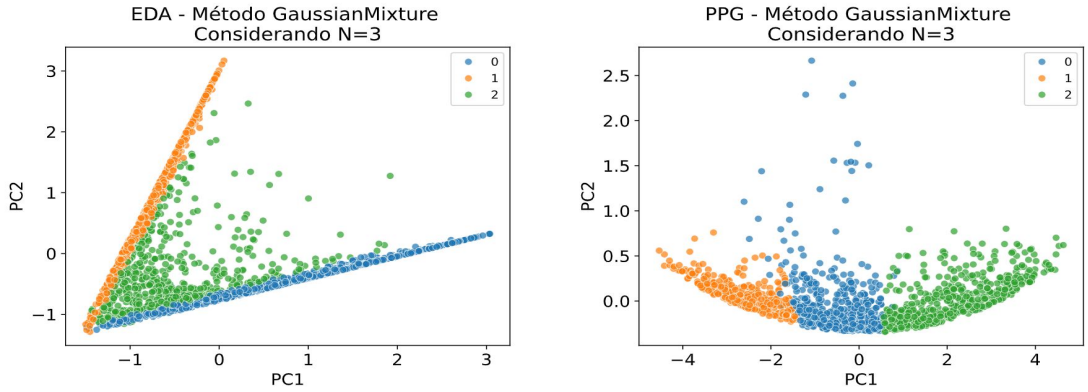


Figura 9: *Visualización de los clústeres obtenidos mediante el método GaussianMixture aplicado a las características de las señales de EDA y PPG, no considerando valores atípicos.*

Los resultados obtenidos con DBSCAN evidencian su capacidad limitada para distinguir entre diferentes niveles de respuesta al estrés utilizando datos no etiquetados. La falta de diferenciación en los clústeres sugiere que los hiper-parámetros y características seleccionados no fueron suficientes para capturar la complejidad de las respuestas fisiológicas. Aunque la eliminación de valores atípicos mejoró

parcialmente la segmentación, la prevalencia de ruido y la dispersión de los datos en clústeres con pocos elementos indican que DBSCAN no alcanzó la precisión necesaria para este tipo de análisis.

Estos resultados resaltan los desafíos del aprendizaje no supervisado en el análisis de estrés y la importancia de explorar enfoques alternativos. Por ejemplo, el modelo Gaussian Mixture, al permitir que cada grupo tenga su propia varianza, demostró ser una alternativa prometedora para analizar datos con distribuciones complejas y variadas [15]. Esta característica le proporciona la flexibilidad necesaria para adaptarse a grupos dispersos o concentrados, superando las limitaciones de DBSCAN en el análisis de señales fisiológicas.

De esta forma, los resultados obtenidos abren oportunidades para futuras investigaciones, que podrían centrarse en el ajuste de hiper-parámetros, la integración de DBSCAN con otros métodos de agrupamiento o la aplicación de técnicas de preprocesamiento para mejorar los resultados. Aunque DBSCAN no identificó los tres niveles de estrés como se esperaba, el estudio resalta la complejidad de las respuestas fisiológicas al estrés y la necesidad de enfoques más robustos, que permitan seguir explorando y perfeccionando técnicas de análisis de señales, especialmente cuando el etiquetado no es factible, en el contexto de la salud mental.

Conclusiones

Este estudio aporta evidencia sobre las limitaciones del método DBSCAN para la segmentación de señales fisiológicas para la identificación de estrés. Aunque la combinación de la reducción de dimensionalidad mediante PCA y DBSCAN es útil para el análisis exploratorio de datos no etiquetados, los resultados muestran que los clústeres identificados no son representativos de las etiquetas del conjunto de datos original, identificando un solo grupo o grupos con pocos elementos. Esto sugiere que, la hipótesis de que el método DBSCAN podría detectar patrones y diferenciar niveles de estrés en señales fisiológicas no etiquetadas no es válida.

Los hallazgos obtenidos subrayan la complejidad inherente en los datos fisiológicos, donde factores como la correcta selección de parámetros y la variabilidad individual en la respuesta al estrés dificultan el uso de métodos de aprendizaje automático. Esto es especialmente relevante en situaciones reales, donde no siempre es posible contar con un experto o realizar un etiquetado constante de los datos.

Por otro lado, los resultados obtenidos cuestionan la efectividad de DBSCAN en este contexto, destacando la necesidad de explorar métodos alternativos, como K-Medoids o Gaussian Mixture, siendo este último utilizado en el estudio, donde se lograron identificar clústeres definidos, aunque no coincidentes con los datos originales Además, es necesario optimizar la selección de los hiper-parámetros de DBSCAN, así como los modelos de aprendizaje empleados, realizando más iteraciones para determinar adecuadamente estos valores, con el fin de formar agrupamientos más consistentes.

Futuras investigaciones deben centrarse en técnicas de preprocesamiento más avanzadas, como la extracción de características en el dominio de la frecuencia, y explorar otros métodos de clustering o enfoques híbridos que combinen varios algoritmos, como aquellos que incluyan la exclusión de valores atípicos o poda, especialmente en contextos de datos no etiquetados o experimentos donde realizar un etiquetado riguroso no sea posible. En resumen, este estudio sugiere que, aunque DBSCAN y otros métodos de aprendizaje no supervisado pueden servir como punto de partida, son necesarios enfoques más robustos y flexibles para clasificar eficazmente las respuestas fisiológicas al estrés.

https://doi.org/10.36677/rici.v3i1.25133 Enero-Julio 2025 ISSN: 2992-7447 págs. 59-71

REFERENCIAS

[1] American Psychological Association. Stress in America. [Online]. Available: https://www.apa.org/news/press/releases/stress. [Accessed: Nov. 28, 2024].

[2] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis. Review on Psychological Stress Detection Using Biosignals. *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 440–460, Jan.–Mar. 2022. DOI: 10.1109/TAFFC.2019.2927337.

[3] L. M. Manosso, C. R. Gasparini, and G. Z. Réus. Definitions and Concepts of Stress. in *Glutamate and Neuropsychiatric Disorders: Current and Emerging Treatments*, Z. M. Pavlovic, Ed. Cham: Springer International Publishing, 2022, pp. 27–63. DOI: 10.1007/978-3-030-87480-3_2.

[4] M. Z. Baig and M. Kavakli. A Survey on Psycho-Physiological Analysis & Measurement Methods in Multimodal Systems. *Multimodal Technologies and Interaction*, vol. 3, no. 2, Article 37, 2019. DOI: 10.3390/mti3020037.

[5] J. A. M. Sidey-Gibbons and C. J. Sidey-Gibbons. Machine learning in medicine: a practical introduction. *BMC Medical Research Methodology*, vol. 19, no. 1, p. 64, Mar. 2019. DOI: 10.1186/s12874-019-0681-4.

[6] M. U. Mustafa, S. A. Buzdar, A. Ikhlaq, M. Nisa, S. Malik, S. Saeed, M. S. Khan, and A. Javid. Machine learning approach for multi-class stress assessment with electroencephalography signals. *Journal of Population Therapeutics and Clinical Pharmacology*, vol. 30, no. 19, pp. 1464–1478, 2023. DOI: 10.53555/jptcp.v30i19.3852.

[7] Scikit-Learn. Clustering. 2024. [Online]. Available: https://scikit-learn/stable/modules/clustering.html. [Accessed: Nov. 28, 2024].

[8] MathWorks. DBSCAN Clustering. 2024. [Online]. Available: https://la.mathworks.com/help/stats/dbscan-clustering.html#mw_0ad71796-4ae1-4620-a0e5-9ea5dcded2c6. [Accessed: Nov. 28, 2024].

[9] O. Kulkarni and A. Burhanpurwala. A Survey of Advancements in DBSCAN Clustering Algorithms for Big Data. in *2024 3rd International Conference on Power Electronics and IoT Applications in Renewable Energy and its Control (PARC)*, 2024, pp. 106–111. DOI: 10.1109/PARC59193.2024.10486339.

[10] D. H. Lamb. On the distinction between physical and psychological stressors. *Motivation and Emotion*, vol. 3, pp. 51–61, Mar. 1979. DOI: 10.1007/BF00994160.

[11] J. Rodríguez-Arce, L. Lara-Flores, O. Portillo-Rodríguez, and R. Martínez-Méndez. Towards an anxiety and stress recognition system for academic environments based on physiological features. *Computer Methods and Programs in Biomedicine*, vol. 190, p. 105408, 2020. DOI: 10.1016/j.cmpb.2020.105408.

[12] E. Mejía-Mejía and P. A. Kyriacou. Duration of photoplethysmographic signals for the extraction of Pulse Rate Variability Indices. *Biomedical Signal Processing and Control*, vol. 80, p. 104214, 2023. DOI: 10.1016/j.bspc.2022.104214.

[13] D. Leiner, A. Fahr, and H. Früh. EDA Positive Change: A Simple Algorithm for Electrodermal Activity to Measure General Audience Arousal During Media Exposure. *Communication Methods and Measures*, vol. 6, no. 4, pp. 237–250, 2012. DOI: 10.1080/19312458.2012.732627.

[14] R. Martinez, A. Salazar-Ramirez, A. Arruti, E. Irigoyen, J. I. Martin, and J. Muguerza. A Self-Paced Relaxation Response Detection System Based on Galvanic Skin Response Analysis. *IEEE Access*, vol. 7, pp. 43730–43741, 2019. DOI: 10.1109/ACCESS.2019.2908445.

[15] Scikit-Learn. Gaussian Mixture Models. 2024. [Online]. Available: `https://scikit-learn.org/1.5/modules/mixture.html`. [Accessed: Nov. 28, 2024].

IDEAS EN CIENCIAS DE LA INGENIERÍA

Uso de YOLOv8 y aumento de datos para la detección de microplásticos en agua

Using YOLOv8 and data augmentation for microplastics detection in water

Villegas-Camacho, O.[1], Francisco-Valencia, I.[2], Alejo-Eleuterio, R.[3], Del-Razo-Lopez, F.[4]*
(*fdelrazol@toluca.tecnm.mx)

Recibido: 2024-11-14 Aceptado: 2024-12-02

Resumen

Los microplásticos presentes en el agua se han convertido en un problema ambiental grave, por lo que ha surgido la necesidad de desarrollar técnicas y herramientas para su detección y posterior eliminación. En este estudio, se utilizó el modelo de visión por computadora *YOLO* (*You Only Look Once*) en su versión 8, para la detección de microplásticos en agua, el cual fue entrenado utilizando un conjunto de 500 imágenes obtenidas mediante microscopio estereoscópico, derivadas de muestras recolectadas en dos zonas del Estado de México: Valle de Bravo y el municipio de Malinalco. Adicionalmente a este conjunto de imágenes se le aplicaron diferentes técnicas de aumento de datos, incluyendo, *Resize, Crop, Flip, Brightness and Contrast, HSV, Gamma Correction, CLAHE, RGB Shift y Gaussian Blur*, con el objetivo de generar un conjunto más diverso. Los resultados obtenidos indican que *YOLO*, entrenado con el conjunto aumentado, alcanzó una precisión de 0.825 y un mAP@50 de 0.73, valores superiores a los logrados con el conjunto original sin aumentos. En conclusión, *YOLO*, en combinado con técnicas de aumento de datos, muestra ser eficaz para la detección de microplásticos, sentando las bases para su aplicación práctica en el monitoreo de cuerpos de agua.

Palabras clave: microplásticos, YOLO, aumento de datos.

[1]Tecnológico Nacional de México/Instituto Tecnológico de Toluca, Av. Tecnológico S/N, Colonia Agrícola Bellavista, Metepec, Estado de México, México. C.P. 52149.
[2]Unidad Académica Profesional Tianguistenco, Universidad Autónoma del Estado de México, Paraje El Tejocote, San Pedro Tlaltizapán, Tianguistenco, Estado de México, CP. 52640.
[3]Tecnológico Nacional de México/Instituto Tecnológico de Toluca, Av. Tecnológico S/N, Colonia Agrícola Bellavista, Metepec, Estado de México, México. C.P. 52149.
[4]Tecnológico Nacional de México/Instituto Tecnológico de Toluca, Av. Tecnológico S/N, Colonia Agrícola Bellavista, Metepec, Estado de México, México. C.P. 52149.

Abstract

The presence of microplastics in water has become a serious environmental issue, creating the need to develop techniques and tools for their detection and subsequent removal. In this study, the computer vision model *YOLO (You Only Look Once)* version 8 was used for the detection of microplastics in water. The model was trained using a dataset of 500 images obtained through a stereoscopic microscope from water samples collected in two areas of the State of Mexico: Valle de Bravo and the municipality of Malinalco. Additionally, various data augmentation techniques were applied to this dataset, including *Resize, Crop, Flip, Brightness and Contrast, HSV, Gamma Correction, CLAHE, RGB Shift, and Gaussian Blur*, to generate a more diverse dataset. The results indicate that YOLO, trained with the augmented dataset, achieved a precision of 0.825 and a mAP@50 of 0.73, values higher than those obtained with the original dataset without augmentation. In conclusion, YOLO, combined with data augmentation techniques, proves to be an effective tool for detection of microplastics, laying the foundation for its practical application in monitoring bodies of water.

Keywords: microplastics, YOLO, data augmentation.

Introducción

Los plásticos son polímeros orgánicos derivados del petróleo, entre los que destacan el polietileno, el polipropileno y el polietileno tereftalato, siendo estos los más comúnmente utilizados a nivel mundial. Estos polímeros, al entrar en contacto con el medio ambiente, se degradan y producen microplásticos y nanoplásticos, los cuales son partículas o fragmentos sólidos no solubles en agua, con una longitud menor a 5 mm para los primeros y menor a 1 μm para los segundos [1, 2]. Actualmente, los microplásticos están ampliamente distribuidos en el ambiente acuático, donde pueden permanecer por cientos o incluso miles de años, representando un grave riesgo para la vida humana y los ecosistemas [3, 4].

La presencia de microplásticos en los cuerpos de agua se debe principalmente a actividades y negligencias humanas. Los desechos provenientes de hogares, industrias y otras fuentes pueden llegar directamente al sistema marino o a cuerpos de agua dulce, liberando microplásticos que estos sistemas no son capaces de eliminar [2].

Iri et al. [5] estiman que para 2025, la acumulación de microplásticos en los océanos alcanzará cientos de millones de toneladas y Sarker et al. [3] indican, que, según las proyecciones del Foro Económico Mundial, los niveles de contaminación por microplásticos se duplicarán para 2030 y, para 2050 el peso total de los plásticos en el océano superará al de los peces.

Los microplásticos han sido encontrados en diversos organismos, incluidos peces, invertebrados y zooplancton. Se acepta ampliamente que la ingesta y la adherencia son la principal vía por la que los animales incorporan microplásticos. Issac et al. [2] establecen tres tipos de efectos nocivos para las especies acuáticas relacionados con los microplásticos:

- **Efectos fisiológicos derivados de la ingestión**: la ingesta de microplásticos reduce el desarrollo y altera los hábitos alimenticios en las especies que los consumen.

https://doi.org/10.36677/rici.v3i1.25148 Enero-Julio 2025 ISSN: 2992-7447 págs. 72-87

- **Reacciones tóxicas por la liberación de sustancias peligrosas:** los microplásticos tienen la capacidad de acumular y liberar en el agua contaminantes orgánicos peligrosos, como el DDT, los éteres de difenilo polibromados y otros aditivos químicos utilizados en su fabricación los cuales pueden filtrarse hacia los tejidos del organismo de quienes los consumen, provocando bioacumulación y cambios en las funciones biológicas.

- **Reacciones nocivas debido a contaminantes adsorbidos involuntariamente por los microplásticos**: debido a sus propiedades, los microplásticos tienden a adsorber contaminantes en su superficie, convirtiéndose en vectores de sustancias dañinas para las especies marinas. Entre los contaminantes más comunes están los hidrocarburos aromáticos policíclicos, los bifenilos policlorados, el dicloro difenil tricloroetano, pesticidas organohalogenados, hexaclorociclohexanos y bencenos clorados, que son tóxicos y pueden alterar los sistemas hormonales de los animales.

Los microplásticos pueden ingresar al organismo humano a través de la cadena alimentaria, especialmente mediante el consumo de organismos marinos como mariscos y peces, o a través de agua potable, afectando principalmente el tracto gastrointestinal. Además, estos contaminantes pueden ser inhalados a través del aire, ingresando por las vías respiratorias y, una vez dentro del cuerpo, afectar diversos órganos, tejidos y sistemas. La presencia de microplásticos en heces humanas evidencia su incorporación en el organismo [6, 7].

Estudios han reportado daños potenciales en seres humanos similares a los observados en animales, tales como respuestas inflamatorias, especialmente en el intestino, efectos citotóxicos en células cerebrales y epiteliales, estrés oxidativo, alteraciones en la respuesta inmunológica y trastornos metabólicos, entre otros [6, 7].

Los microplásticos se han convertido en contaminantes emergentes de creciente preocupación, lo que hace urgente el desarrollo de tecnologías avanzadas para su detección y eliminación, con el fin de reducir su presencia en el medio ambiente [8].

Entre las técnicas para la detección de microplásticos se encuentran la inspección visual, el pesaje, la medición de absorbancia, turbidez, carbono orgánico total, generalmente combinadas con la espectroscopía infrarroja por transformada de Fourier (FTIR). Estas técnicas se aplican comúnmente a los tipos de microplásticos más comunes, como el poliestireno, el polietileno, el cloruro de polivinilo y el tereftalato de polietileno [8].

Actualmente, las técnicas basadas en inteligencia artificial y visión por computadora han ganado popularidad en la detección de microplásticos, ya que superan varias limitaciones de los métodos tradicionales. Estos métodos convencionales suelen requerir largos tiempos de procesamiento, presentan altos costos, tienen una baja precisión y dependen de equipos costosos y personal especializado. En contraste, las nuevas técnicas de inteligencia artificial y visión por computadora ofrecen soluciones más eficientes y accesibles para enfrentar estos desafíos[3, 9, 10].

En este estudio se explora el uso de la técnica de detección de objetos mediante el modelo de visión por computadora YOLO y cómo el aumento de datos mejora su rendimiento para la detección de microplásticos en muestras de agua. Este artículo, está estructurado de acuerdo con las siguientes secciones:

- **Trabajo Relacionados**: donde se describen el estado del arte relacionados con la detección y eliminación de microplásticos.

- **Conjunto de datos**: se describen las fuentes y características de las muestras de agua utilizadas, así como la generación de las imágenes de microscopio empleadas para entrenar el modelo.
- **Aumento de datos**: se explican las técnicas empleadas para enriquecer y diversificar el conjunto de imágenes, con el objetivo de mejorar la precisión del modelo.
- **YOLO**: se describe el funcionamiento del modelo YOLO, especializado en la detección de microplásticos.
- **Experimentación**: se detalla el proceso realizado para ajustar y evaluar el rendimiento del modelo.
- **Análisis de resultados**: se presentan y analizan los resultados obtenidos.
- **Conclusión**: se reflexiona sobre el beneficio de combinar el aumento de datos con el modelo YOLO para la detección de microplásticos en el agua.

Trabajos relacionados

La contaminación por microplásticos se ha consolidado como un desafío ambiental crítico. Recientes investigaciones han explorado metodologías avanzadas para la identificación y caracterización de microplásticos, utilizando sistemas automatizados y técnicas de aprendizaje profundo, como las redes neuronales convolucionales (CNN). Estas herramientas permiten segmentar y clasificar microplásticos en imágenes microscópicas con distintos niveles de precisión. Por ejemplo, el modelo Mask R-CNN ha reportado métricas F1-Score del 61 % y 76 % en segmentación y clasificación de microplásticos [11]. De manera similar, estudios en aguas urbanas de China emplearon redes neuronales preentrenadas, como UNet, UNet2plus y UNet3plus, logrando precisiones de hasta el 91 % [12]. Estas metodologías no solo reducen el tiempo y esfuerzo del análisis manual, sino que también contribuyen a la estandarización de la identificación de microplásticos a través de imágenes.

Respecto al uso de YOLOv8 en la detección de residuos plásticos, destaca el trabajo de Kryvenchuk y Marusyk [10] donde se exploran múltiples configuraciones de modelos YOLO entrenados con el dataset Plastic in River de Kili Technology. Este conjunto de datos está compuesto por 4259 imágenes de alta resolución categorizadas en cuatro clases: "bolsas de plástico", "botellas de plástico", .ºtros desechos plásticosz "desechos no plásticos". Los resultados obtenidos muestran que YOLOv8 alcanza una tasa de precisión cercana al 80 % tras entrenar el modelo durante 500 épocas.

Sarker et al. [3] desarrollaron un sensor de cámara para la detección de microplásticos en movimiento, midiendo su tamaño y velocidad mediante modelos YOLOv5 combinados con DeepSORT para seguimiento. El modelo YOLOv5n fue entrenado durante 300 épocas con una tasa de aprendizaje de 0.01, logrando precisiones del 97 % en laboratorio y 96 % en campo. Para construir el conjunto de datos, se recopilaron 484 videos que abarcan ocho distancias (190-330 mm) y cuatro tamaños de microplásticos (2-5 mm). De estos videos se extrajeron fotogramas redimensionados a 1280 × 720 píxeles, generando 17,069 imágenes etiquetadas.

Respecto a la eliminación de microplásticos, la revisión realizada por Padervand et al [13]. presenta diversos métodos siendo los procesos de sorción y filtración los más destacados, seguidos, en menor

https://doi.org/10.36677/rici.v3i1.25148 Enero-Julio 2025 ISSN: 2992-7447 págs. 72-87

medida, por el uso de lodos activados. También se resaltan la degradación fotocatalítica, la electrocoagulación y la aglomeración, aunque estas requieren combinarse con otras técnicas para aumentar su efectividad. Entre las técnicas menos convencionales se encuentran la degradación biológica y la ingestión por organismos, como la almeja gigante del Mar Rojo.

Conjunto de Datos

En esta sección se describe el proceso de obtención de imágenes, los lugares de recolección de las muestras de agua, el procedimiento de generación de imágenes y, finalmente, el etiquetado de estas.

Obtención de Muestras de Aguas

En esta subsección se describe el proceso de obtención de las muestras de agua, las cuales se utilizaron para generar las imágenes empleadas en el entrenamiento de YOLOv8.

Las muestras de agua se recolectaron de dos sitios diferentes:

- La presa Miguel Alemán, también conocida como presa Valle de Bravo, ubicada en el municipio de Valle de Bravo en el Estado de México, México.
- El Santuario del Señor de Chalma, ubicado en el municipio de Malinalco, Estado de México, México.

En cada una de estas localidades se seleccionaron entre 7 y 9 puntos de muestreo, abarcando tanto zonas turísticas como no turísticas. En cada punto, se recolectaron tres litros de agua, almacenados en contenedores de cristal. Estos contenedores fueron transportados a laboratorio, donde, mediante de una bomba de succión, el agua fue filtrada a través de membranas (ver Figura 1) con orificios no mayores a 0.45 micras. Los microplásticos y otros elementos presentes en el agua quedaron retenidos en estos filtros de membrana para su posterior análisis.

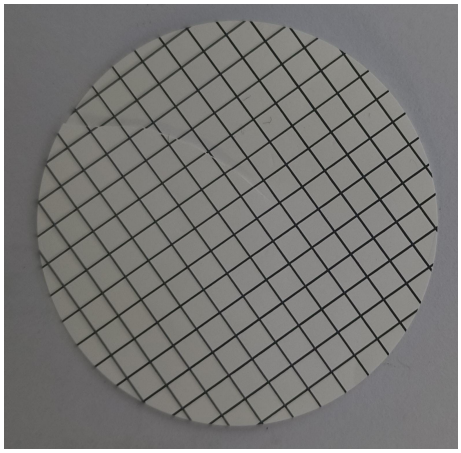


Figura 1: *Membras Filtro utilizadas para retener los microplásticos de las muestras de agua.*

Estas muestras de agua pueden contener los siguientes tipos de microplásticos: polietileno tereftalato (PET), polietileno de alta densidad (HDPE), cloruro de polivinilo (PVC), polietileno de baja densidad (LDPE), polipropileno (PP) y poliestireno (PS).

Generación de Imágenes
Las membranas fueron analizadas mediante un microscopio estereoscópico digital modelo VE-S5C de la marca Velab. A través de una inspección visual, un experto examinó las membranas por secciones y localizó los microplásticos atrapados en estos filtros. Una vez identificados uno o varios microplásticos, se capturó una imagen de la sección correspondiente y se guardó en formato TIFF, con un tamaño de 1016 por 760 píxeles o de 2040 por 1528 píxeles. Como resultado, se obtuvieron 500 imágenes con microplásticos. Adicionalmente, se generaron 504 imágenes sin microplásticos; para ello, se utilizaron las membranas filtro sin materiales sólidos retenidos, que fueron analizadas con el mismo microscopio VE-S5C y guardadas en el mismo formato y tamaño.

Etiquetado de Microplásticos
Las imágenes obtenidas se procesaron mediante el software labelme[1] para etiquetar los microplásticos presentes, con el fin de entrenar y validar el modelo YOLO. Cada microplástico en las imágenes fue etiquetado individualmente, asegurándose de cubrirlo por completo y de que la etiqueta se ajustara lo más posible a su contorno. Finalmente, y como resultado, se generó un archivo JSON para cada imagen, que contiene la información de la localización de cada microplástico en ella. Cabe mencionar que solo se utilizó una única etiqueta, identificada como *MP*, para englobar todos los tipos posibles de microplásticos contenidos en las imágenes. La Figura 2 muestra un ejemplo de una imagen con los microplásticos etiquetados.

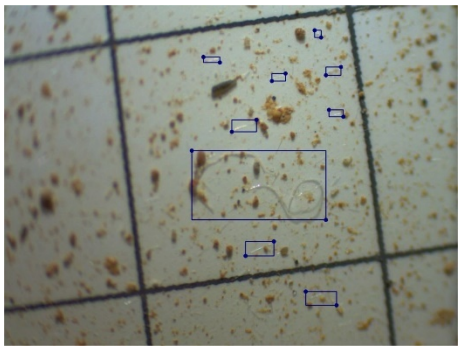


Figura 2: *Imagen de microscopio de una sección de una membrana con 9 microplásticos etiquetados por un experto.*

Del proceso anterior se obtuvieron 500 imágenes con plásticos etiquetados a las que se les denomina imágenes positivas. Para las 504 imágenes generadas con membranas limpias, que no contienen microplásticos, se crearon archivos JSON sin localización de microplásticos. A estas imágenes se las denomina imágenes negativas.

[1] https://github.com/wkentaro/labelme

https://doi.org/10.36677/rici.v3i1.25148 Enero-Julio 2025 ISSN: 2992-7447 págs. 72-87

Aumento de Datos

De acuerdo con Kumar et al. [14], las redes neuronales profundas, como YOLO, dependen de una gran cantidad de datos para entrenarse de manera efectiva. Cuando los datos son insuficientes, como en nuestro caso, donde se cuenta con solo 500 imágenes positivas (imágenes que contienen microplásticos), estas redes son susceptibles al sobreajuste. Esto significa que el modelo funciona bien en el conjunto de entrenamiento, pero falla en el conjunto de prueba. El aumento de datos es una solución a este problema y consiste en aplicar una serie de transformaciones a las imágenes con el fin de aumentar su diversidad. Estas transformaciones pueden incluir transformaciones geométricas, transformaciones en el espacio de color, aplicación de filtros con ventanas de convolución (kernel), mezcla de imágenes, borrado aleatorio, transformaciones en el espacio de características, así como técnicas más avanzadas, como redes generativas adversariales o meta-aprendizaje, entre otras [14]. Estas transformaciones pueden aplicarse antes o durante el entrenamiento. En este documento, nos referimos al *aumento de datos offline* cuando las transformaciones se aplican antes del entrenamiento, y al *aumento de datos online* cuando se realizan durante el proceso de entrenamiento.

En el contexto de los microplásticos, la aplicación de técnicas de aumento de datos puede simular condiciones que podrían ocurrir en el mundo real, como microplásticos en diversas posiciones, colores, tamaños, formas e incluso bajo diferentes escenarios de iluminación. Esto permitirá que YOLOv8 pueda mejorar su capacidad de generalización, logrando una detección más precisa de microplásticos en agua en distintos entornos.

Para nuestro conjunto de datos, aplicamos aumento de datos offline utilizando únicamente transformaciones geométricas, cambios en el espacio de color y aplicación de una ventana de filtrado:

1. **Resize**: es una de las transformaciones más básicas, la cual consiste en redimensionar la imagen a un tamaño mayor o menor. En el caso de las imágenes de microplásticos, se redimensionaron a un tamaño estandar de 640 x 640 píxeles.

2. **Crop**: este aumento consiste en recortar una parte de la imagen original, lo cual es útil cuando existen áreas de interés en la imagen; sin embargo, se corre el riesgo de perder información relevante [14]. En nuestro caso, se recortó, de manera aleatoria, una sección cuadrada de la imagen original con un tamaño aleatorio entre 400 y 500 píxeles por lado; posteriormente, la sección recortada se redimensionó a 640 x 640. Este aumento se implementa con el propósito de que el aprendizaje de YOLOv8 se enfoque en partes o secciones de los microplásticos, en lugar de depender exclusivamente de su forma completa, lo que fomenta una mayor capacidad de generalización del modelo. Además, al redimensionar las imágenes recortadas, las coordenadas de las cajas contenedoras se ajustan automáticamente, lo que permite a YOLOv8 detectar microplásticos en diferentes áreas de la imagen.

3. **Flip**: este aumento consiste en voltear la imagen horizontal o verticalmente, lo que genera variaciones en la orientación de los microplásticos. Esto permite que YOLOv8 mejore su capacidad de generalización y desempeño en escenarios reales, donde los microplásticos pueden adoptar diferentes posiciones y orientaciones.

4. **Brightness y Contrast**: transformaciones que consisten en ajustar el brillo y el contraste de la imagen, aumentando o disminuyendo sus valores. En las imágenes de microplásticos, el brillo y el

contraste se modificaron de manera aleatoria. Estos aumentos se implementaron con el objetivo de simular condiciones de iluminación provenientes de fuentes externas, como las lámparas utilizadas en los laboratorios donde se analizan las muestras de agua

5. **Hue, Saturation, Value (HSV)**: aumento que modifica los canales de Tono (Hue), Saturación (Saturation) y Valor (Value) en la imagen. Para las imágenes, el tono se ajustó de forma aleatoria con valores entre ± 20 grados, y la Saturación y Valor se modificaron ± 30 y ± 20 unidades respetivamente. Similar al aumento anterior, este se aplicó para simular condiciones de iluminación que pueden presentarse en escenarios reales.

6. **Gamma Correction**: la corrección gamma, es una operación no lineal utilizada para ajustar los valores de luminancia o estímulos de color de sistemas de imágenes, siendo útil para simular distintas condiciones de iluminación o corregir las características de visualización [15]. Para las imágenes de microplásticos, se aplicó una corrección gamma con un límite inferior de 80 y un límite superior de 120. De igual forma que los aumentos anteriores, este aumento está diseñado para simular condiciones de iluminación, por ejemplo, las originadas por las diferentes configuraciones del microscopio utilizado.

7. **Contrast Limited Adaptive Histogram Equalization (CLAHE):** transformación diseñada para mejorar el contraste de una imagen, produciendo una ecualización más equilibrada y evitando la sobre-amplificación del contraste en áreas con un contraste inicial bajo [15]. Para las imágenes, se aplicó un límite de contraste de 4 unidades y se utilizó un tamaño de mosaicos de 8x8. Este aumento se utilizó para simular imágenes de microplásticos capturadas bajo niveles elevados de iluminación.

8. **RGB Shift**: aumento que modifica los valores de los canales de Rojo (Red), Verde (Green) y Azul (Blue) de la imagen original, introduciendo variaciones de color. Para las imágenes los tres canales se modificaron de manera aleatoria en ± 20. Este aumento se utilizó para que YOLOv8 sea capaz de detectar microplásticos de distintos colores, como ocurre en escenarios reales.

9. **Gaussian Blur**: este aumento aplica un desenfoque en la imagen utilizando un filtro Gaussiano, lo que produce una imagen más borrosa. Para las imágenes de microplásticos, se usaron filtros Gaussianos con un tamaño de kernel aleatorio entre 5 y 7. Este aumento fue pensando para simular condiciones de desenfoque comúnmente provocadas por diferentes configuraciones del microscopio o por su manipulación durante el uso por parte del experto.

La Figura 3 muestra cada uno de los aumentos mencionados anteriormente aplicados a una imagen de microplásticos.

Estos aumentos se aplicaron a las imágenes con microplásticos (tanto positivas como negativas) utilizando la librería Albumentations [15]. Para lograr un conjunto diverso de imágenes de microplásticos y simular imágenes realistas, los aumentos se aplicaron con una probabilidad de 0.5, a excepción de CLAHE, que se aplicó con una probabilidad de 0.1 debido a los cambios drásticos que puede producir en las imágenes, y de Crop, que se aplicó con una probabilidad de 0.75. Cada aumento, con sus respectivas probabilidades, se aplicó 10 veces a cada imagen, generando así 10 imágenes sintéticas por cada imagen

https://doi.org/10.36677/rici.v3i1.25148 Enero-Julio 2025 ISSN: 2992-7447 págs. 72-87

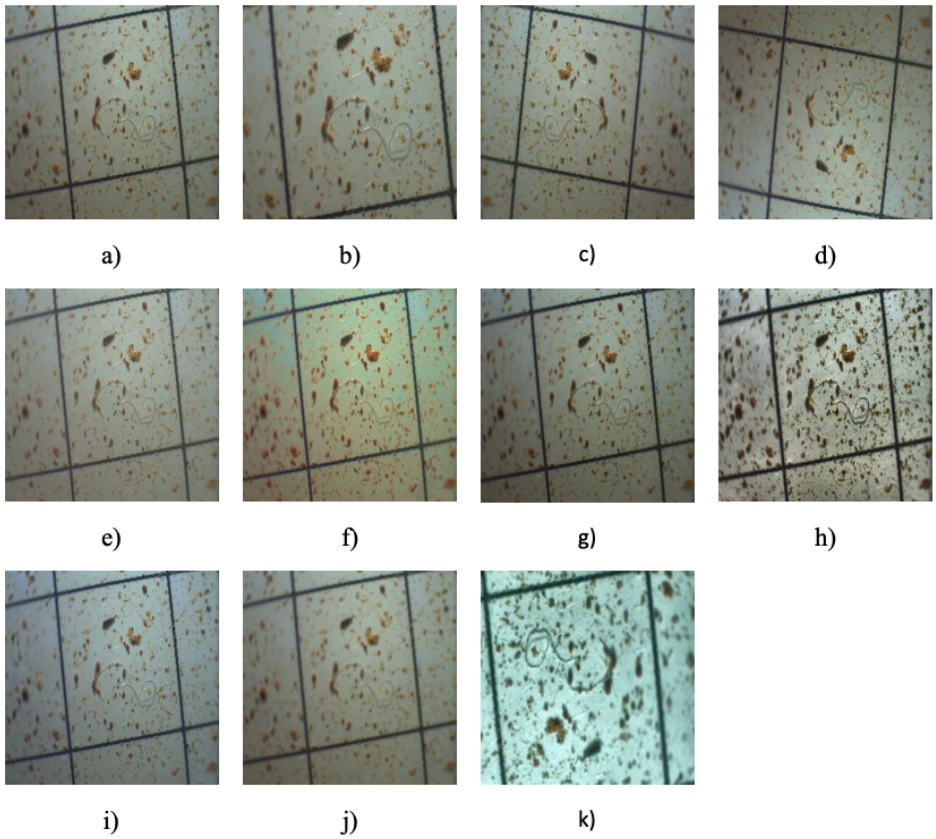


Figura 3: *Aumento de datos aplicados a las imágenes de microplásticos. a) Resize, b) Crop, c) Horizontal Flip, d) Vertical Flip, e) Brightness y Contrast, f) HSV, g) Gamma Correction, h) CLAHE, i) RGB Shift, j) Gaussian Blur y k) imagen con todos los filtros aplicados.*

original. Como resultado, el conjunto de datos final consta de 5000 imágenes positivas y 5040 imágenes negativas.

You Look Only Once

La visión por computadora, también conocida como visión artificial o visión computacional, es un campo del aprendizaje profundo cuyo objetivo es proporcionar simulaciones eficientes del sistema de visión humana a las computadoras [16]. Este campo se compone de tres subcampos: la *clasificación de imágenes*, la *detección de objetos* y la *segmentación de objetos*. El enfoque de este estudio se centra en la detección de objetos, que, a diferencia de la clasificación de imágenes, no solo determina si una imagen contiene uno o varios objetos de interés (en este caso, los microplásticos), sino que también determina su ubicación dentro de la imagen.

En la detección de objetos, los métodos se clasifican en dos tipos [16, 17]:

- **De dos etapas**: en este enfoque, el proceso de detección se divide en dos partes. Primero, una búsqueda selectiva genera propuestas de regiones, y luego redes neuronales convolucionales las clasifican y refinan para obtener la salida final. Este método ofrece una alta precisión, pero a costa de un alto costo computacional. Ejemplos de estos métodos incluyen R-CNN (Redes Neuronales Convolucionales Basadas en Regiones), Fast R-CNN y Faster R-CNN

- **De una etapa**: Estos métodos combinan ambas partes en un solo paso, lo que permite obtener la salida reduciendo la demanda computacional. Ejemplos de estos métodos incluyen SSD (Single-Shot MultiBox Detector) y YOLO.

Desde su lanzamiento en 2015, con la primera versión de YOLO (YOLOv1) y las mejoras subsecuentes, este modelo se ha convertido en un referente en la detección de objetos y ha sido aplicado en diversos campos, como sistemas de vehículos autónomos, vigilancia, análisis deportivo, detección y clasificación de cultivos, identificación de plagas y enfermedades en plantas, detección de cáncer, sistemas de reconocimiento facial, detección y clasificación de objetos en imágenes satelitales y aéreas, planificación urbana, y monitorización ambiental, entre muchos otros casos [17].

En enero de 2023, Ultralytics lanzó la última versión de YOLO, YOLOv8 [18], el cual soporta múltiples tareas, incluyendo detección de objetos, segmentación, estimación de pose, seguimiento y clasificación. Para la tarea de detección YOLOv8 se ofrece en cinco versiones escaladas YOLOv8 nano, YOLOv8 small, YOLOv8 medium, YOLOv8 large, YOLOv8 extra-large.

Según Jocher et al. [18] YOLOv8 es una versión mejorada de YOLOv5, también desarrollado por Ultralytics. Incorpora un módulo llamado C2F (cuello de botella parcial en etapas cruzadas con dos convoluciones), que mejora la precisión en la detección. YOLOv8 utiliza un modelo sin anclas (anchor-free) para manejar de manera independiente las tareas de objetividad, clasificación y regresión, lo que contribuye a una mayor precisión. En su capa final, emplea una función de activación sigmoide para determinar la probabilidad de que una región contenga un objeto y una función softmax para calcular la probabilidad de que el objeto pertenezca a una clase específica. Además, utiliza las funciones de pérdida CIoU, DFL y entropía cruzada binaria para mejorar el rendimiento en la detección de objetos, especialmente en el caso de objetos pequeños. Para más detalles sobre el funcionamiento y la arquitectura de YOLOv8, se puede consultar [18].

https://doi.org/10.36677/rici.v3i1.25148 Enero-Julio 2025 ISSN: 2992-7447 págs. 72-87

EXPERIMENTACIÓN

Para la experimentación utilizamos YOLOv8 de Ultralytics (YOLOv8.2.78) en su versión nano. Con el fin de evaluar el impacto del aumento de datos en el rendimiento de YOLOv8, se emplearon dos conjuntos de datos: uno con las imágenes originales (*OriginalDataset*) y otro compuesto únicamente por las imágenes aumentadas (*AumentoDataSet*).

Experimentación con OriginalDataSet

En el primer experimento, el conjunto de datos OriginalDataSet, compuesto por 1004 imágenes (500 positivas y 504 negativas), se dividió en un 70 % para entrenamiento, 10 % validación y 20 % para prueba. YOLOv8 se configuró para entrenarse por 500 épocas, con un batch de 32 y un early stopping de 30 épocas, el cual detendrá el entramiento si no se observa una mejora en durante 30 épocas consecutivas. Se utilizó el optimizador SGD con una taza de aprendizaje de 0.01. Además, se desactivaron los aumentos online que YOLOv8 de Ultralytics implementa, con la excepción del aumento Mosaic el cual combina cuatro imágenes para formar una sola.

Experimentación con AumentoDataSet

Para el caso de *AumentoDataSet*, compuesto por 10040 imágenes (5000 positivas y 5040 negativas), se dividió un 80 % para entrenamiento y un 20 % para validación. De esta forma, YOLOv8 fue entrenado exclusivamente con las imágenes aumentadas, mientras que las 1004 imágenes originales se utilizaron como conjunto de prueba. Al igual que con *OriginalDataSet,* YOLOv8 se configuró para entrenarse por 500 épocas, un batch de 32 y un early stopping de 30 épocas. Se utilizó el optimizador SGD con una tasa de aprendizaje de 0.01, y se desactivaron los aumentos online, a excepción del aumento Mosaic.

ANÁLISIS DE RESULTADOS

En esta sección se analizarán los resultados del entrenamiento de YOLOv8 con los dos conjuntos, *OriginalDataSet* y *AumentoDataSet,* utilizando el conjunto de prueba de ambos. Para ello se emplearon las siguientes métricas:

- Intersection over Union (IoU): Métrica que mide la superposición entre la caja real y la caja predicha por el modelo. La métrica IoU se calcula de la siguiente manera, (Ec. (1)):

$$IoU = \frac{rea\ de\ la\ interseccin}{rea\ de\ la\ unin} \tag{1}$$

IoU se utiliza con un umbral, que representa el valor mínimo que el modelo debe alcanzar para que la predicción se clasifique como una detección correcta.

- Precision: Métrica que calcula la proporción de verdaderos positivos respecto al total de predicciones positivas, Ec. (2):

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

donde TP es el número de verdaderos positivos y FP el número de falsos positivos. Esta métrica indica que tanto de los microplásticos detectados por el modelo son realmente son microplásticos.

- Recall: Métrica que indica la proporción de verdaderos positivos detectados con respecto al total de los realmente positivos, Ec. (3):

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

donde FN es el número de falsos negativos. Esta métrica muestra la porción de microplásticos que el modelo es capaz de detectar respecto al total de microplásticos presentes en la imagen.

- Mean Average Precision (mAP): Es una de las métricas más usadas para medir el rendimiento de un modelo en detección objetos. La métrica mAP se calcula de la siguiente manera, Ec. (4):

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP(i) \tag{4}$$

donde N es el total de clases y $AP(i)$ es la presión promedio de la clase i. Para YOLOv8 de Ultralytics, la métrica mAP se presenta en dos variantes: una con un umbral de IoU de 0.5 (mAP@50) y otra con un rango de umbrales que va de 0.5 a 0.9 (mAP@90) [18].

Resultados de OriginalDataSet

Al entrenar YOLOv8 con el OriginalDataset, el early stopping se activó en la época 133, obteniendo el mejor rendimiento en la época 103. Esto indica que entrenar el modelo por más épocas no mejora su rendimiento y podría llevar al sobreajuste. Con los pesos de la época 103 se evaluó el modelo utilizando el conjunto de prueba de OriginalDataset. Los resultados (ver Tabla 1) muestran que el rendimiento de YOLOv8 con este conjunto es deficiente; la precisión revela que el modelo produce muchos falsos positivos, y el recall indica que detecta muy pocos de los microplásticos presentes en las imágenes. Esto sugiere que el modelo no es capaz de detectar los microplásticos de manera efectiva.

Resultados de AumentoDataSet

En el caso de *AumentoDataSet,* el early stopping no se activó, por lo que se completaron las 500 épocas de entrenamiento. Con los pesos obtenidos, se evaluó el modelo utilizando el conjunto de prueba de *AumentoDataSet,* los resultados se muestran en la Tabla 1. Se observa que YOLOv8 alcanza una presión de 0.82, lo que significa que gran parte de los microplásticos detectados por el modelo corresponden a microplásticos reales. Sin embargo, el recall bajo sugiere que, aunque el modelo detecta la mayoría de los microplásticos, aún no es capaz de identificar por completo todos los microplásticos en una imagen. Además, el valor de mAP@50 es de 0.73 lo que indica una buena superposición entre los microplásticos detectados y las cajas reales. Aunque con mAP@90 es bajo, el modelo podría ser viable para su uso en aplicaciones reales. Finalmente, al no activarse el early stopping, se infiere que un entrenamiento con un mayor número de épocas podría mejorar el rendimiento del modelo.

https://doi.org/10.36677/rici.v3i1.25148 Enero-Julio 2025 ISSN: 2992-7447 págs. 72-87

Tabla 1: Resultados del entrenamiento de YOLOv8 con los conjuntos OriginalDataSet y AumentoDataSet

Conjunto de Datos	**Precision**	**Recall**	**mAP@50**	**mAP@90**
OriginalDataSet	0.571	0.446	0.463	0.245
AumentoDataSet	0.825	0.602	0.730	0.449

Análisis de resultados.
Además, ambos modelos se probaron con 86 imágenes nuevas, con un umbral de confianza mínimo de 0.5, generadas a partir de las mismas muestras de agua utilizadas en *OriginalDataset* y *AumentoDataset*, todas contienen microplásticos y presentaban diferentes tamaños y formatos. Estas imágenes no forman parte de ningún conjunto de entrenamiento ni de validación, no están etiquetadas ni han sido sometidas a aumentos. Las imágenes están pensadas para mostrar el comportamiento de los modelos si fuesen usadas en un laboratorio con imágenes completamente nuevas.

La Figura 4 compara el desempeño de ambos modelos en tres imágenes seleccionadas al azar de las 86 nuevas, referidas como muetras. En columnas se muestran las etiquetas realizadas por el experto (izquierda), las de modelo entrenado con el *OriginalDataset* (centro) y del modelo entrenado con el *AumentadoDataset* (derecha), junto con los niveles de confianza de YOLOv8. En la primera muestra, el modelo con *OriginalDataset* no detecta el microplástico identificado por el experto, mientras que el modelo con *AumentadoDataset* sí lo hace, demostrando la efectividad de los aumentos. En la segunda muestra, el modelo con *AumentadoDataset* no detecta los microplásticos pequeños que el experto identificó, mientras que el modelo con *OriginalDataset* detecta uno, pero genera varios falsos positivos, esto puede debe ser a que los microplásticos presentes en esta muestra son demasiados pequeños por lo que YOLOv8 no logra identificarlos y los aumentos no fueron beneficiosos. En la tercera muestra, ambos modelos identifican correctamente los microplásticos del experto, pero el modelo con *AumentadoDataset* lo hace con mayor confianza.

CONCLUSIONES

En este artículo se mostró cómo el modelo YOLOv8 de Ultralytics se beneficia del uso de aumento de datos para la detección de microplásticos en agua. Al aplicar técnicas de aumento de datos como Resize, Crop, Flip, Brightness and Contrast, HSV, Gamma Correction, CLAHE, RGB Shift y Gaussian Blur a las imágenes de microplásticos, YOLOv8 alcanzó una precisión de 0.825 y un mAP@50 de 0.73, en contraste con los valores de precisión de 0.571 y mAP@50 de 0.463 obtenidos sin los aumentos de datos.

Por lo tanto, se considera que el modelo YOLOv8 presentado en este estudio, junto con técnicas de aumento de datos, puede ser aplicado para la detección de microplásticos en el agua, contribuyendo a su identificación y potencial eliminación del medio ambiente.

Durante la experimentación, se observó que el modelo aún puede mejorar su rendimiento, mediante un mayor número de épocas de entrenamiento, ya que el early stopping de 30 épocas no se activó durante las 500 épocas previstas. Se cree que existen dos líneas para seguir optimizando el modelo: ampliar la cantidad de zonas de muestreo de las muestras de agua y aplicar técnicas de aumento de datos más avanzadas.

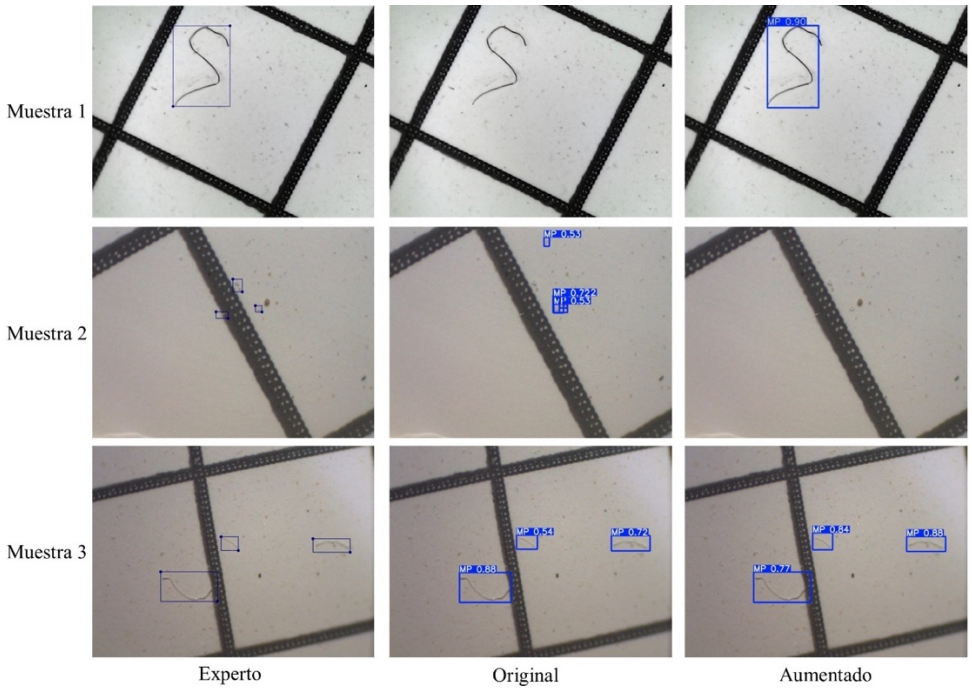


Figura 4: *Predicciones de los modelos de YOLOv8 en tres imágenes nuevas comparado con el etiquetado realizado por un experto.*

Las imágenes de microplásticos utilizadas para entrenar YOLOv8 se generaron a partir de muestras tomadas de dos zonas del Estado de México; incluir más zonas permitiría diversificar el conjunto de datos y mejorar la generalización del modelo. Asimismo, aplicar técnicas avanzadas de aumento de datos, como redes generativas adversariales, meta-aprendizaje, aumentos basados en características o técnicas de transferencia de estilo neuronal (Neural Style Transfer), podría mejorar la generación del modelo.

https://doi.org/10.36677/rici.v3i1.25148 Enero-Julio 2025 ISSN: 2992-7447 págs. 72-87

Declaración de contribución de autores y colaboradores

Octavio Villegas Camacho: Conceptualización, Curación de datos, Metodología, Redacción - Preparación del borrador original. **Iván Francisco Valencia**: Conceptualización, Software, Metodología, Redacción - Preparación del borrador original. **Roberto Alejo Eleuterio**: Conceptualización, Metodología, Supervisión, Redacción – revisión y edición. **Federico del Razo López**: Metodología, Supervisión, Redacción – revisión.

Referencias

[1] H. Kye, J. Kim, S. Ju, J. Lee, C. Lim, y Y. Yoon. Microplastics in water systems: A review of their impacts on the environment and their potential hazards. *Heliyon*, vol. 9, no. 3, p. e14359, 2023. DOI: 10.1016/j.heliyon.2023.e14359.

[2] M. N. Issac y B. Kandasubramanian. Effect of microplastics in water and aquatic systems. *Environmental Science and Pollution Research*, vol. 28, pp. 19544–19562, 2021. DOI: 10.1007/s11356-021-13184-2.

[3] M. A. B. Sarker, M. H. Imtiaz, T. M. Holsen, y A. B. M. Baki. Real-Time Detection of Microplastics Using an AI Camera. *Sensors*, vol. 24, no. 13, 2024. DOI: 10.3390/s24134394.

[4] M. A. B. Sarker, U. Butt, M. H. Imtiaz, y A. B. Baki. Automatic Detection of Microplastics in the Aqueous Environment. *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 768–772, 2023. DOI: 10.1109/CCWC57344.2023.10099253.

[5] A. H. Iri, M. H. A. Shahrah, A. M. Ali, S. A. Qadri, T. Erdem, I. T. Ozdur, y K. Icoz. Optical detection of microplastics in water. *Environmental Science and Pollution Research*, vol. 28, pp. 63860–63866, 2021. DOI: 10.1007/s11356-021-12358-2.

[6] Q. Zhang, E. G. Xu, J. Li, Q. Chen, L. Ma, E. Y. Zeng, y H. Shi. A review of microplastics in table salt, drinking water, and air: direct human exposure. *Environmental Science & Technology*, vol. 54, no. 7, pp. 3740–3751, 2020. DOI: 10.1021/acs.est.9b04535.

[7] Z. Yan, Y. Liu, T. Zhang, F. Zhang, H. Ren, y Y. Zhang. Analysis of Microplastics in Human Feces Reveals a Correlation between Fecal Microplastics and Inflammatory Bowel Disease Status. *Environmental Science & Technology*, vol. 56, no. 1, pp. 414–421, 2022. DOI: 10.1021/acs.est.1c03924.

[8] W. Gao, Y. Zhang, A. Mo, J. Jiang, Y. Liang, X. Cao, y D. He. Removal of microplastics in water: Technology progress and green strategies. *Green Analytical Chemistry*, vol. 3, p. 100042, 2022. DOI: 10.1016/j.greeac.2022.100042.

[9] E. K. Dal y R. Kilic. Design and Implementation of a Microplastic Detection and Classification System Supported by Deep Learning Algorithm. DOI: 10.21203/rs.3.rs-3890356/v1.

[10] Y. Kryvenchuk y A. Marusyk. Plastic Waste on Water Surfaces Detection Using Convolutional Neural Networks. Apr. 2024. [Online]. Available: https://ceur-ws.org/Vol-3668/paper13.pdf.

[11] H. Huang, H. Cai, J. U. Qureshi, S. R. Mehdi, H. Song, C. Liu, Y. Di, H. Shi, W. Yao, y Z. Sun. Proceeding the categorization of microplastics through deep learning-based image segmentation. *Science of The Total Environment*, vol. 2023, p. 165308, 896. DOI: 10.1016/j.scitotenv.2023.165308.

[12] J. Xu y Z. Wang. Efficient and accurate microplastics identification and segmentation in urban waters using convolutional neural networks. *Science of The Total Environment*, vol. 911, p. 168696, 2024. DOI: 10.1016/j.scitotenv.2023.168696.

[13] M. Padervand, E. Lichtfouse, D. Robert, y C. Wang. Removal of microplastics from the environment. A review. *Environmental Chemistry Letters*, vol. 18, no. 3, pp. 807–828, 2020. DOI: 10.1007/s10311-020-00983-1.

[14] T. Kumar, R. Brennan, A. Mileo, y M. Bendechache. Image Data Augmentation Approaches: A Comprehensive Survey and Future Directions. *IEEE Access*, vol. 4, 2024. DOI: 10.48550/arXiv.2301.02830.

[15] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, y A. A. Kalinin. Albumentations: Fast and Flexible Image Augmentations. *Information*, vol. 11, no. 2, 2020. DOI: 10.3390/info11020125.

[16] M. Hussain. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines*, vol. 11, no. 7, p. 677, 2023. DOI: 10.3390/machines11070677.

[17] J. Terven, D.-M. Córdova-Esparza, y J.-A. Romero-González. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, 2023. DOI: 10.3390/make5040083.

[18] G. Jocher, A. Chaurasia, y J. Qiu. Ultralytics YOLOv8. 2023. [Online]. Available: https://github.com/ultralytics/ultralytics. [Último acceso: Nov. 6, 2024].

IDEAS EN CIENCIAS DE LA INGENIERÍA

Detección de eventos delictivos usando Deep Learning: un caso de estudio en el TecNM/TEST

Crime events detection using Deep Learning: a case study on TecNM/TEST

Ballesteros-Reyes, L.F.[1] , Cleofas-Sánchez, L.[2]*, Guzmán-Ponce, A.[3], Yañez-Badillo, H.[4], Pineda-Briseño, A.[5]
(*laura_cs@test.edu.mx)

Recibido: 2024-11-15 Aceptado: 2024-12-09

Resumen

En 2023, se observó un incremento en la incidencia y prevalencia de actos delictivos en México, con variaciones significativas entre las diferentes entidades federativas. El Estado de México destacó particularmente por registrar un aumento en los niveles de eventos delictivos con respecto al año anterior. Los delitos más comunes incluyeron robos perpetrados mediante distintas modalidades, destacando el uso de armas blancas y de fuego, así como agresiones físicas directas. Por otro lado, la violencia en los entornos escolares en México ha mostrado tendencias alarmantes en varios tipos de agresión. De lo antes mencionado, se destaca la importancia de implementar técnicas avanzadas que permitan la identificación oportuna de hechos delictivos en entornos escolares a fin de salvaguardar la integridad de los estudiantes. Este estudio explora el uso del modelo YOLOv5s para la detección de arrebatos, armas blancas y de fuego, específicamente en el Tecnológico de Estudios Superiores de Tianguistenco utilizado imágenes digitales obtenidas a través de cámaras fotográficas. Los resultados obtenidos indican que este modelo logra clasificar incidentes que involucran el uso de armas blancas, armas de fuego y arrebatos, alcanzando un promedio de precisión media del 60 %.

Palabras clave: YOLOv5, Reconocimiento de actos violentos, Deep Learning.

[1]Tecnológico Nacional de México/Tecnológico de Estudios Superiores de Tianguistenco, Carretera Tenango, Santiago-La Marquesa 22, 52650 Santiago Tilapa, México.
[2]Tecnológico Nacional de México/Tecnológico de Estudios Superiores de Tianguistenco, Carretera Tenango, Santiago-La Marquesa 22, 52650 Santiago Tilapa, México.
[3]Departamento de Lenguajes y Sistemas Informáticos, Instituto de Nuevas Tecnologías de Imagen, Universitat Jaume I, 12071 Castelló de la Plana, España.
[4]Tecnológico Nacional de México/Tecnológico de Estudios Superiores de Tianguistenco, Carretera Tenango, Santiago-La Marquesa 22, 52650 Santiago Tilapa, México.
[5]Tecnológico Nacional de México/Tecnológico Nacional de México / Instituto Tecnológico de Matamoros, H. Matamoros, Tamps., México.

Abstract

In 2023, an increase in the incidence and prevalence of criminal acts was observed in Mexico, with significant variations across different federal entities. The State of Mexico particularly stood out for recording a rise in crime levels compared to the previous year. The most common offences included thefts committed through various methods, notably involving the use of bladed weapons and firearms, as well as direct physical assaults. Moreover, violence in school settings in Mexico has shown alarming trends in various types of aggression. From the aforementioned, the importance of implementing advanced techniques for the opportune identification of criminal events in school settings to safeguard the integrity of students is highlighted. This study explores the potential use of the YOLOv5s model for detecting snatch thefts, bladed weapons, and firearms, specifically at the Technological Institute of Higher Studies in Tianguistenco, using digital images captured by photographic cameras. The results obtained indicate that this model is capable of classifying achieves a classification of incidents involving bladed weapons, firearms, and snatch thefts, reaching an average precision of 60 %.

Keywords: YOLOv5, Recognition of violent acts, Deep Learning.

Introducción

En 2023, se registró un aumento en las tasas de criminalidad del 83 % de la población global[1], este fenómeno ha afectado significativamente a nivel internacional. Los reportes de la Interpol[2] destacaron como las principales actividades delictivas el crimen organizado, el tráfico de drogas, la ciberdelincuencia y la explotación sexual.

Lankford [1] argumentó que los tiroteos masivos y otros ataques públicos se originaron por algún tipo de violencia, donde los tiradores atacan a víctimas inocentes en espacios públicos. En Estados Unidos, estos tiradores suelen actuar solos y son autodestructivos, representando solo el 1.25 % de los casos globales. En otros países, los ataques generalmente involucran grupos, como organizaciones terroristas o milicias, que tienen como motivación la lucha y la supervivencia.

En las últimas cuatro décadas, los incidentes con armas en México han aumentado debido a la producción y el tráfico de armas desde **E**stados Unidos, impulsados por su política de 1791 que favorece el derecho a portar armas [1, 2]. El tráfico de armas hacia México ha incrementado la violencia reflejándose en los años 2011 y 2017, con 24 y 25 homicidios por cada 100,000 ciudadanos [3] (ver Figura 1).

Al respecto, México ocupa el sexto lugar mundial en mortalidad por el uso y portación de armas, con un aumento en las muertes entre los años 2015 y 2022. Las víctimas incluyen principalmente hombres (95.5 %), pero también mujeres (53.5 %) y niños. Los tiroteos ocurren en espacios públicos, calles y hogares, relacionados con el crecimiento de organizaciones delictivas y el tráfico de armas, las cuales cada año, se adquieren 213,000 armas, muchas de ellas provenientes de Estados Unidos [5].

En 2023, *México* registró 15,082 homicidios, con una tasa de 12 por cada 100,000 habitantes, de los cuales el 71.3 % fueron causados por armas de fuego, seguido por arma blanca (9.1 %) y ahorcamiento, estrangulamiento o sofocación (6.7 %) [4].

[1] https://globalinitiative.net/analysis/ocindex-2023/

[2] https://www.interpol.int/en/How-we-work/Criminal-intelligence-analysis/Our-analysis-reports

https://doi.org/10.36677/rici.v3i1.25167 Enero-Julio 2025 ISSN: 2992-7447 págs. 88-100

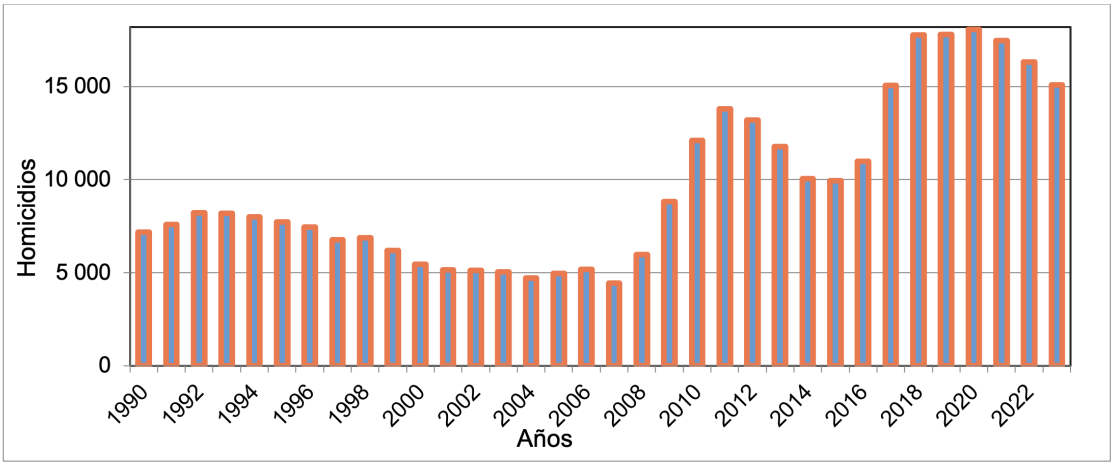


Figura 1: *Homicidios en México, entre los años 1990 y 2023 [4].*

El estudio realizado por Weigend [5] analizó que el 26 % de los incidentes violentos con armas de fuego ocurridos en escuelas y oficinas en México entre 2015 y 2022, impactan negativamente en la economía del país. La inseguridad reduce las oportunidades de inversión, empleo y desarrollo local. Asimismo, estos eventos tienen un efecto considerable en la salud mental de los mexicanos, generando un ambiente de miedo y estrés crónico.

Derivado de lo anterior, según la Encuesta Nacional de Victimización y Percepción sobre la Seguridad Pública del Instituto Nacional de Estadística y Geografía (INEGI)[3], en los últimos años, el Estado de México y el país han experimentado un aumento en los índices de violencia. En 2006, la criminalidad era el segundo problema más grave, pero entre 2012 y 2014 ascendió al primer lugar de las preocupaciones nacionales, afectando tanto a los residentes del Estado de México como al resto del país [6].

De igual modo, en 2023, el INEGI reportó 2,849 homicidios en el Estado de México, con un 85.3 % de víctimas masculinas y un 10.31 % femeninas [5]. En 2024, los homicidios por arma de fuego aumentaron un 5.8 % en comparación con 2022, una tendencia atribuida al tráfico de armas entre Estados Unidos y el norte de México. En ese mismo año, más del 70 % de los homicidios en México fueron cometidos con armas de fuego, reflejando una grave situación de seguridad, exacerbada por la proliferación de armamento ilegal en el país.

En la Tabla 1, se describen las casusas socioeconómicas del incremento de violencia en México y Estado de México entre los años 2006 y 2014. Como puede verse, en el año 2006, el desempleo agravó la inseguridad, mientras que en 2014 se observó un incremento en delitos. En términos económicos, en el 2010 se presentó una situación relativamente más estable.

La criminalidad en México ha afectado diversas áreas, incluido el sector educativo, en donde se requiere atención para prevenir escaladas de violencia como las ocurridas en Estados Unidos, en el cual la violencia escolar involucra armas blancas y otras agresiones físicas [7].

El uso de aprendizaje profundo ha mejorado significativamente en la detección de delitos. Investiga-

[3]https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2024/ENVIPE/ENVIPE_24.pdf

Tabla 1: Causas de violencia en el Estado de México y a nivel nacional en México: a.- Desempleo (México), b.- Desempleo (Estado de México), c.- Delincuencia y Crimen (México), d.- Delincuencia y Crimen (Estado de México), e.- Economía (México), f.- Economía (Estado de México), g.- Encuestados (México), h.- Encuestados (Estado de México) [3].

	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**
2014	13.5 %	11.3 %	25.7 %	29.6 %	21.3 %	13.8 %	1519	203
2012	12.2 %	13.9 %	21.7 %	18.8 %	13.4 %	9.9 %	778	101
2010	20.0 %	23.4 %	15.4 %	22.9 %	33.1 %	24.9 %	1540	201
2008	16.7 %	15.9 %	15.8 %	16.9 %	16.5 %	23.4 %	1530	201
2006	21.5 %	21.8 %	14.9 %	14.4 %	11.3 %	9.9 %	1533	202

dores han explorado diversos métodos para optimizar los protocolos de seguridad y reducir actividades ilegales [8, 9, 10]. Mukto et al [11] argumentaron que los algoritmos de aprendizaje profundo se están implementando en sistemas de monitoreo criminal en tiempo real, lo que permite identificar y responder rápidamente a incidentes delictivos.

Has [12] y Sandhya [13] afirmaron que los modelos de aprendizaje profundo para el reconocimiento de objetos son efectivos en la detección de intenciones delictivas y mejoran las operaciones policiales al identificar sospechosos, en tanto que, Jenga et al. [14] y Kumar [15] subrayaron la necesidad de adaptar estas técnicas para entornos educativos. Stalidis et al [16] exploraron cómo las arquitecturas de aprendizaje pueden predecir delitos en contextos educativos, proponiendo un marco para implementar medidas de detección específicas.

Por otro lado, Divya et al [17] desarrollaron un sistema que utiliza VGG-19, conocido por su eficacia en el reconocimiento de imágenes de armas, para ello emplearon redes convolucionales de regiones (RCNN) para delimitar objetos en imágenes relevantes, logrando una precisión de entrenamiento superior al 80 %. Anandhi, et al [18] propusieron un sistema automatizado para detectar actos de violencia en tiempo real usando cámaras de vigilancia mediante algoritmos de aprendizaje profundo.

La implementación de aprendizaje profundo en el contexto mexicano, particularmente en el sector educativo, crea una brecha significativa en las estrategias preventivas que permitan identificar armas portadas por individuos que busquen cometer actos de violencia, como tiroteos en instituciones educativas.

El Estado de México enfrenta desafíos de seguridad en el ámbito educativo, especialmente en áreas de crecimiento como Tianguistenco[4]. El TecNM/TEST se encuentra en el municipio de Tianguistenco (Tilapa), en donde se han reportado fosas clandestinas en un sembradío de maíz en el año 2024[5]. Asimismo, en el 2023 se suscitaron 18 feminicidios en un lapso de dos meses en los municipios de Otzolotepec, Almoloya de Juárez, Toluca y Tianguistenco, dichos delitos fueron cometidos con arma de fuego (2), arma blanca (5) y otros (11). Uno de los casos más mencionados fue el homicidio de Ameyali de 17 años quien

[4] https://www.ieem.org.mx/maxima_publicidad/maxima17_18/InfGral/PLATAFORMAS-2018/docs/Plataformas _Mpales/PRI/102_Tianguistenco.pdf?utm_source=chatgpt.com

[5] https://www.milenio.com/policia/hallan-3-cuerpos-en-fosa-clandestina-de-tianguistenco-edomex?cx_testId=32&cx_testVariant=cx_1&cx_artPos=0&cx_experienceId=EXZC9VQ5Y2SQ&cx_experienceActionId=showRecommendationsI8BLQC9TX2D091#cxrecs_s

https://doi.org/10.36677/rici.v3i1.25167 Enero-Julio 2025 ISSN: 2992-7447 págs. 88-100

vivía en Tianguistenco, Tilapa, a ella se le encontró sin vida en el paraje llamado las ratoneras rumbo al Ajusco[6].

Esta situación de violencia en Santiago Tianguistenco y su comunidad de Tilapa, Estado de México, ha sido motivo de preocupación en los últimos meses. Entre los hechos recientes, destaca el asesinato de tres comerciantes en septiembre de 2024. Estas personas fueron secuestradas y posteriormente encontradas sin vida en una calle de la localidad. Este suceso impactó profundamente a la comunidad, evidenciando los desafíos en materia de seguridad pública que enfrenta la región [7].

Por lo antes mencionado, la falta de tecnologías avanzadas para detectar violencia en el Tecnológico de Estudios Superiores de Tianguistenco (TecNM/TEST) es preocupante. Este estudio propone implementar YOLO v5 para el reconocimiento de armas en el TecNM/TEST, buscando mitigar los riesgos de violencia mediante detección temprana y efectiva, y así hacer eficiente los protocolos de seguridad, pudiendo ser adaptable para otras instituciones, fomentando un entorno académico seguro.

You Only Look Once

YOLO es un modelo de redes neuronales convolucionales para la detección de objetos de una sola pasada (*single-shot detection*), esto lo hace ideal para aplicaciones en problemas que requieren respuesta en tiempo real, como sistemas de vigilancia o vehículos autónomos Funciona transformando el proceso de detección de objetos en una tarea de regresión lineal, en la cual se analizan directamente las imágenes de entrada para determinar las coordenadas de cuadros delimitadores y las probabilidades asociadas a cada clase de objeto detectado [19] .

YOLO presenta varias ventajas con respecto a otros modelos de aprendizaje automático, particularmente en tareas de detección de objetos, debido a que se requiere menos recursos computacionales en comparación con modelos como Faster R-CNN o SSD, lo que permite implementar el modelo de YOLO en dispositivos con hardware limitado.

En su funcionamiento YOLO [19] se puede describir en los siguientes pasos:

1. Segmentar una imagen de entrada en una cuadrícula $S \times S$.

2. Para cada celda de la cuadrícula, el modelo genera predicciones para B cuadros delimitadores y asigna a cada uno un valor de confianza. Este valor de confianza refleja la precisión con la que el cuadrado delimitador encierra un objeto.

3. Cada celda de la cuadrícula estima probabilidades condicionales asociadas a C clases posibles de objetos.

4. Para obtener la predicción final, la red convolucional combina estas probabilidades condicionales con los valores de confianza correspondientes, multiplicándolos para evaluar la clase más probable de objeto presente dentro de cada cuadro delimitador.

YOLOv5 [20] optimiza el modelo YOLO en tres componentes principales:

[6] https://adnoticias.mx/ameyali-alvarado-localizan-sin-vida/

[7] https://paralelo19.tv/nacional/2024/09/11/tragedia-en-santiago-tianguistenco-tres-comerciantes-asesinados y https://www.milenio.com/policia/matan-a-3-personas-en-santiago-tianguistenco-eran-polleros

- *Backbone*: Procesa las imágenes de entrada *reduciendo* sus dimensiones espaciales y aumentando la profundidad de sus canales para extraer características detalladas. Se beneficia de los módulos *Focus,* que optimiza la captura de información espacial en menos operaciones y CSPDarknet, mejora la eficiencia de la extracción de características con un diseño eficiente que reduce el coste computacional y mejora el flujo de información. Estas mejoras hacen que YOLOv5 sea más rápido y preciso en la detección de objetos [20].

- *Neck*: En este componente se utilizan los módulos PAN (Path Aggregation Network) y FPN (Feature Pyramid Network) para integrar características detectadas en múltiples escalas espaciales, mejorando notablemente la precisión en la localización de objetos de diversos tamaños. FPN propaga características ricas en contexto de niveles superiores a niveles inferiores, mientras que PAN mejora la precisión de localización optimizando la fusión de características de bajo nivel con alta resolución [20].

- *Head:* En este componente se optimiza la detección simultánea de objetos grandes y pequeños, aumentando la precisión del modelo.

El modelo utiliza la métrica de intersección sobre la unión (IoU, por sus siglas en inglés) de la Ec. (1) para seleccionar los cuadros con mayor precisión en la detección de objetos [20].

$$IoU = \frac{\text{mín}\,(W_a, W_b) * min(H_a, H_b)}{W_a H_a + W_b H_b - \text{mín}\,(W_a, W_b) * min(H_a, H_b)} \tag{1}$$

Donde de la Ec. (1), (W_a, H_a) y (W_b, H_b) representan el alto y ancho de los cuadros.

YOLOv5s, por su tamaño reducido y baja demanda de recursos, es ideal para dispositivos con hardware limitado, como Raspberry Pi y teléfonos móviles [19].

METODOLOGÍA

En respuesta a los desafíos de seguridad en el TecNM/TEST, se propone una metodología para desarrollar un sistema de detección de armas blancas, armas de fuego y arrebatos con el modelo YOLOv5.

Conjunto de datos

Ante la falta de conjuntos de datos que involucren armas y arrebatos, se creó un conjunto de datos especializado. Se recopilaron 2400 imágenes de la web mediante el motor de búsqueda de Google[8], utilizando palabras clave como pistolas de comando, mausser, luger, revólveres, rifles, escopetas, fusiles mosquetones para categorizar armas de fuego y punzocortantes como cuchillos de cocina, navajas, machetes zapapicos, hachas, para armas blancas y comportamientos de arrebatos. Las imágenes se seleccionaron cuidadosamente de sitios públicos, priorizando el formato JPEG, dimensiones de 640x640 y visibilidad clara de las armas. Todo el conjunto recolectado se destinó al entrenamiento del modelo, respetando los derechos de autor y utilizando material que cumple con las normativas de uso y distribución de imágenes públicas.

[8]https://es.m.wikipedia.org/wiki/Archivo:Colt_Official_Police_38_Revolver-NMAH-AHB2015q021271.jpg

https://doi.org/10.36677/rici.v3i1.25167 Enero-Julio 2025 ISSN: 2992-7447 págs. 88-100

Para el conjunto de prueba, se recopilaron 400 imágenes autorizadas por autoridades del TecNM/TEST. Se usaron réplicas de juguete para las armas de fuego y utensilios de oficina como armas blancas (cutters). Las imágenes se tomaron bajo iluminación controlada, protegiendo la identidad de los participantes con cubrebocas. Además, las imágenes se mantuvieron consistentes en tamaño y formato con respecto al conjunto de entrenamiento. Todos los participantes otorgaron su consentimiento informado previo a la inclusión en el estudio.

Figura 2: *Representación de objetos en un entorno controlado para pruebas de reconocimiento en el TecNM/TEST.*

El etiquetado de las imágenes se llevó a cabo manualmente utilizando la herramienta LabelImg, que está desarrollada en Python. Este proceso implicó la selección cuidadosa de las regiones de interés en las imágenes, donde se identificaban claramente armas de fuego, armas blancas y/o acciones de arrebato. Estas regiones seleccionadas constituyen las clases dentro del conjunto de datos.

Modelo YoloV5s

La red neuronal convolucional *YOLOV5s* cuenta con 117 capas, 5447688 parámetros, 11.4 GFLOPs que realizan 11.4 mil millones de operaciones de punto flotante para procesar una imagen de entrada, un batch_size de 4 imágenes, 100 épocas, una función de activación llamada LeakyReLU(0.1), considera hasta 80 clases de un entrenamiento previo con el conjunto de datos llamado COCO (pesos yolov5s.pt), utiliza una profundidad de modelo de 0.33 (depth_multiple : 0.33), un número de canales en cada capa convolucional del modelo de 0.50 (width_multiple: 0.50).

Métricas de validación

Con el fin de medir la eficiencia del modelo YOLOv5s la métrica del promedio de precisión media (mAP, por sus siglas en inglés) es comúnmente adoptada debido a su capacidad para integrar las medidas de precisión y sensibilidad a lo largo de diferentes umbrales de IoU.

Formalmente la mAP se calcula promediando los valores de precisión en varios niveles de sensibilidad, lo que da como resultado un único valor escalar que indica la precisión general de detección, como se describe en la Ec. (2).

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i \tag{2}$$

Donde C denota el número de clases y AP_i (Ec. (3)) representa el área bajo la curva de precisión-sensibilidad para la clase i [20].

$$AP_i = \sum_i (R_i - R_{i-1})\, P_i \tag{3}$$

Donde R denota la sensibilidad, definida como la proporcion de verdaderos positivos identificados en relación con el total de los verdaros positicos rales (Ec. (4)) y P denota precisión, que se calcula como la proporcion de verdaderos positivos en comparación con el total de predicciones positivas (Ec. (5)).

$$R = \frac{TP}{TP - FN} \tag{4}$$

$$P = \frac{TP}{TP + FP} \tag{5}$$

Donde TP representa el número de verdaderos positivos, FP indica el número de falsos positivos y FN señala el número de falsos negativos.

Escenario Experimental

En la experimentación se usaron 1,920 imágenes para entrenamiento y 480 para validación, configurando 100 épocas con un tamaño de lote de 4 imágenes. El entrenamiento se realizó en Google Colab, con Python 3.10, 12 GB de RAM y una GPU NVIDIA® T4. La fase de prueba con 400 imágenes se realizó usando un umbral de confianza del 60 % para evaluar la métrica de la mAP. Los experimentos se realizaron en una computadora HP EliteBook 820, procesador Intel Core i5 6ta generación, velocidad (3.0 Ghz), cache 3 MB, memoria (8 GB), unidad de estado sólido 256 GB, pantalla LED 12.5", gráficos HD 520 y video integrado. En un entorno de Windows y Anaconda, instalando en el entorno las librerías como Pytorch usando el toolkit de CUDA.

RESULTADOS

En la Figura 3, se muestran los resultados de la función de pérdida en el contexto de entrenamiento (train/cls_loss) y validación (val/cls_loss), de los 10 (a, b,c, d, e, f, g, h, i, j) modelos construidos con validación cruzada. En estos resultados se observa que el error de entrenamiento tiende a disminuir de forma constante a medida que aumentan las épocas, lo cual está ajustándose adecuadamente a los datos de entrenamiento, mientras que el error de validación también tiene un descenso progresivo indicando que el modelo está generalizando bien, sin embargo, las oscilaciones del error son normales en los datos de validación debido a que son un conjunto más reducido o diverso.

En la Figura 4, el mAP de los resultados de entrenamiento varía entre 83.8 % y 90.4 %, lo que indica que los modelos son capaces detectar y clasificar todas las clases adecuadamente. Este rango indica un rendimiento bastante sólido, ya que un valor más alto de mAP refleja una mayor precisión en la detección de objetos, mientras que los valores cercanos al 100 % son ideales.

En la Figura 5, el mAP de los resultado de prueba que varía entre 53.16 % y 60.59 %, significa que el modelo logra detectar y clasificar correctamente los objetos en las imágenes con un rendimiento moderado.

https://doi.org/10.36677/rici.v3i1.25167 Enero-Julio 2025 ISSN: 2992-7447 págs. 88-100

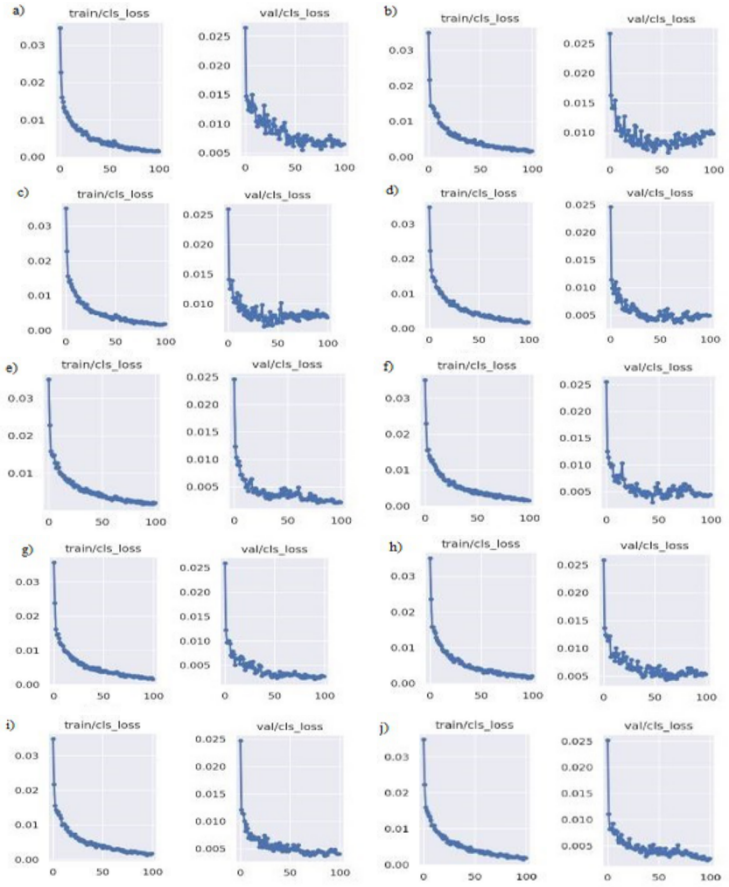


Figura 3: *Resultados de la función de pérdida del entrenamiento y la validación, considerando una validación cruzada de 10.*

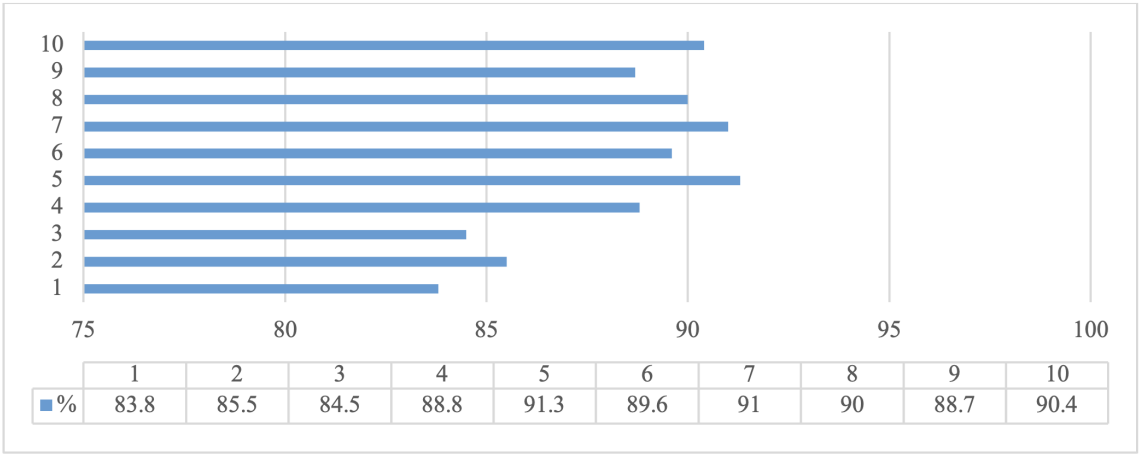


Figura 4: *Resultados del mAP en porcentaje de cada entrenamiento, considerando una validación cruzada de 10.*

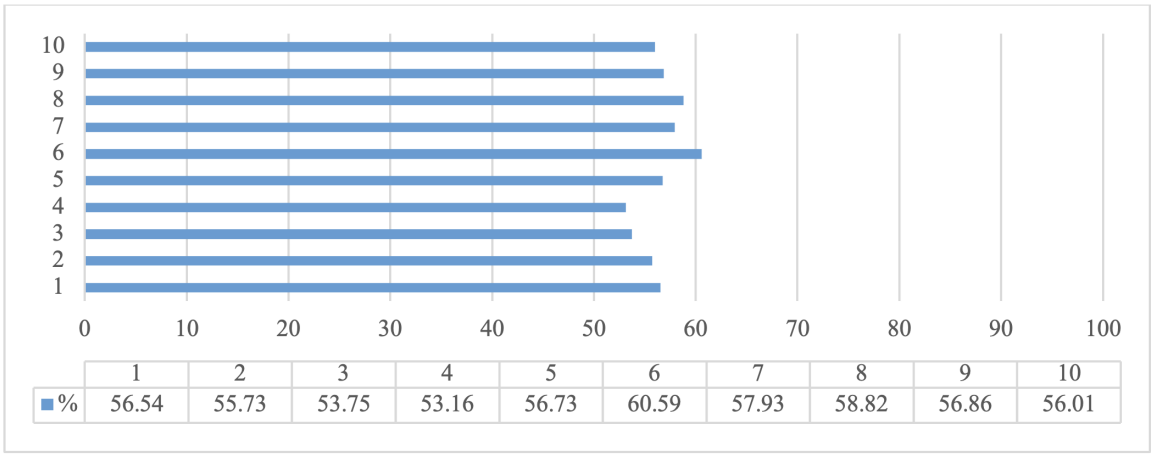


Figura 5: *Resultados del mAP en porcentaje de cada prueba, considerando una validación cruzada de 10.*

https://doi.org/10.36677/rici.v3i1.25167 Enero-Julio 2025 ISSN: 2992-7447 págs. 88-100

Conclusiones

En 2023, el Estado de México experimentó un aumento significativo en delitos como robos y asaltos con armas de fuego y blancas, además de agresiones directas, lo que generó preocupación sobre la posible expansión de estos incidentes a entornos escolares. Ante esta situación, este estudio evaluó la implementación del modelo YOLOv5s en el Tecnológico de Estudios Superiores de Tianguistenco. La investigación se centró en la capacidad del modelo para detectar y clasificar eventos relacionados con armas blancas y de fuego, así como actos de arrebato, buscando mejorar los protocolos de seguridad y ofrecer respuestas más eficientes a los incidentes, contribuyendo a un ambiente académico más seguro y favorable para el aprendizaje.

Los resultados de mAP indica que el modelo YOLOv5s muestra un desempeño notable en la detección de armas y acciones de arrebato durante las fases de entrenamiento, con valores entre 83.8 % y 90.4 %. Estos resultados destacan la capacidad del modelo para identificar correctamente los objetos de interés. Sin embargo, en las pruebas realizadas en escenarios no conocidos por el modelo en el entrenamiento, el mAP disminuye a un rango de 53.16 % a 60.59 %, lo que refleja una reducción significativa en la precisión general y la recuperación del modelo bajo condiciones nuevas y variables.

El contraste en el rendimiento sugiere que, aunque el modelo es eficaz en entornos controlados, necesita ajustes para mejorar su consistencia y precisión en escenarios no previstos durante la fase de entrenamiento, donde las diferencias en el contexto y las condiciones presentan desafíos adicionales para la detección efectiva de violencia.

Como trabajo futuro se encuentra la implementación de técnicas de aprendizaje incremental y transferencia de aprendizaje. Esto permitirá al modelo adaptarse mejor a nuevas condiciones y variabilidades que no fueron presentes en la fase de entrenamiento inicial. Además, sería beneficioso incrementar la diversidad y cantidad de datos en las etapas de entrenamiento para cubrir un espectro más amplio de situaciones potenciales, incluyendo diferentes entornos, tipos de armas y comportamientos de arrebato.

Declaración de contribución de autores y colaboradores

Ballesteros-Reyes, L. F.: Ejecución de experimentos y redacción. Cleofas-Sánchez, L.: Diseño experimental, redacción y supervisión. Guzmán-Ponce, A.: Revisión y redacción. Yañez-Badillo, H.: Revisión. Pineda-Briseño, A.: Revisión

Referencias

[1] A. Lankford. The importance of analyzing public mass shooters separately from other attackers when estimating the prevalence of their behavior worldwide. *Econ Journal*, vol. 17, no. 1, pp. 40–55, 2020. https://journaltalk.net/articles/5999/.

[2] D. Arboleda Andrade. Tiroteo en Stoneman Douglas: discursos políticos, funcionamiento y materialidades discursivas del suceso. *Revista Científica de Comunicación*, vol. 10, no. 1, pp. 49–68, 2019. DOI: 10.31207/rch.v10i1.177.

[3] D. Perez Esparza, S. D. Johnson, y P. Gill. Why did Mexico become a violent country? Assessing the role of firearms trafficked from the US. *Security Journal*, vol. 33, pp. 179–209, 2020. DOI: 10.1057/s41284-019-00178-6.

[4] E. Weigend Vargas, M. Degli Esposti, y S. Hargarten. Examining firearm-related deaths in Mexico. *Econ Journal*, vol. 11, no. 33, pp. 2–7, 2024. DOI: 10.1186/s40621-024-00519-z.

[5] Instituto Nacional de Estadística y Geografía (INEGI). 2023. [Online]. Available: `https://en.www.inegi.org.mx/inegi/acercade.html`. [Consultado el 1 de noviembre de 2024].

[6] A. Pulido Gómez y B. J. Almaraz Calderón. Violencia y comportamiento electoral: el caso del Estado de México. *Apuntes Electorales*, vol. 16, no. 56, pp. 9–38, 2017. Recuperado de `https://aelectorales.ieem.org.mx/index.php/ae/article/view/80`.

[7] J. E. Goldstick, R. M. Cunningham, y P. M. Carter. Current Causes of Death in Children and Adolescents in the United States. *New England Journal of Medicine*, vol. 386, no. 20, pp. 1955–1956, 2022. DOI: 10.1056/nejmc2201761.

[8] F. Dakalbab, M. A. Talib, O. A. Waraga, A. B. Nassif, S. Abbas, y Q. Nasir. Artificial intelligence & crime prediction: A systematic literature review. *Social Sciences & Humanities Open*, vol. 6, no. 1, pp. 1–23, 2022. DOI: 10.1016/j.ssaho.2022.100342.

[9] A. M. Al-Madani, V. Mahale, y A. T. Gaikwad. Real-time detection of crime and violence in video surveillance using deep learning. in *First International Conference on Advances in Computer Vision and Artificial Intelligence Technologies (ACVAIT 2022)*, Atlantis Press, 2023, pp. 431–441. DOI: 10.2991/978-94-6463-196-8_33.

[10] P. Sivakumar, V. Jayabalaguru, R. Ramsugumar, y S. Kalaisriram. Real-time crime detection using deep learning algorithm. in *2021 International Conference on System, Computation, Automation and Networking (ICSCAN)*, IEEE, 2021, pp. 1–5. DOI: 10.1109/ICSCAN53069.2021.9526393.

[11] M. M. Mukto, M. Hasan, M. M. Al Mahmud, I. Haque, M. A. Ahmed, T. Jabid, M. S. Ali, M. R. A. Rashid, M. M. Islam, y M. Islam. Design of a real-time crime monitoring system using deep learning techniques. *Intelligent Systems with Applications*, vol. 21, pp. 1–17, 2024. DOI: 10.1016/j.iswa.2023.200311.

[12] A. O. Hashi, A. A. Abdirahman, M. A. Elmi, y O. E. R. Rodriguez. Deep learning models for crime intention detection using object detection. *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 4, pp. 300–306, 2023. DOI: 10.14569/IJACSA.2023.0140434.

[13] S. Sandhya, A. Balasundaram, y A. Shaik. Deep learning-based face detection and identification of criminal suspects. *Computers, Materials & Continua*, vol. 74, no. 2, pp. 2331–2343, 2023. DOI: 10.32604/cmc.2023.032715.

https://doi.org/10.36677/rici.v3i1.25167 Enero-Julio 2025 ISSN: 2992-7447 págs. 88-100

[14] K. Jenga, C. Catal, y G. Kar. Machine learning in crime prediction. *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 3, pp. 2887–2913, 2023. DOI: 10.1007/s12652-023-04530-y.

[15] J. R. R. Kumar, D. Dhabliya, y S. S. Dari. A comparative study of machine learning algorithms for image recognition in privacy protection and crime detection. *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 9s, pp. 482–490, 2023. ISSN: 2147-679921.

[16] P. Stalidis, T. Semertzidis, y P. Daras. Examining deep learning architectures for crime classification and prediction. *Forecasting*, vol. 3, no. 4, pp. 741–762, 2021. DOI: 10.3390/forecast3040046.

[17] M. Divya, G. S. Priya, R. Abitha, K. Sirisha, A. Manikanta, y K. Jayanth. Automated crime intention detection using deep learning. *International Research Journal of Modernization in Engineering, Technology, and Science*, vol. 4, no. 6, pp. 5670–5676, 2022. e-ISSN: 2582-5208.

[18] R. J. Anandhi. Edge Computing-Based Crime Scene Object Detection from Surveillance Video Using Deep Learning Algorithms. in *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2023, pp. 1159–1163. DOI: 10.1109/ICIRCA57980.2023.10220800.

[19] M. Alruwaili *et al.* Deep Learning-Based YOLO Models for the Detection of People With Disabilities. *IEEE Access*, vol. 12, pp. 2543–2566, 2024. DOI: 10.1109/ACCESS.2023.3347169.

[20] K. Jun-Hwa, K. Namho, P. Yong Woon, y W. Chee Sun. Object Detection and Classification Based on YOLO-V5 with Improved Maritime Dataset. *Journal of Marine Science and Engineering*, vol. 10, no. 3, pp. 2–14, 2022. DOI: 10.3390/jmse10030377.