

Herramienta automatizada de segmentación y clasificación del paisaje urbano para la vigilancia eco-epidemiológica con drones

Sarahi Ventura-Angoa^A, Víctor Muñiz-Sánchez^A,
Francisco Javier Hernández-López^B, Rogelio Danis-Lozano^C,
Kenia Mayela Valdez-Delgado^C

^A Centro de Investigación en Matemáticas, A.C., CIMAT, Unidad Monterrey
Parque de Investigación e Innovación Tecnológica (PIIT)

^B SECIHTI - Centro de Investigación en Matemáticas A.C., CIMAT Unidad Mérida
Parque Científico y Tecnológico de Yucatán (PCTY)

^C Centro Regional de Investigación en Salud Pública (CRISP),
Instituto Nacional de Salud Pública (INSP)

sarahi.ventura@cimat.mx, victor_m@cimat.mx, fcoj23@cimat.mx, rdanis@insp.mx,
Autor de Correspondencia: kenia.valdez@insp.mx

RESUMEN

*Este estudio propone una herramienta automatizada para la segmentación y clasificación del paisaje urbano asociado a criaderos de *Aedes aegypti*, utilizando imágenes aéreas RGB tomadas por dron en Tapachula, Chiapas, México. Bajo un enfoque de Análisis de Imágenes Basado en Objetos (OBIA), las imágenes fueron segmentadas en superpíxeles mediante el algoritmo SLIC, y se extrajeron características estadísticas, de borde y de textura para cada objeto. Se construyó el conjunto de datos SetDroneDataset-TapachulaRGB (109,956 superpíxeles, 10 clases). De los clasificadores evaluados (KNN, SVM y MLP), la red neuronal MLP obtuvo el mejor desempeño global (F1-score ponderado) y el menor tiempo de predicción.*

PALABRAS CLAVE

Paisaje urbano, *Aedes aegypti*, superpíxeles, aprendizaje automático, drones.

ABSTRACT

*This study proposes an automated tool for the segmentation and classification of urban land associated with *Aedes aegypti* breeding sites using RGB aerial imagery captured by drone in Tapachula, Chiapas, México. Under an Object-Based Image Analysis (OBIA) framework, images were segmented into superpixels using the SLIC algorithm, and statistical, edge, and texture features were extracted for each object. The dataset SetDroneDataset-TapachulaRGB was constructed (109,956 labeled superpixels, 10 classes). Among the classifiers evaluated (KNN, SVM, and MLP), the MLP neural network achieved the best overall performance (weighted F1-score) and the shortest prediction time.*

KEYWORDS

Urban landscape, *Aedes aegypti*, superpixels, machine learning, drones.

INTRODUCCIÓN

El dengue es una enfermedad viral de importancia mundial, transmitida a los humanos por la picadura de hembras infectadas del mosquito *Aedes aegypti*.¹ El mosquito *Ae. aegypti* es el vector principal de dengue en México.² En México, entre 2010 y 2020, el Sistema Nacional de Vigilancia Epidemiológica registró 1,620,872 casos probables y 336,991 casos confirmados, con una tasa de

mortalidad hospitalaria que aumentó de 0.7% a 2.6% en ese período.³ Esta enfermedad representa no solo una amenaza para la salud pública, sino también una carga económica significativa, estimada en 170 millones de dólares anuales para México; de los cuales, 49.5 millones corresponden a la atención médica y 82 millones al programa de vigilancia y control de vectores.⁴

Frente a esta situación, la identificación y el monitoreo de los sitios de reproducción de este mosquito son fundamentales para priorizar las intervenciones en áreas de alto riesgo. Por ello, el desarrollo de herramientas que permitan su identificación y seguimiento es clave para generar datos consistentes que coadyuven en los programas de control. Ante esta necesidad, el uso de drones o vehículos aéreos no tripulados (VANT) ha cobrado relevancia en salud pública, debido a su eficiencia operativa y bajo costo. Experiencias en países como Brasil, Perú y Tanzania han demostrado su potencial para mapear criaderos de mosquitos,⁵ identificar hábitats larvarios⁶ e incluso aplicar larvicidas.⁷ En Tapachula, Chiapas, se implementó por primera vez una metodología basada en drones para el monitoreo de criaderos de *Ae. aegypti*⁸ validándolos como una herramienta complementaria en programas de control vectorial. Además, se demostró que los drones pueden utilizarse en zonas urbanas para generar mapas especializados⁹ y obtener datos para la generación de indicadores de riesgo.¹⁰

Desde una perspectiva integradora, la información generada por drones puede incorporarse al análisis del paisaje urbano para una mejor comprensión de los determinantes ambientales que favorecen el desarrollo del ciclo de vida del mosquito *Ae. aegypti*.⁹ Esta integración es relevante ya que la estructura y la composición del paisaje influyen en la disponibilidad de hábitats de reproducción, de refugio y de alimentación de los mosquitos.¹¹ El mosquito *Ae. aegypti* requiere agua estancada para reproducirse (en recipientes diversos de uso doméstico),^{2,12} mientras que la vegetación densa proporciona hábitats para su alimentación y reposo.¹³ Asimismo, las superficies pavimentadas y los techos de lámina modifican las condiciones microclimáticas, influyendo indirectamente en su abundancia y distribución.¹⁴

No obstante, las herramientas actuales para identificar y clasificar características del paisaje asociadas a la abundancia de mosquitos carecen de la eficiencia y la escalabilidad necesarias para manejar grandes volúmenes de datos de imágenes aéreas. Escasos estudios proponen metodologías efectivas para estos fines.¹⁵ En este contexto, el objetivo de este estudio fue desarrollar una herramienta automatizada para la detección y clasificación de características del paisaje relevantes para la distribución y abundancia de mosquitos *Ae. aegypti* vectores de dengue, utilizando imágenes aéreas de alta resolución captadas por drones en Tapachula, Chiapas. El sistema integró técnicas de análisis de imágenes y de aprendizaje automático para segmentar y clasificar regiones de interés, lo que permitió generar un conjunto de datos robustos para el entrenamiento de modelos predictivos.

Para alcanzar el objetivo de este trabajo, se desarrolló una herramienta de segmentación y etiquetado de objetos de interés en imágenes, basada en el enfoque de Análisis de Imágenes Basado en Objetos (OBIA, *Object-Based Image Analysis*).¹⁶ Esta técnica consiste en fraccionar una imagen en objetos o regiones de imagen formados por grupos de píxeles contiguos, en lugar de analizar píxeles individuales.^{17,18} La implementación de OBIA se centró en dos piezas fundamentales: (i) la segmentación de la imagen en objetos de imagen (superpíxeles) mediante el método SLIC¹⁹; y (ii) la extracción de características a nivel de objeto/superpíxel, con el propósito de construir un conjunto de datos de entrada para los algoritmos de clasificación automática.

DATOS

Se emplearon 140 imágenes RGB de alta resolución (1536×1152 píxeles) correspondientes a las colonias El Vergel y las Delicias, Municipio de Tapachula, Chiapas, México; de las cuales se obtuvieron un total de 109,956 superpíxeles. Las imágenes se adquirieron con un dron DJI Matrice 600 (DJI, Shenzhen, China). Un vehículo aéreo no tripulado tipo hexacóptero, equipado con un sistema de estabilización (gimbal) y una cámara DJI Zenmuse X5. Los vuelos se realizaron entre las 12:00 y las 14:00 h durante el mes de noviembre de 2019, a una altitud aproximada de 100 m sobre el nivel del terreno. La planificación de los vuelos se basó en información pública de elevación

obtenida del modelo ASTER GDEM (*Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model*).

SEGMENTACIÓN BASADA EN SUPERPIXELES MEDIANTE EL ALGORITMO SLIC

La segmentación es la división de una imagen en regiones que representan objetos o partes de objetos; el grado de partición depende de la aplicación.²⁰ En imágenes RGB, los atributos básicos para segmentar incluyen componentes de color, textura, forma y contexto.

En OBIA, la segmentación se utiliza para derivar las unidades iniciales: una partición de la imagen en regiones mutuamente excluyentes y conjuntamente exhaustivas, relativamente homogéneas en su interior y diferenciables respecto de las regiones adyacentes.¹⁸ Estos segmentos pueden considerarse como objetos-imagen candidatos: son discretos, internamente coherentes (dado que han superado el criterio de homogeneidad del algoritmo) y diferentes en cierta medida de su entorno; de lo contrario, se habrían fusionado con algún vecino, sobre los cuales se extraen características y se realiza la clasificación.

La segmentación se realizó mediante superpíxeles que se puede definir como un conjunto de píxeles contiguos que comparten características similares, agrupados en regiones homogéneas respetando en lo posible los límites naturales de los objetos presentes en la imagen,^{21,22} que fueron generados con el algoritmo SLIC (*Simple Linear Iterative Clustering*), -cuya técnica agrupa píxeles considerando su similitud de color y su proximidad espacial, y opera en un espacio de cinco dimensiones (5-D) $[l, a, b, x, y]$, donde $[l, a, b]$ es el vector de color en el espacio *CIELAB* y $[x, y]$ es la posición del píxel en la imagen.⁹ En *CIELAB*, los canales de color se interpretan como:

- l^* : luminosidad (0 = negro, ≈ 100 = blanco)
- a^* : eje verde (−) $\leftrightarrow$ rojo (+)
- b^* : eje azul (−) $\leftrightarrow$ amarillo (−)

Dado que la distancia espacial en el plano (x, y) depende del tamaño de la imagen, no es apropiado usar directamente la distancia euclidiana de cinco dimensiones (5-D) sin antes normalizar la componente espacial.¹⁹ Entonces, sea i un píxel y k el centro de un superpíxel. Se definen la distancia de color y la distancia espacial como:

$$d_{lab} = \sqrt{(l_i - l_k)^2 + (a_i - a_k)^2 + (b_i - b_k)^2} \quad (1)$$

$$d_{xy} = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2} \quad (2)$$

$$d_s = d_{lab} + \frac{m}{S} d_{xy} \quad (3)$$

Donde d_s es la suma de la distancia lab y la distancia espacial normalizada xy en el intervalo de la cuadrícula S , definido como $S \approx \sqrt{N/K}$ es el paso entre centros iniciales para una imagen con N píxeles y K superpíxeles, y m es el parámetro de compacidad que controla el equilibrio entre color y espacio; a mayor m , más se enfatiza la proximidad espacial y más compactos y uniformes son los superpíxeles; a menor m , se sigue más fielmente el contorno radiométrico.²¹

En la práctica, SLIC restringe la búsqueda al vecindario $2S \times 2S$ alrededor de cada centro, asigna píxeles al centro más cercano según d_s , actualiza los centros como promedios de sus asignaciones y repite iterativamente hasta converger, seguido de una corrección de conectividad. En este trabajo se utilizó la librería *OpenCV* mediante el módulo *ximgproc* (*createSuperpixelSLIC*), cuya implementación del método SLIC recibe como entrada cuatro parámetros: (i) tamaño promedio, que corresponde al tamaño aproximado de cada superpíxel; valores mayores generan menos regiones, ya que cada superpíxel cubre un área más grande; (ii) *ruler*, que controla la compacidad de los superpíxeles; valores bajos permiten que los bordes sigan mejor los contornos naturales de los objetos; (iii) iteraciones, que define el número de iteraciones del algoritmo; un mayor número puede

mejorar la segmentación, aunque incrementa el tiempo de procesamiento; y (iv) tamaño mínimo, que establece el tamaño mínimo permitido para un superpixel, evitando la generación de segmentos demasiado pequeños o aislados que puedan considerarse ruido.²³

EXTRACCIÓN DE CARACTERÍSTICAS

La extracción de características es el proceso de transformar los datos brutos de una imagen en atributos numéricos que conserven la información esencial para las tareas posteriores. En este trabajo, la unidad de análisis son los superpíxeles, lo que permite reducir la complejidad (al procesar menos unidades que a nivel de píxel), disminuir la redundancia, preservar bordes y facilitar tanto la extracción de características como la clasificación de coberturas.

Para cada superpíxel se extrajeron diferentes características en el espacio de color RGB, que se refieren a las propiedades relacionadas con la medición de la intensidad de luz registrada por cada uno de los canales Rojo (R), Verde (G) y Azul (B) en una imagen digital.²⁰ Estas características describen la distribución de color e intensidad en la imagen.

Características estadísticas

En este trabajo se calculó la media y la desviación estándar por canal y por superpíxel. Sea $|S|$ el conjunto de píxeles que componen un superpíxel $|S|$ su cardinalidad y $C \in \{R, G, B\}$ un canal; si $C(x, y)$ denota la intensidad del canal C en el píxel (x, y) , entonces:

$$\bar{C}_s = \frac{1}{|S|} \sum_{(x,y) \in S} C(x, y), \quad \sigma_{\{C,S\}} = \sqrt{\frac{1}{|S|} \sum_{(x,y) \in S} (C(x, y) - \bar{C}_s)^2} \quad (4)$$

Estas métricas resumen el brillo (intensidad promedio por canal) y su variabilidad dentro de cada superpíxel. En resumen, se extraen las siguientes características por superpíxel:

- $\bar{R}_S, \bar{G}_S, \bar{B}_S$: media por canal
- $\sigma_{R,S}, \sigma_{G,S}, \sigma_{B,S}$: desviación estándar por canal.

Características de bordes

Además de las características de color, se extrajeron características de bordes a nivel de superpíxel. Se calculó la magnitud de gradiente mediante el operador *Sobel* y la densidad de bordes detectados con *Canny*, ambas medidas por superpíxel.^{20,24}

Gradiente Sobel: Sea $I(x, y)$ la imagen en escala de grises. El gradiente se aproxima convolucionando con máscaras discretas de Sobel:

$$G_x = I * K_x, G_y = I * K_y \quad (5)$$

donde K_x, K_y son los kernels de Sobel horizontal y vertical, respectivamente, y $*$ denota la convolución. La magnitud y la dirección del gradiente se definen como:

$$M(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}, \quad \theta(x, y) = \text{atan2}(G_y(x, y), G_x(x, y)) \quad (6)$$

En la implementación se usa una combinación lineal equivalente y eficiente computacionalmente:

$$\tilde{M}(x, y) = \alpha |G_x(x, y)| + (1 - \alpha) |G_y(x, y)|, \quad \alpha = 0.5 \quad (7)$$

que correlaciona estrechamente con M . Dado un superpíxel S y su cardinalidad $|S_\ell|$, se resumen las intensidades de borde mediante:

$$\bar{M}_S = \frac{1}{|S|} \sum_{(x,y) \in S} \tilde{M}(x,y), \quad \sigma_{M,S} = \sqrt{\sum_{(x,y) \in S} (\tilde{M}(x,y) - \bar{M}_S)^2} \quad (8)$$

Bordes Canny: El detector Canny se aplica sobre la imagen en escala de grises para obtener un mapa binario de bordes $E(x,y) \in \{0,1\}$. A partir de este mapa, la densidad de bordes por superpixel se define como:

$$\rho^{edges} = \frac{\sum_{(x,y) \in S} E(x,y)}{|S|} \quad (9)$$

En resumen, para cada superpixel S se extraen:

- $\bar{M}_S$ media de magnitud de gradiente
- $\sigma_{M,S}$ desviación estándar de la magnitud de gradiente
- ρ^{edges} densidad de bordes

Características de textura

Para extraer características de textura existen diversos métodos; en este trabajo se empleó la matriz de co-ocurrencia de niveles de grises (*GLCM*, por sus siglas en inglés *Gray-Level Co-occurrence Matrix*), una herramienta estadística de segundo orden, que captura relaciones espaciales entre pares de píxeles.²⁵

Aunque no existe una definición formal para la textura, en este contexto se entiende como un patrón formado por elementos mutuamente relacionados. Es decir, consideramos grupos de píxeles cuya apariencia conjunta describe la textura observada; dicha descripción depende del número de píxeles considerados y de sus relaciones.²⁵

Se considera una ventana rectangular de tamaño $M \times N$ de una imagen con textura. Una matriz de frecuencias relativas $P_{\phi,d}(a,b)$ que corresponde a la frecuencia con la cual dos píxeles con escala de grises a y b aparece en la ventana con una distancia d en dirección ϕ , describe una ocurrencia de una ventana de una imagen en escala de grises.

Formalmente, las frecuencias de co-ocurrencia no normalizadas en función del ángulo y la distancia se definen como¹⁴:

$$P_{0^\circ,d}(a,b) = |\{(k,l), (m,n) \in D: k-m=0, l-n=d, f(k,l)=a, f(m,n)=b\}|, \quad (10)$$

$$P_{45^\circ,d}(a,b) = |\{(k,l), (m,n) \in D: (k-m=d, l-n=-d), (k-m=-d, l-n=d), f(k,l)=a, f(m,n)=b\}|, \quad (11)$$

$$P_{90^\circ,d}(a,b) = |\{(k,l), (m,n) \in D: k-m=d, l-n=0, f(k,l)=a, f(m,n)=b\}|, \quad (12)$$

$$P_{135^\circ,d}(a,b) = |\{(k,l), (m,n) \in D: (k-m=d, l-n=d), (k-m=-d, l-n=-d), f(k,l)=a, f(m,n)=b\}|, \quad (13)$$

donde f es la imagen original y $D = (M \times N) \times (M \times N)$. Con las definiciones anteriores se obtiene una matriz simétrica. A partir de estas matrices $P_{\phi,d}(a,b)$ se calcularon 6 descriptores:

- Disimilitud: mide la diferencia media entre niveles de gris en vecindad.

$$\sum_{a,b} |a-b| P_{\phi,d}(a,b) \quad (14)$$

- Energía: también denominada segundo momento angular, (*ASM*, por sus siglas en inglés *Angular Second Moment*) la cual es una medida de la homogeneidad de la imagen. Se define como sigue:

$$\sum_{a,b} P_{\phi,d}^2(a,b) \quad (15)$$

- Entropía:

$$-\sum_{a,b} P_{\phi,d}(a,b) \log_2 P_{\phi,d}(a,b) \quad (16)$$

- Contraste: una medida de la variación local de la imagen (Típicamente $\kappa = 2, \lambda = 1$).

$$\sum_{a,b} |a-b|^\kappa P_{\phi,d}^\lambda(a,b), \quad (17)$$

- Correlación: una medida de linealidad de la imagen.

$$\frac{\sum_{a,b} [(ab)P_{\phi,d}(a,b)] - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (18)$$

donde $\mu_x = \sum_a a \sum_b P_{\phi,d}(a,b)$, $\mu_y = \sum_b b \sum_a P_{\phi,d}(a,b)$, $\sigma_x^2 = \sum_a (a - \mu_x)^2 \sum_b P_{\phi,d}(a,b)$, $\sigma_y^2 = \sum_b (b - \mu_y)^2 \sum_a P_{\phi,d}(a,b)$

- Homogeneidad (Momento de diferencia inversa):

$$\sum_{a,b; a \neq b} \frac{P_{\phi,d}(a,b)}{|a-b|} \quad (19)$$

Para la extracción de estas características se consideraron las orientaciones $\phi \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ y las distancias $d \in \{1, 2, 3, 4, 5\}$. En cada combinación (4 orientaciones $\times$ 5 distancias = 20) se evaluaron seis descriptores GLMC (disimilitud, energía, entropía, contraste, correlación y homogeneidad), de modo que en total se obtienen 120 características de textura por superpíxeles. Mediante la herramienta de etiquetado, se extrajeron en total 129 características por superpíxel. Con estas características se construyó un conjunto de datos compuesto por $n = 109,956$ superpíxeles etiquetados denominado SetDroneDataset-TapachulaRGB, el cual se encuentra disponible en el repositorio de GitHub.²⁶ En la figura 1 se muestra la imagen segmentada (izquierda) y la imagen etiquetada (derecha) con la interfaz diseñada para ello, considerando las categorías: árbol, pavimento, techo de lámina, arbusto, hierba y pasto verde, suelo desnudo, cuerpo de agua, techo de losa, hierba y pasto seco y sombra. En la figura 2 se presenta la distribución de clases utilizada para el entrenamiento y la evaluación de los modelos.

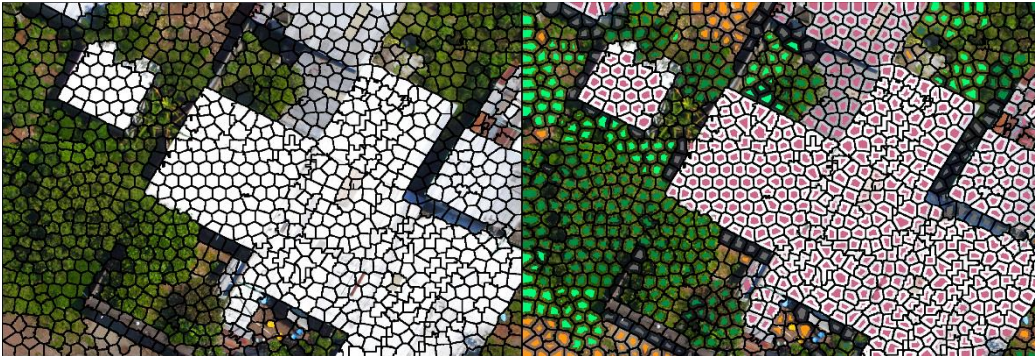


Fig. 1. Segmentación y etiquetado de superpíxeles. Izquierda: imagen segmentada mediante SLIC. Derecha: imagen etiqueta mediante la interfaz de la herramienta de etiquetado con superpíxeles clasificados por cobertura, disponible en el repositorio de Github.²⁶

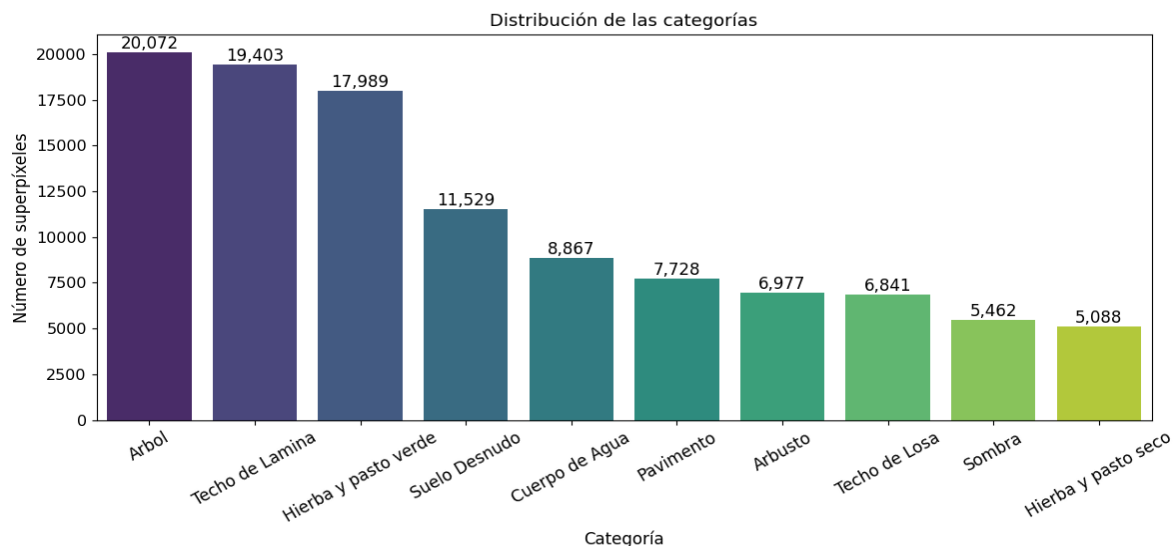


Fig. 2. Distribución de superpíxeles por categoría.

Antes del entrenamiento se desarrolló un módulo de limpieza y visualización para inspeccionar la calidad del etiquetado y detectar confusiones entre coberturas. Para cada imagen, el módulo genera una visualización por categoría, resaltando únicamente los superpíxeles de esa clase y dejando el resto en negro, y guarda los resultados en carpetas por categoría (por ejemplo, `image_x5_1_arbol.jpg` y `image_x5_1_techo_de_lamina.jpg`). Esto permitió revisar rápidamente el etiquetado y corregir manualmente los casos con mayor ruido o con mezcla de clases. En la figura 3 se muestran ejemplos de esta visualización para dos categorías.

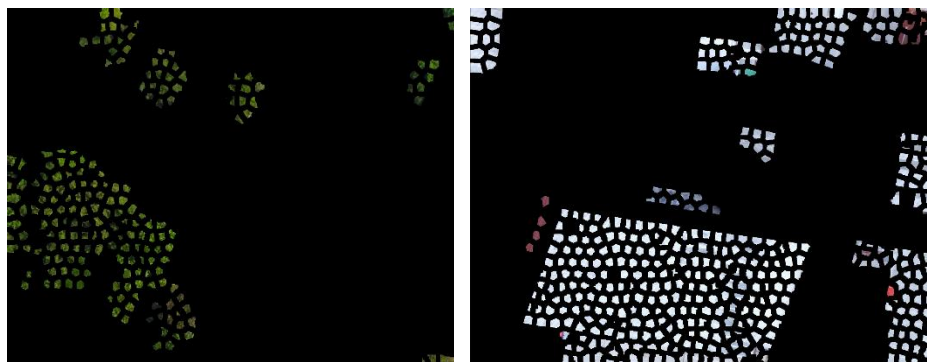


Fig. 3. Ejemplos generados por el módulo de visualización: a la izquierda, se resalta la categoría de árbol; a la derecha, la de techo de lámina.

APRENDIZAJE AUTOMÁTICO PARA LA CLASIFICACIÓN DE SUPERPIXELES

El aprendizaje automático (*machine learning*) es un subcampo de la inteligencia artificial que busca aprender patrones a partir de datos para realizar predicciones en ejemplos no observados.²⁷ En este trabajo se emplea aprendizaje supervisado: cada superpíxel cuenta con un vector de características $x \in \mathbb{R}^p$ (medidas estadísticas, borde y de textura), denotando a $\mathcal{X} \subseteq \mathbb{R}^p$ como el espacio de instancias. A cada instancia se le asignó una etiqueta manual correspondiente a una cobertura del paisaje urbano del sur de México (árbol, suelo desnudo, pavimento, cuerpo de agua,

techo de lámina, techo de losa, arbusto, hierba y pasto seco, hierba y pasto verde y sombra). El conjunto de clases se denota por $\mathcal{Y}$ y se codificó como $\mathcal{Y} \in \{0, \dots, 9\}$, donde p.ej., 0: árbol, 1: suelo desnudo y así sucesivamente.

Se evaluaron tres modelos de clasificación: k vecinos más cercanos (*KNN: k-nearest neighbors*) (nuestro modelo de referencia o baseline) máquinas de soporte vectorial (*SVM: support vector machines*) y una red neuronal tipo perceptrón multicapa (*MLP: multilayer perceptron*).²⁷⁻²⁹ Para el entrenamiento se utilizó el conjunto SetDroneDataset-TapachulaRGB, dividido en 70% para entrenamiento (76,969 superpíxeles) y 30% para validación (32,987 superpíxeles). Debido al desbalance entre clases (ver figura 2), la partición se realizó de forma estratificada para preservar las proporciones de cada categoría en ambos subconjuntos.

El flujo de entrenamiento se implementó mediante *pipelines* para integrar el preprocesamiento y el modelado en una sola ejecución. En primer lugar, las variables predictoras se normalizaron con *StandardScaler* para asegurar la misma escala en todas las características. Posteriormente, se aplicó una búsqueda exhaustiva de hiperparámetros mediante *GridSearch* con el fin de identificar la configuración que maximizara el desempeño de cada modelo; el detalle de los hiperparámetros explorados se encuentra en los scripts disponibles en el repositorio.²⁶ El desempeño se evaluó con métricas multiclase a nivel global y por clase (*precision*, *recall* y *F1-score*), además de matrices de confusión para analizar las confusiones entre coberturas, así como los tiempos de entrenamiento y predicción.

RESULTADOS Y DISCUSIÓN

En esta sección se reporta el desempeño de los modelos *KNN*, *SVM* y *MLP* en el conjunto de validación. Para cada modelo se presentan: i) matrices de confusión para identificar patrones de error entre coberturas, así como las métricas por clase por modelo; ii) comparación de las métricas por clase y por modelo; y iii) métricas globales de desempeño, así como los tiempos de entrenamiento y predicción.

En conjunto, los modelos *KNN*, *SVM* y *MLP* muestran mejor desempeño en las clases de cuerpo de agua y techo de lámina, lo cual se refleja en las matrices de confusión, con una alta concentración de aciertos en la diagonal y pocas confusiones. Las principales confusiones se concentran en la vegetación viva (árbol, arbusto y hierba y pasto verde), donde la mezcla de coberturas dentro de los superpíxeles y la similitud visual provocan confusiones entre estas clases. Por otra parte, las coberturas pavimento, suelo desnudo y techo de losa también muestran confusiones notorias, ya que los tonos grises similares y los cambios de iluminación incrementan la ambigüedad.

De acuerdo con la tabla I, se observa que *KNN* presenta un comportamiento razonable *baseline*: en cuerpo de agua y en techo de lámina, donde la separabilidad en el espacio de características es alta. En su matriz de confusión (ver figura 4), cuerpo de agua presenta errores mínimos, mientras que el techo de lámina se confunde ocasionalmente con pavimento o suelo desnudo, lo cual es coherente con superficies claras o grises bajo variaciones de iluminación.

Tabla I. Métricas por clase para KNN en el conjunto de validación.

Clase	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Árbol	0.6820	0.8219	0.7454
Suelo desnudo	0.6281	0.7136	0.6681
Pavimento	0.6515	0.6559	0.6537
Cuerpo de agua	0.9688	0.9568	0.9627
Techo de lámina	0.9373	0.8820	0.9088
Techo de losa	0.6680	0.5052	0.5753
Arbusto	0.6937	0.5156	0.5915
Hierba y pasto seco	0.7309	0.6168	0.6690
Hierba y pasto verde	0.7013	0.7454	0.7227
Sombra	0.7960	0.5854	0.6747
Promedio ponderado	0.7535	0.7481	0.7466

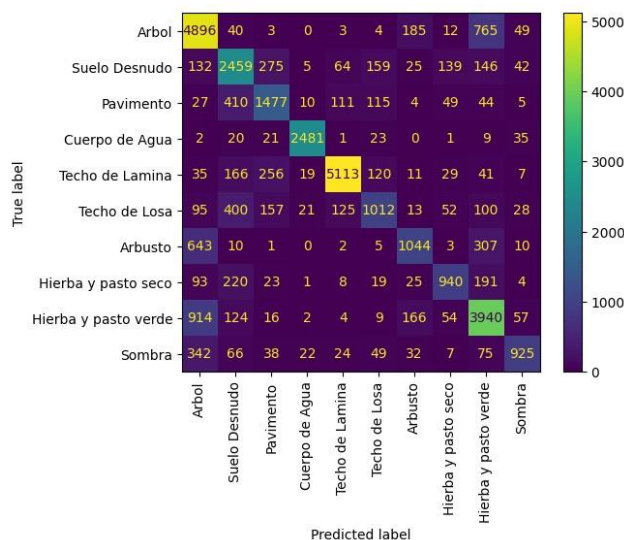


Fig. 4. Matriz de confusión del modelo KNN sobre el conjunto de validación.

Comparativamente, la tabla II muestra que SVM tiene mejor desempeño que el KNN en la gran mayoría de las clases. En la matriz de confusión (ver figura 5) se aprecia una mayor concentración en la diagonal para suelo desnudo, pavimento y techo de losa respecto a KNN, lo que sugiere fronteras de decisión más efectivas en un espacio de alta dimensión tras el escalamiento.

Tabla II. Métricas por clase para SVM en el conjunto de validación.

Clase	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Arbol	0.7668	0.7489	0.7577
Suelo desnudo	0.7813	0.8323	0.8060
Pavimento	0.7378	0.7598	0.7486
Cuerpo de agua	0.9583	0.9657	0.9620
Techo de lámina	0.9546	0.9250	0.9395
Techo de losa	0.7292	0.7369	0.7331
Arbusto	0.4845	0.5862	0.5305
Hierba y pasto seco	0.8379	0.8379	0.8379
Hierba y pasto verde	0.7978	0.7425	0.7692
Sombra	0.7551	0.7203	0.7373
Promedio ponderado	0.8030	0.7981	0.7999

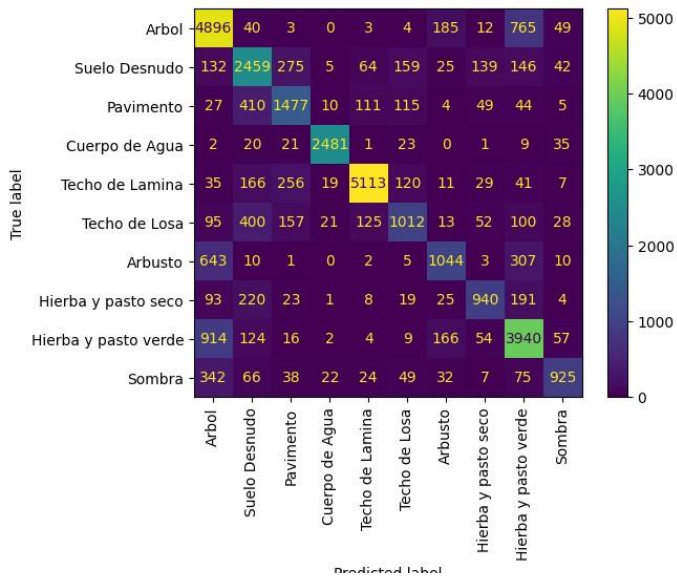


Fig. 5. Matriz de confusión del modelo SVM sobre el conjunto de validación.

Por su parte, MLP supera el desempeño global de KNN y SVM; sin embargo, mantiene dificultades para identificar correctamente la clase arbusto. En particular, presenta un *recall* bajo (ver tabla III), lo que indica que rara vez recupera esta clase y, con frecuencia, la confunde con árbol y hierba y pasto verde como se muestra en la figura 6.

Tabla III. Métricas por clase para MLP en el conjunto de validación.

Clase	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
Árbol	0.7837	0.8848	0.8312
Suelo desnudo	0.8400	0.8622	0.8509
Pavimento	0.8115	0.7993	0.8054
Cuerpo de agua	0.9895	0.9780	0.9837
Techo de lámina	0.9596	0.9589	0.9593
Techo de losa	0.7630	0.7938	0.7781
Arbusto	0.8972	0.5472	0.6798
Hierba y pasto seco	0.8869	0.8898	0.8883
Hierba y pasto verde	0.8249	0.8309	0.8279
Sombra	0.8377	0.7709	0.8029
Promedio ponderado	0.8594	0.8564	0.8546

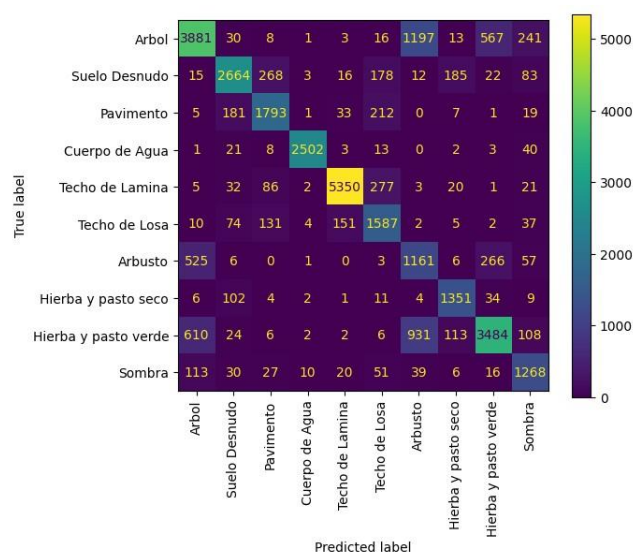


Fig. 6. Matriz de confusión del modelo MLP sobre el conjunto de validación.

Para comparar las métricas por clase y por modelo, el análisis se centra en la gráfica del *F1-score* como se muestra en la figura 7, ya que esta medida resume en un solo valor el equilibrio entre *precision* y *recall*; se observa que MLP alcanza los valores más altos de *F1-score* en la mayoría de las categorías, con excepción de la clase arbusto. En general, este desempeño indica que MLP es el modelo con mejor equilibrio entre *precision* y *recall*. SVM se sitúa en segundo lugar, con valores cercanos a los de MLP; sin embargo, presenta un mejor rendimiento en la clase arbusto en comparación con MLP, lo que sugiere una mayor capacidad para separar esta cobertura frente a vegetación similar.

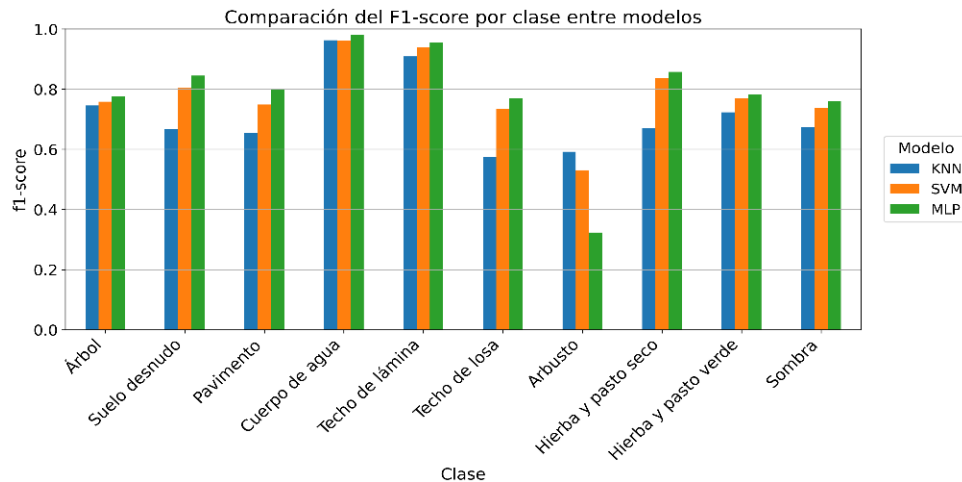


Fig. 7. *F1-score* por clase y por modelo. Valores altos indican un buen equilibrio entre *precision* y *recall* en la categoría, mientras que valores bajos señalan dificultad del modelo para esa clase.

Por su parte, KNN es el modelo con el menor desempeño global frente a SVM y MLP. Aun así, muestra un comportamiento notable en la clase arbusto, donde supera a ambos modelos, lo cual es consistente con su naturaleza local y su sensibilidad a vecindarios representativos en el espacio de características.

Además de las métricas por clase y por modelo, se calcularon métricas globales ponderadas en el conjunto de validación para comparar el desempeño general de cada modelo, como se muestra en la tabla IV. De acuerdo con estos resultados, MLP presenta el mejor desempeño global en las tres métricas, seguido de SVM. Por su parte, KNN obtiene los valores más bajos en todas las métricas. No obstante, a nivel global los resultados de MLP y SVM son cercanos, lo que sugiere que ambos modelos logran un desempeño competitivo en este conjunto de validación.

Tabla IV. Comparación de las métricas globales de modelos en el conjunto de validación

Modelo	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
KNN	0.7535	0.7481	0.7466
SVM	0.8030	0.7981	0.7999
MLP	0.8199	0.8193	0.8086

En la figura 8 se compara el tiempo de entrenamiento y de predicción (en segundos) utilizando la mejor configuración de hiperparámetros encontrada para cada modelo; se observa que el mayor costo computacional en entrenamiento corresponde a SVM (10922.875 s), seguido de MLP (254.770 s), mientras que KNN presenta el menor tiempo (0.119 s). En la fase de predicción, MLP es el más eficiente (0.0545 s), seguido de KNN (4.2652 s), y SVM es el más costoso (86.7473 s). En conjunto, estos resultados indican que, aunque SVM puede alcanzar un desempeño competitivo, su costo computacional es considerable, especialmente en entrenamiento, mientras que MLP ofrece un mejor balance entre desempeño y tiempo de inferencia.

Modelo recomendado. Considerando el desempeño global (*F1-score ponderado*), la estabilidad por clase y el costo computacional en inferencia, MLP presenta el mejor rendimiento general y el comportamiento más consistente entre las categorías evaluadas. Por ello, se considera la opción más adecuada para la clasificación automática de superpíxeles, manteniendo además tiempos de predicción razonables frente a alternativas más costosas como SVM.

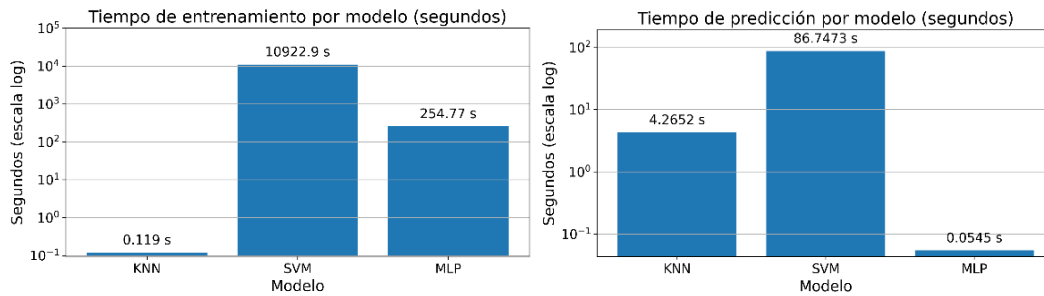


Fig. 8. Comparación de tiempos en segundos para entrenamiento (izquierda) y predicción (derecha).

Finalmente, en la figura 9 a la izquierda se muestra la imagen original y a la derecha la imagen segmentada en superpíxeles del lado y en la figura 10 se muestra la imagen clasificada por el modelo MLP previamente entrenado. De acuerdo con las figuras 9 y 10, en general, la predicción es correcta para la mayoría de las categorías. No obstante, se observan dificultades para distinguir el techo de losa del suelo desnudo: en algunos casos, el techo de losa se confunde con el suelo desnudo y la sombra, mientras que el suelo desnudo se confunde con el pavimento. Este patrón era esperable debido a la similitud entre estas coberturas en términos de color y textura bajo variaciones de iluminación.



Fig. 9. Imagen original (izquierda) y la misma imagen segmentada en *superpíxeles* mediante SLIC (derecha).

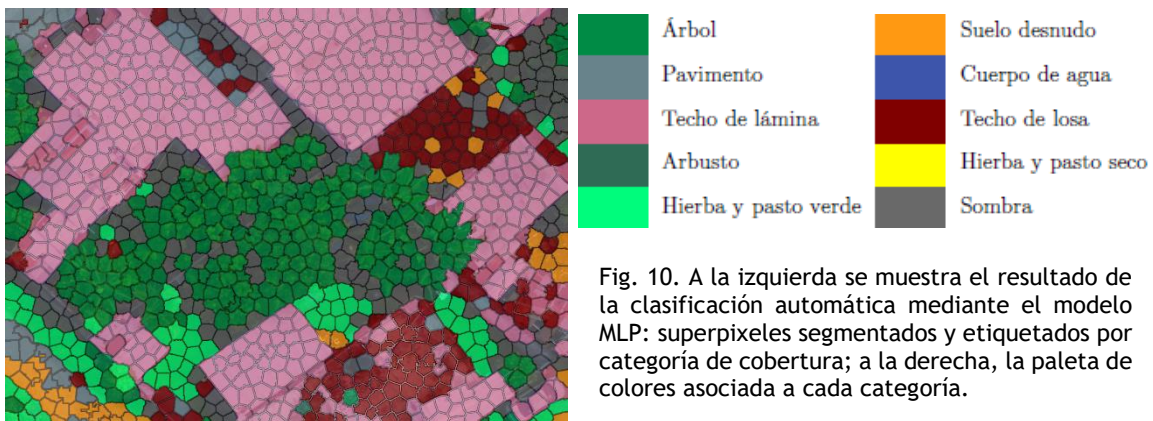


Fig. 10. A la izquierda se muestra el resultado de la clasificación automática mediante el modelo MLP: superpíxeles segmentados y etiquetados por categoría de cobertura; a la derecha, la paleta de colores asociada a cada categoría.

CONCLUSIONES

En este estudio se demostró la viabilidad de desarrollar una herramienta automatizada para la detección y clasificación de características del paisaje urbano relevantes para la distribución y abundancia de mosquitos *Ae. aegypti* vectores de dengue, a partir de imágenes aéreas de alta resolución captadas por drones. La implementación de un flujo de trabajo basado en segmentación por superpíxeles y aprendizaje automático permitió generar información detallada sobre diferentes elementos del paisaje urbano.

De los algoritmos evaluados (KNN, SVM y MLP), el clasificador MLP obtuvo el mejor desempeño global (*precision* = 0.8199, *recall* = 0.8193 y *F1-score* = 0.8086), además de presentar el menor costo computacional en la etapa de predicción (aprox. 0.0545 s). No obstante, se identificaron limitaciones para discriminar ciertas coberturas del paisaje con características espectrales y texturas similares. Las principales confusiones ocurrieron entre categorías de vegetación (árbol, arbusto y hierba y pasto verde) y entre superficies artificiales con tonos y texturas similares (pavimento, suelo desnudo y techo de losa); siendo la categoría arbusto la más compleja de identificar, debido a su frecuente confusión con las categorías árbol, hierba y pasto verde. El MLP fue seleccionado como clasificador final para la identificación automática de coberturas del paisaje, debido a su desempeño superior frente a los demás algoritmos evaluados.

En conjunto, los resultados respaldan el uso combinado de la segmentación por superpíxeles y técnicas de aprendizaje automático para la clasificación de elementos del paisaje urbano a partir de imágenes aéreas. Se reconoce la necesidad de incrementar la robustez del modelo mediante su entrenamiento y validación con imágenes aéreas de otras colonias de Tapachula, Chiapas. Las categorías del paisaje urbano identificadas en este estudio pueden ser empleadas en el análisis del paisaje donde ocurre la transmisión de dengue, aportando así una base técnico-operativa con aplicaciones potenciales en el monitoreo y el control de mosquitos vectores.

AGRADECIMIENTOS

Los autores agradecen a la Facultad de Ciencias Biológicas (FCB) de la Universidad Autónoma de Nuevo León (UANL), por el uso de las imágenes obtenidas durante la tesis doctoral “Uso de drones para la asociación de factores de riesgo con la abundancia de mosquitos *Aedes aegypti* (Linnaeus) Diptera: Culicidae, en áreas de transmisión de dengue de la ciudad de Tapachula, Chiapas” de la doctora Kenia Mayela Valdez Delgado <http://eprints.uanl.mx/id/eprint/25097>. Al Centro Nacional de Prevención de Desastres (CENAPRED) por su asesoría y apoyo en la obtención de imágenes aéreas. Al Programa de Maestría en Cómputo Estadístico del Centro de Investigación en Matemáticas A. C. (CIMAT) por el trabajo conjunto en la formación de estudiantes mediante el desarrollo de herramientas basadas en machine learning. A Reyes Maya, K. M. (2022) y a Méndez Vázquez, V. I. (2024) por su contribución al desarrollo de la herramienta de etiquetado para la segmentación y la extracción de características por superpíxel. Este trabajo forma parte del proyecto SECIHTI CBF-2025-G-1351, proyecto apoyado por la Secihti en el año 2025.

REFERENCIAS

1. Hunsperger, E. A., Sharp, T. M., Lalita, P., Tikomaidraubuta, K., Rebello Cardoso, Y., Naivalu, T., Khan, A. S., Marfel, M., Hancock, W. T., Tomashek, K. M., & Margolis, H. S. (2016). Use of a rapid test for diagnosis of dengue during suspected dengue outbreaks in resource-limited regions. *Journal of Clinical Microbiology*, 54(8), 2042–2047. <https://doi.org/10.1128/JCM.00521-16>.
2. Betanzos-Reyes AF, Rodríguez-López MH, Huerta-Jiménez H, Mejía-Mendoza TG, Dzul-Manzanilla FA, Correa-Morales F, Larios-Morales J. (2025). *Aedes aegypti* y arbovirosis. Factores que determinan la presencia del vector y pautas para su control integral sostenido. Cuernavaca, Morelos: *Instituto Nacional de Salud Pública*, https://spmediciones.mx/libro/aedes-aegypti-y-arbovirosis_162599/.
3. Pando-Robles, V., Alvarez-Obregón, J., Díaz del Castillo-Flores, G., González-Roldán, J. F., Saúl Raga-Sarabia, E., González-Acosta, C., & Méndez-Galván, J. F. (2025). Epidemiological and

- Economic Burden of Dengue in Mexico: Data Analysis from 2010 to 2020. *The American Journal of Tropical Medicine and Hygiene*, 113(6), 1363-1374. Retrieved Jan 28, 2026, from <https://doi.org/10.4269/ajtmh.24-0777>.
4. Pando Robles, V. (2024, noviembre 20). Vacuna contra el dengue: una herramienta adicional para la prevención. Instituto Nacional de Salud Pública. <https://portal.insp.mx/informacion-relevante/vacuna-contra-el-dengue-una-herramienta-adicional-para-la-prevencion>.
5. Carrasco-Escobar, G., Manrique, E., Ruiz-Cabrejos, J., Saavedra, M., Álava, F., Bickersmith, S., Prussing, C., Vinetz, J. M., Conn, J. E., Moreno, M., & Gamboa, D. (2019). High-accuracy detection of malaria vector larval habitats using drone-based multispectral imagery. *PLoS Neglected Tropical Diseases*, 13(1), e0007105. <https://doi.org/10.1371/journal.pntd.0007105>.
6. Aragão, F. V., Cavicchioli Zola, F., Nogueira Marinho, L. H., de Genaro Chiroli, D. M., Braghini Júnior, A., & Colmenero, J. C. (2020). Choice of unmanned aerial vehicles for identification of mosquito breeding sites. *Geospatial Health*, 15(1), 810. <https://doi.org/10.4081/gh.2020.810>.
7. Mukabana, W. R., Welter, G., Ohr, P., Tingitana, L., Makame, M. H., Ali, A. S., & Knols, B. G. J. (2022). Drones for Area-Wide Larval Source Management of Malaria Mosquitoes. *Drones*, 6(7), 180. <https://doi.org/10.3390/drones6070180>.
8. Valdez-Delgado, K.M.; Moo-Llanes, D.A.; Danis-Lozano, R.; Cisneros-Vázquez, L.A.; Flores-Suarez, A.E.; Ponce-García, G.; Medina-De la Garza, C.E.; Díaz-González, E.E.; Fernández-Salas, I. (2021). Field Effectiveness of Drones to Identify Potential *Aedes aegypti* Breeding Sites in Household Environments from Tapachula, a Dengue-Endemic City in Southern Mexico. *Insects*. 12, 663 <https://doi.org/10.3390/insects12080663>.
9. Valdez-Delgado, K. M., Garcia-Salazar, O., Moo-Llanes, D. A., Izcapa-Treviño, C., Cruz-Pliego, M. A., Domínguez-Posadas, G. Y., Armendáriz-Valdez, M. O., Correa-Morales, F., Cisneros-Vázquez, L. A., Ordóñez-González, J. G., Fernández-Salas, I., & Danis-Lozano, R. (2023). Mapping the Urban Environments of *Aedes aegypti* Using Drone Technology. *Drones*, 7(9), 581. <https://doi.org/10.3390/drones7090581>.
10. Muñoz-Sánchez, V., Valdez-Delgado, K. M., Hernández-López, F. J., Moo-Llanes, D. A., González-Farías, G., & Danis-Lozano, R. (2022). Use of unmanned aerial vehicles for building a house risk index of mosquito-borne viral diseases. *Machines*, 10(12), 1161. <https://doi.org/10.3390/machines10121161>.
11. Ferraguti, M., Martínez-de la Puente, J., Roiz, D., Ruiz S., Soriguer R., Figuerola J. (2016). Effects of landscape anthropization on mosquito community composition and abundance. *Sci Rep* 6, 29002 <https://doi.org/10.1038/srep29002>.
12. Fernandez-Salas I. (2009). Biología y control de *Aedes aegypti*, Manual de operaciones. Universidad Autonoma de Nuevo León. 2da. Edición. Monterrey, Nuevo León, México. ISBN 978-607-433-320-6.
13. Ortiz, D. I., Piche-Ovares, M., Romero-Vega, L. M., Wagman, J., & Troyo, A. (2022). The Impact of Deforestation, Urbanization, and Changing Land Use Patterns on the Ecology of Mosquito and Tick-Borne Diseases in Central America. *Insects*, 13(1), 20. <https://doi.org/10.3390/insects13010020>.
14. Cofone, L., Sabato, M., Di Paolo, C., Di Giovanni, S., Donato, M. A., & Paglione, L. (2025). Urban, Architectural, and Socioeconomic Factors Contributing to the Concentration of Potential Arbovirus Vectors and Arbovirolos in Urban Environments from a One Health Perspective: A Systematic Review. *Sustainability*, 17(9), 4077. <https://doi.org/10.3390/su17094077>.
15. Stanton, M. C., Kalonde, P., Zembere, K., Spaans, R. H., & Jones, C. M. (2021). The application of drones for mosquito larval habitat identification in rural environments: a practical approach for malaria control? *Malaria Journal*, 20(1), 244. <https://doi.org/10.1186/s12936-021-03759-2>.
16. Hay, G.J., Castilla, G. (2008). Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline. In: Blaschke, T., Lang, S., Hay, G.J. (eds) Object-Based Image Analysis. Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-77058-9_4.

17. Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3–4), 239–258.
18. Blaschke, T., Lang, S., & Hay, G. J. (Eds.). (2008). *Object-based image analysis: Spatial concepts for knowledge-driven remote sensing applications*. Springer. <https://doi.org/10.1007/978-3-540-77058-9>.
19. OpenCV contributors. (2025). OpenCV: ximgproc::SuperpixelSLIC [Documentación]. *OpenCV.org*. Recuperado el 14 de octubre de 2025, de https://docs.opencv.org/3.4/d3/da9/classcv_1_1ximgproc_1_1SuperpixelSLIC.html.
20. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11), 2274–2282.
21. Ren, X., & Malik, J. (2003). Learning a classification model for segmentation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 10–17). IEEE.
22. Hernández Girón, A., González Navarro, Y. E., & Morales Blas, A. (2017, julio). Implementación de algoritmo de superpíxeles para la segmentación de imágenes a color. *Boletín UPIITA*, (61). <https://www.boletin.upiita.ipn.mx/Archivos/Articulos/2026/05/c050ad7eb4054a42bc772f842effcbb4.pdf>.
23. Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing* (4th ed.). Pearson.
24. Valverde Rebaza, J. (2007). Detección de bordes mediante el algoritmo de Canny [Manuscrito]. Universidad Nacional de Trujillo. <https://www.researchgate.net/publication/267240432>.
25. Sonka, M., Hlavac, V., & Boyle, R. (2014). *Image processing, analysis, and machine vision* (4th ed.). Cengage Learning.
26. Ventura Angoa, S. (2025). *segmentacion-automatizada-paisaje-urbano-dengue* [Repositorio de código]. GitHub. Recuperado el 16 de octubre de 2025, de <https://github.com/SarahiVenturaaaa/segmentacion-automatizada-paisaje-urbano-dengue.git>.
27. Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.
28. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
29. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.