

TECNOLÓGICO NACIONAL DE MÉXICO
INSTITUTO TECNOLÓGICO DE CD. VICTORIA

SECRETARÍA DE
EDUCACIÓN PÚBLICA

SEP

TecnoINTELECTO

Órgano de Divulgación Científica

Una Publicación del Tecnológico Nacional de México
– Instituto Tecnológico de Cd. Victoria

Volumen 22

No. 1

Julio 2025

ISSN 1665-983X

INGENIERÍA Y TECNOLOGÍA

Desarrollo de una herramienta para la detección de HLB en árboles de cítricos. Jesús Carlos Carmona-Frausto, Adriana Mexicano-Santoyo, Kevin Eduardo Bee-Cruz, Salvador Cervantes-Álvarez & Lilia Mexicano-Santoyo 1

Clasificación, agrupamiento y descubrimiento de tópicos con inteligencia artificial en textos sobre la enfermedad de SJÖGREN. Ana Laura Lezama-Sánchez & Mireya Tovar-Vidal 12



DIRECTORIO

Dr. Ramón Jiménez López

Director General del Tecnológico Nacional de México

Mtra. Deysi Yesica Álvarez Vergara

Directora

M.I.I. Ana María Magdalena Ortiz Elizalde

Subdirectora Académica

Arq. Evaristo Ángel Romero Nava

Subdirector de Planeación y Vinculación

Mtro. José Ángel Nieto Meza

Subdirectora de Servicios Administrativos

COMITÉ EDITORIAL

Instituto Tecnológico de Cd. Victoria

División de Estudios de Posgrado e Investigación

COORDINACIÓN EDITORIAL

Adriana Mexicano-Santoyo, Dra.

Jésus Carlos Carmona-Frausto, Dr.

Asistencia Editorial:

Jonatan Axel Carrizales-López, Ing.

COMITÉ REVISOR

Dr. Franco Rojas-López. Universidad Politécnica Metropolitana de Puebla, Ingeniería Sistemas Computacionales.

Dra. Nelva Nely Almanza-Ortega. Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), SNII Nivel 1.

Dr. Vítervo López-Caballero. Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET). Departamento de Comunicación y Eventos, SNII Nivel C.

Dra. Hilda Castillo-Zacatelco. Benemérita Universidad Autónoma de Puebla. Facultad Ciencias de la Computación, Perfil Deseable PRODEP.

TecnoINTELECTO (ISSN 1665-983X y reserva: 04-2004-072626452400-102) es un órgano de divulgación científica de forma semestral del Tecnológico Nacional de México-Instituto Tecnológico de Cd. Victoria. Avenida Tecnológico No. 1301, C. P. 87010, Cd. Victoria, Tamaulipas, México; Tels. (834) 153 20 00 Ext. 364. El contenido y la sintaxis de los artículos presentados son responsabilidad del autor (es). Editor Principal: División de Estudios de Posgrado e Investigación. Apoyo editorial-informático: Dr. Jesús Carlos Carmona-Frausto. Envío de documentos, consultas y sugerencias al correo electrónico: tecnointelecto@cdvictoria.tecnm.mx. Todos los derechos son reservados y propiedad del Tecnológico Nacional de México-Instituto Tecnológico de Cd. Victoria. TecnoINTELECTO, Vol. 22 No. 1. Cd. Victoria, Tamaulipas, México.

Consúltanos en el Índice Latinoamericano www.latindex.org y en el
Índice de Revistas Latinoamericanas en Ciencias PERIÓDICA
www.dgb.unam.mx/periodica.html



DESARROLLO DE UNA HERRAMIENTA PARA LA DETECCIÓN DE HLB EN ÁRBOLES DE CÍTRICOS

Jesús Carlos Carmona-Frausto¹, Adriana Mexicano-Santoyo¹, Kevin Eduardo Bee-Cruz¹, Salvador Cervantes-Álvarez², Lilia Mexicano-Santoyo³

¹Tecnológico Nacional de México, Instituto Tecnológico de Cd. Victoria. Avenida Tecnológico No. 1301 C. P. 87010 Ciudad Victoria, Tamaulipas, México.

jesus.cf@cdvictoria.tecnm.mx
adriana.ms@cdvictoria.tecnm.mx^{*}
m23380019@cdvictoria.tecnm.mx

²Centro Universitario de los Valles, Universidad de Guadalajara. Carretera Guadalajara-Ameca Km 45.5, C.P. 46600, Ameca, Jalisco, México
salvador.cervantes7964@academicos.udg.mx

³Universidad de Guanajuato, Campus Celaya-Salvatierra. Av. Ing. Barros Sierra No. 201, Ejido de Santa María del Refugio, C.P. 38140, Celaya, Guanajuato, México
l.mexicano@ugto.mx

Resumen. El Huanglongbing (HLB) es una enfermedad que afecta de manera crítica a los cultivos de cítricos, reduciendo su productividad y calidad. Con el propósito de apoyar en su detección, se desarrolló una herramienta basada en visión artificial y aprendizaje supervisado. La metodología se dividió en tres fases: cálculo de índices de vegetación, análisis de datos y clasificación. Se emplearon imágenes multiespectrales RGB y NIR obtenidas con una cámara RedEdge-P, sobre huertos de cítricos de la variedad Naranja Valencia. A partir de estas imágenes se calcularon veinte índices de vegetación, aplicando técnicas de reducción de datos (AC, ACP, AE, AC-ACP, AC-AE, mRMR y Transformers) y clasificadores SVM, kNN, NB y Ensemble Learning, con sus variantes. El modelo que tuvo mejor desempeño fue Transformers NB Multinomial con sus métricas de evaluación 70% de Exactitud, 69% de Precisión, 70% de Recuerdo, 69% de F1 y 55% de Cohen Kappa. Los resultados muestran que la herramienta permite identificar de forma visual y diferenciada las clases Fondo, Sanos y Enfermos, proporcionando una alternativa práctica para el monitoreo fitosanitario. Asimismo, ofrece una base para el desarrollo de futuros proyectos orientados a la detección de enfermedades en distintos cultivos.

PALABRAS CLAVE: Visión Artificial, Agricultura de Precisión, Huanglongbing, Aprendizaje Supervisado.

Abstract. Huanglongbing (HLB) is a disease that critically affects citrus crops, reducing their productivity and quality. To aid in its detection, a tool based on computer vision and supervised learning was developed. The methodology was divided into three phases: calculation of vegetation indices, data analysis, and classification. Multispectral RGB and NIR images obtained with a RedEdge-P camera were used, covering Valencia orange orchards. From these images, twenty vegetation indices were calculated, applying data reduction techniques (AC, PCA, AE, AC-PCA, AC-AE, mRMR, and Transformers) and SVM, kNN, NB, and Ensemble Learning classifiers, with their variants. The best-performing model was Transformers NB Multinomial, with the following evaluation metrics: 70% Accuracy, 69% Precision, 70% Recall, 69% F1, and 55% Cohen's Kappa. The results show that the tool allows for the visual and distinct identification of Background, Healthy, and Diseased classes, providing a practical alternative for phytosanitary monitoring. It also offers a foundation for the development of future projects focused on disease detection in various crops.

KEYWORDS: Artificial Vision, Precision Agriculture, Huanglongbing, Supervised Learning.

1. INTRODUCCIÓN

El **Huanglongbing (HLB)** es una enfermedad que afecta a los árboles de cítricos, deteriorando su calidad fisiológica y reduciendo de manera significativa la producción de frutos a lo largo de su ciclo de vida (Bové, 2006). Esta enfermedad se caracteriza por alterar la coloración de las hojas hacia tonos amarillos, provocar malformaciones en los frutos y convertirse en el hábitat del insecto vector que la transmite, ocasionando finalmente la muerte del árbol. El HLB es propagado por el psílido asiático, una plaga difícil de controlar; una vez que un árbol resulta infectado, no existe una cura disponible (Carmona et al., 2025).

Frente a esta situación, se han implementado diversas técnicas para contrarrestar sus efectos, como la prueba de **Polymerase Chain Reaction (PCR)** (Nazir, 2020), medidas de cuarentena, eliminación de árboles infectados y control del vector de la enfermedad (Halbert & Manjunath, 2004). Sin embargo, los resultados obtenidos han sido limitados, lo que ha impulsado la incorporación de la **Agricultura de Precisión (AP)** (Madhukar, 2021). Esta disciplina aplica enfoques tecnológicos al manejo de los cultivos con el fin de mejorar el rendimiento, optimizar las condiciones de vida y controlar plagas de manera más eficiente. Entre sus herramientas se encuentra la **teledetección o Remote Sensing (RS)**, que permite obtener y analizar datos de la superficie terrestre sin necesidad de contacto directo (Graham, 1999).

El RS se lleva a cabo mediante equipos conformados por drones, cámaras espectrales o multiespectrales y sistemas de **Visión Artificial (VA)**, disciplina que constituye el eje central de este proyecto. La VA dota a las computadoras de la capacidad de percibir y analizar el mundo real a través de imágenes (García & Caranqui, 2015). En este campo, se han desarrollado estudios relevantes, tales como: la clasificación de hojas afectadas por HLB mediante espectroscopía de fluorescencia y aprendizaje automático (Wetterich et al., 2013); la detección utilizando una longitud de onda de 591 nm, filtros polarizados y características texturales de las hojas (Pourreza et al., 2013); el uso de imágenes de banda estrecha y filtros polarizados para identificar acumulaciones de almidón en hojas infectadas (Pourreza et al., 2014); el diseño de

sensores de visión para detección en tiempo real (Pourreza et al., 2015); la aplicación de redes neuronales de retropropagación (BPNN) (Xiaoling et al., 2016); y la detección temprana de la enfermedad mediante sensores especializados (Pourreza et al., 2017). Asimismo, se han evaluado arquitecturas de redes neuronales profundas como AlexNet, VGG16, VGG19, ResNet18, GoogLeNet e Inception-V3 Gómez et al., 2022), y se han desarrollado sistemas integrados de visión para la identificación rápida en campo (Xu et al., 2023). También se han empleado imágenes térmicas e infrarrojas (Sankaran et al., 2013), cámaras multiespectrales como la **Micasense RedEdge** instaladas en UAVs (Dadras et al., 2019), así como comparaciones entre plataformas aéreas multiespectrales e hiperespectrales para evaluar su rendimiento en la detección del HLB (García et al., 2013). Finalmente, se han realizado estudios que confirman la viabilidad de la teledetección a diferentes alturas de vuelo en huertos de cítricos (Lan et al., 2020).

En este contexto, el presente proyecto propone la construcción de una herramienta para la detección de HLB en árboles de cítricos. Dicha herramienta se fundamenta en el uso de aprendizaje **supervisado** y segmentación píxel a píxel, con el fin de ubicar con precisión las zonas afectadas, clasificadas en Fondo, Árboles Enfermos y Árboles Sanos. Para ello, se emplearon principalmente Índices de Vegetación (IV) definidos por la reflectancia de las imágenes. Posteriormente, se generaron conjuntos de datos mediante técnicas como **Análisis de Correlación (AC)**, **Análisis de Componentes Principales (ACP)**, **AutoEncoder (AE)**, **AC-ACP**, **AC-AE**, **mínima Redundancia Máxima Relevancia (mRMR)** y **Transformers**. Finalmente, la clasificación de las imágenes se llevó a cabo utilizando algoritmos como **Máquinas de Vectores de Soporte (SVM)**, **k-Vecinos más Cercanos (kNN)**, **Bayes Ingenuo (NB)** y **Ensemble Learning (EL)**.

Construcción de la herramienta

La herramienta fue desarrollada en dos etapas: la etapa de entrenamiento, en la cual se construyó el modelo del algoritmo, y la etapa de prueba, enfocada en la implementación de los modelos de clasificación con las variantes entrenadas, mostrando los resultados de manera visual en la interfaz gráfica. La Figura 1 presenta

el diseño general de la composición de la herramienta.

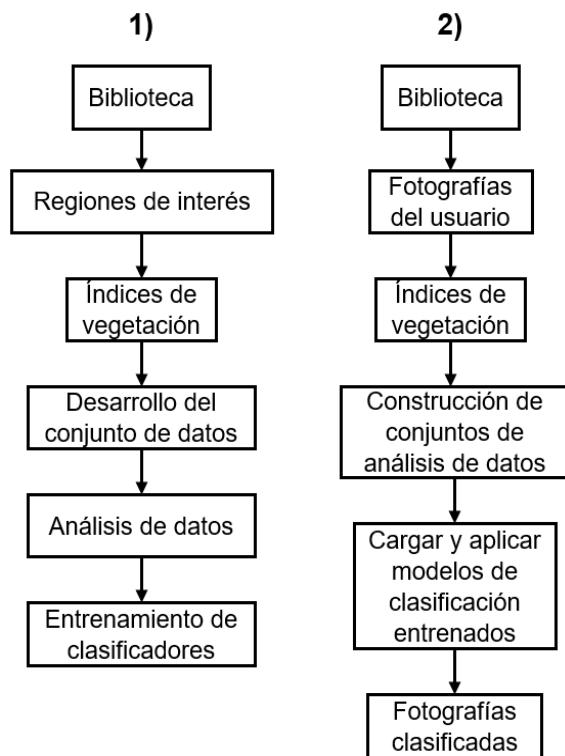


Fig. 1. Etapas de desarrollo de la herramienta: A) Etapa de entrenamiento; B) Etapa de prueba.

A) Etapa de entrenamiento:

Durante esta etapa se instalaron las bibliotecas y módulos Scikit-learn, Pandas, Tensorflow, Seaborn, Matplotlib, Numpy, Joblib, Pickle, OpenCV, Random, Time y Pymrnr, los cuales constituyen la base para la construcción del modelo entrenado. Posteriormente, se segmentaron y cargaron las regiones de interés (ROI) construidas en el software ENVI, consistentes en fotografías multiespectrales con los canales Red, Green y NIR.

Estas fotografías fueron capturadas en zonas de Ciudad Victoria, Tamaulipas, y alrededores, mediante un drone Matrice RTK 350, una cámara multiespectral RedEdge-P y un código de ajuste en Python para normalizar las dimensiones. Para el entrenamiento del modelo se utilizaron 141 fotografías de distintas huertas enfocadas en la fruta cítrica Naranja Valencia. Del total de imágenes, se destinó un 80% a entrenamiento y 20% a prueba en los clasificadores SVM, KNN y

NB. En el caso de EL, la distribución fue 90% entrenamiento y 10% prueba.

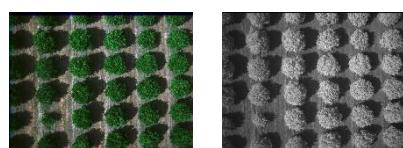
Posteriormente, se calcularon los índices de vegetación (IV) de las fotografías, determinando las clases Fondo, Enfermos y Sanos, que representan los objetos de interés clave en este trabajo. Luego, se estableció la clase para cada grupo de datos generado a partir de los IV, obteniéndose un conjunto unificado que albergaba las tres clases.

A este conjunto se le aplicaron los análisis AC, ACP, AE, AC-ACP, AC-AE, mRMR y Transformers, con el objetivo de reducir y eliminar datos redundantes, mejorando el entrenamiento de los clasificadores al disminuir el ruido y el tiempo de procesamiento. Finalmente, los siete nuevos conjuntos de datos obtenidos mediante dichos análisis fueron procesados por los clasificadores:

- SVM, con kernel polinomial, gaussiano y lineal.
- kNN, con variantes de pesos uniformes y pesos ponderados.
- NB, con modelos gaussiano, multinomial y Bernoulli.
- EL, compuesto por AdaBoost con DecisionTree, XGB y RandomForest, utilizando los métodos Soft Voting y Hard Voting.

B) Etapa de prueba:

La etapa de prueba comprendió la implementación de las bibliotecas, la importación de las imágenes multiespectrales de los cultivos cítricos (Figura 2), el uso de los IV, el desarrollo de los conjuntos de análisis de datos configurados en la etapa de entrenamiento, la importación de los modelos de clasificación previamente entrenados y, finalmente, la obtención de las fotografías clasificadas.



A) Imagen RGB B) Imagen NIR

Fig. 2. Ejemplo de las imágenes que emplea la herramienta. A) Imagen en canales RGB; B) Imagen en canal NIR.

2. METODOLOGÍA

Para implementar la herramienta se definieron tres fases principales, representadas en la Figura 3:

A) Cálculo de los índices de vegetación, orientado a identificar el estado actual de las plantas a partir de un rango de valores numéricos.

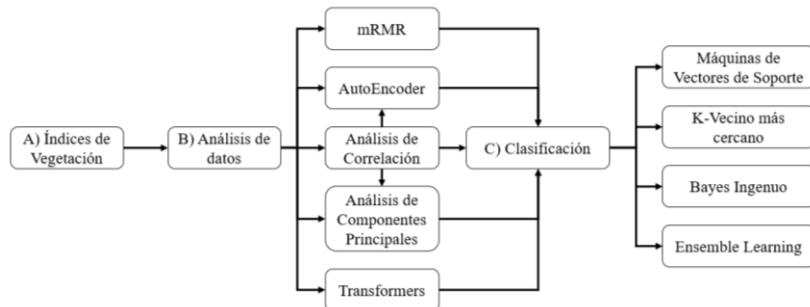


Fig. 3. Fases de la metodología empleada en la herramienta

B) Análisis de datos, enfocado en la construcción de los conjuntos de datos que servirán como base para la clasificación.

C) Clasificación, consistente en la aplicación de diversos algoritmos sobre los conjuntos de datos generados en la fase anterior generados en la fase anterior.

A) Índices de vegetación

Los índices de vegetación tienen como objetivo mostrar, mediante una gama de colores, las regiones que forman parte de una planta u objetos de interés en una imagen. Esto se debe a la reflectancia que emiten: si el estado de salud de una vegetación es sano o enfermo, refleja una radiación distinta, lo cual permite desarrollar proyectos enfocados en la identificación del estado en que se encuentra un cultivo.

Para calcular un IV es necesario utilizar canales espectrales específicos provenientes de una cámara. Un ejemplo es el Índice de Vegetación de Diferencia Normalizada (NDVI), definido en la Ecuación 1:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (1)$$

El NDVI utiliza los canales espectrales NIR y RED para estudiar el estado de los cultivos. El objetivo de este índice, al igual que los otros veinte considerados en este trabajo, es aportar atributos a una clase [19]. Dicho de otro modo, cada IV puede otorgar valores distintos a cada

elemento de interés, por lo cual fueron indispensables para el entrenamiento de los clasificadores, ya que permiten identificar a qué clase pertenecen los píxeles de las fotografías. La identificación de las distintas clases (Fondo, Enfermos, Sanos) se logra a través de la gama de colores de cada índice de vegetación, representada en diferentes valores. Una vez aplicados los IV sobre las fotografías multiespectrales, se obtiene un conjunto de datos que integra las tres clases, el cual fue evaluado mediante **AC, ACP, AE, AC-ACP, AC-AE, mRMR y Transformers**.

B) Análisis de datos

El empleo de los veinte IV puede resultar redundante, debido a que algunos presentan valores numéricos similares. Por ello, el análisis de datos ofrece una alternativa para reducir la cantidad de índices y, al mismo tiempo, disminuir el tiempo de procesamiento de la herramienta. En este trabajo se utilizaron **AC, ACP, AE, AC-ACP y AC-AE**.

Análisis de Correlación (AC)

El AC analiza cada variable mediante el coeficiente producto-momento de Pearson. Su particularidad consiste en señalar cuáles variables tienen un alto índice de correlación (entre 0 y 1), eliminando aquellas que aportan información redundante y que incrementan el tiempo de procesamiento y el ruido durante la clasificación. En la etapa de entrenamiento se estableció un rango de eliminación de correlación de 0.94–1, lo cual descartó los

índices DVI, OSAVI, IPVI, MTVI2, RDVI, GRNDVI, Norm NIR, SR y MTVI1.

Análisis de Componentes Principales (ACP)

El ACP selecciona las variables con mayor varianza, representando el conjunto de datos sin necesidad de conservar todas las variables originales. En este proyecto se eligieron tres componentes principales ya que obtuvieron varianzas de 0.7455, 0.1224 y 0.0411, que en conjunto alcanzaron una varianza total de 0.9091. Esto se debe a que, al solo ser 3 variables, reducían el tiempo de procesamiento y presentaban mejores resultados que utilizando más variables para el ACP.

AutoEncoder (AE)

El AE es una red neuronal diseñada para reducir el número de variables, mejorar la velocidad de ejecución y conservar los datos fundamentales. El proceso consta de un **Codificador**, que comprime las variables de entrada hasta alcanzar un “código latente”, y un **Decodificador**, que reconstruye la información original a partir de ese código. Durante las veinte épocas de entrenamiento, la red registró pérdidas de 5.4569e-04, 5.2696e-04, 5.3073e-04, 5.3097e-04, hasta llegar a 4.9005e-04 en la última iteración, valores cercanos a cero que indican buen rendimiento.

AC-ACP

Esta técnica combina AC y ACP para reducir aún más el número de variables. Con los once IV restantes tras aplicar AC, se seleccionaron tres componentes principales, con varianzas de 0.6252, 0.2020 y 0.0687, sumando una varianza total de 0.8960. Al igual que con ACP, la mejor opción fue utilizar 3 variables y no una mayor cantidad.

AC-AE

Combina AC y AE con el propósito de minimizar las variables seleccionadas tras el análisis de correlación mediante un AutoEncoder. En este caso, las tres neuronas resultantes alcanzaron pérdidas de 0.0138, 0.0117, 0.0120 y 0.0047, hasta llegar a 0.0030 en la última época.

Mínima Redundancia Máxima Relevancia (mRMR)

Esta técnica, propuesta por Peng et al., 2005, selecciona las variables más relevantes del conjunto inicial. En este proyecto se eligieron tres índices: Banda Roja Normalizada (NormR), Banda Verde Normalizada (NormG) e Índice de Cociente Simple (SR).

Transformers

Este modelo de transducción de secuencias, basado en la atención, reemplaza las capas tradicionales de codificadores y decodificadores utilizadas en los AE por un mecanismo de **multi-headed self-attention**. Según Vaswani et al., 2017 esta arquitectura optimiza el procesamiento de secuencias. En este proyecto, el modelo alcanzó pérdidas entre 0.6779 y 0.6775, consideradas óptimas por su proximidad a cero.

C) Clasificación

Máquinas de Vectores de Soporte (SVM)

Se aplicó SVM a los conjuntos de datos obtenidos en el análisis, dado que su aplicación se orienta a la clasificación [22]. El algoritmo define un hiperplano que divide las características en distintas regiones. Los kernels empleados fueron **gaussiano**, **lineal** y **polinomial**.

k-Vecino más cercano (kNN)

El objetivo de kNN es clasificar un nuevo punto de datos estimando su pertenencia a una clase según su cercanía con los vecinos. En este proyecto se utilizaron las variantes de **pesos uniformes** y **pesos ponderados** de la biblioteca Scikit-learn.

Bayes Ingenuo (NB)

NB se denomina “ingenuo” porque asume que cada variable contribuye de manera independiente a la clasificación, aplicando el teorema de Bayes. En este trabajo se emplearon las variantes **gaussianas**, **multinomial** y **Bernoulli** para identificar las clases Fondo, Enfermos y Sanos.

Ensemble Learning (EL)

EL combina varios clasificadores y toma la decisión de clasificación según el consenso: si coinciden dos o tres, la predicción se acepta. En este proyecto se implementaron **AdaBoost** con

DecisionTree como estimador base, XGB y RandomForest, utilizando las técnicas de votación **Soft Voting (SV)** y **Hard Voting (HV)**.

Métricas de evaluación

Para la evaluación de los clasificadores se utilizó La Ecuación 2 mide la Exactitud (%) que consiste en el porcentaje de las predicciones correctas divididas con el total de predicciones efectuadas. La Ecuación 3 mide el Recall (%) que consiste en la habilidad del clasificador para detectar de manera eficiente las instancias positivas. La Ecuación 4 mide la Precisión (%) que consiste en la capacidad del clasificador para detectar de manera eficaz las instancias positivas divididas con el total de instancia que predijo como positivas. La Ecuación 5 mide el F1 que consiste en una mezcla entre las medidas Precisión y Recall para evaluar el rendimiento del clasificador.

$$\text{Exactitud} (\%) = \frac{TN + TP}{TN + TP + FN + FP} \times 100 \quad (2)$$

$$\text{Recall} (\%) = \frac{TP}{TP + FN} \times 100 \quad (3)$$

$$\text{Precisión} (\%) = \frac{TP}{TP + FP} \times 100 \quad (4)$$

$$F1(\%) = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \quad (5)$$

3. IMPLEMENTACIÓN DE LA HERRAMIENTA

El objetivo principal de la herramienta es generar las clases Fondo, Enfermos y Sanos a partir de las fotografías insertadas. Para su uso es necesario emplear fotografías multiespectrales, que deben incluir los canales RGB combinados y la imagen del canal NIR, como se muestra en la Figura 4.

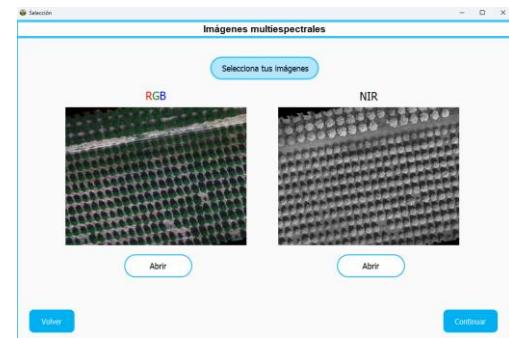


Fig. 4. Selección de imágenes multiespectrales (RGB y NIR) necesarias para el funcionamiento de la herramienta.

Posteriormente, la herramienta aplica los 20 índices de vegetación (IV) al conjunto de imágenes insertadas, utilizando los canales separados de las fotografías en diferentes operaciones matemáticas. Después, se aplica AC al conjunto de datos obtenido a partir de los IV, ya que no corresponde a un modelo implementado, a diferencia de ACP, AE, AC-ACP, AC-AE, mRMR y Transformers, los cuales sí fueron implementados porque su configuración se estableció durante la etapa de entrenamiento.

Una vez cargado cada modelo, estos se aplican sobre el conjunto de datos de los IV. A continuación, se cargan los modelos de clasificación SVM, kNN, NB y EL, junto con cada una de las variantes construidas en la etapa de entrenamiento. Finalmente, se ejecutan los modelos de clasificación y, como resultado, se obtienen las imágenes clasificadas, las cuales pueden ser visualizadas a través de la herramienta.

4. PRUEBA

Una vez insertadas y procesadas las imágenes, la herramienta proporciona dos opciones para visualizar los resultados: "Índices de vegetación" e "Imagenes clasificadas", como se observa en la Figura 5.

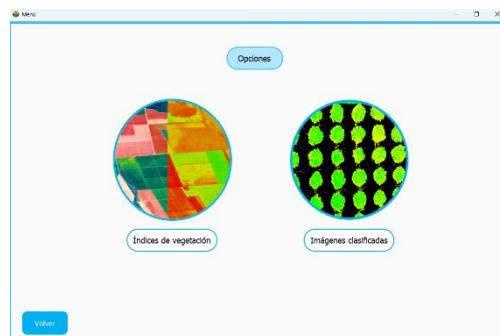
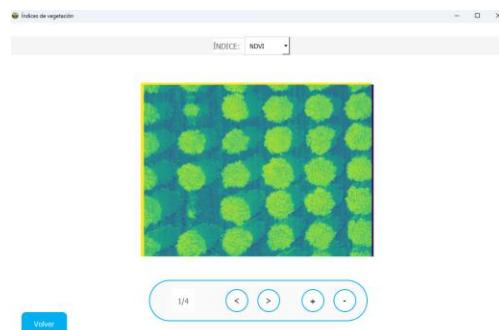


Fig. 5. Menú principal de la herramienta con las opciones de visualización: "Índices de vegetación" e "Imágenes clasificadas".

En el menú "Índices de vegetación" es posible visualizar las imágenes generadas al aplicar los veinte IV sobre cada imagen ingresada. Estos índices permiten analizar cómo se representan ciertas regiones mediante una gama de colores ajustada a cada índice, como se muestra en la Figura 6. La gama de colores de la Figura 6 indica que los árboles presentan un tono verde con algunas pigmentaciones amarillas, lo que demuestra que no están completamente enfermos, a diferencia del Fondo, que en su mayoría se muestra en tonos azulados.



Por otro lado, se encuentra la opción de "Imágenes clasificadas". Esta opción presenta los resultados de los cuatro clasificadores desarrollados durante la etapa de entrenamiento e implementados en la herramienta: SVM, kNN, NB y EL, lo cual se observa en la Figura 7. Cada uno de estos clasificadores genera una clasificación de acuerdo con las operaciones definidas en su arquitectura.

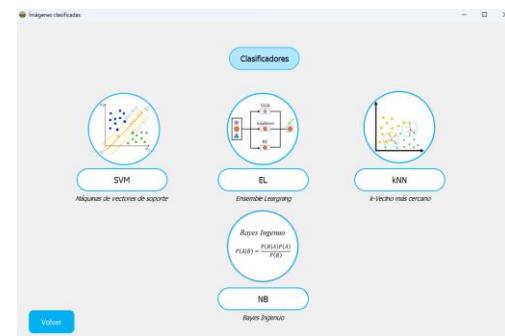


Fig. 7. Interfaz de selección de clasificadores disponibles en la herramienta: SVM, kNN, NB y EL.

El análisis de los resultados permite considerar la alternativa de que, si varias clasificaciones coinciden en señalar la misma región, esa área corresponde a la enfermedad de interés de este proyecto, es decir, HLB.

5. RESULTADOS

Para alcanzar los resultados se emplearon fotografías en formato RGB y NIR. Estas fotografías contenían los objetos de interés, es decir, Fondo, Enfermos y Sanos. Una vez clasificadas, la herramienta proporcionó la visualización de los resultados mediante los 70 conjuntos de imágenes elaborados a partir del uso de SVM con sus tres variantes, kNN con sus dos variantes, NB con sus tres variantes y EL con sus dos técnicas de votación, ejecutando cada clasificador sobre los siete conjuntos de datos analizados. Los mejores clasificadores obtuvieron los resultados que se presentan en la Tabla 1.

Tabla 1. Clasificadores con mejores resultados.

Métrica	AC-AE EL SV	Transformers NB Multinomial	ACP SVM Kernel Gaussiano	ACP kNN Pesos Uniformes
Exactitud	0.70	0.70	0.69	0.66
Precisión	0.69	0.69	0.68	0.66
Recall	0.69	0.70	0.698	0.66
F1	0.69	0.69	0.68	0.66

Después de aplicar la prueba del Estadístico de Friedman con resultados de Friedman = 8.0000, p-valor = 0.0183 indicando que hay diferencias

significativas entre los clasificadores de la Tabla 1. Luego, analizar cada uno de los resultados de los mejores modelos. Se concluyó que el modelo Transformers NB multinomial quien tiene los mejores resultados. En la Tabla 2 es posible observar su matriz de confusión.

Tabla 2. Matriz de confusión del modelo Transformers NB Multinomial.

Métrica	Fondo	Sanos	Enfermos
Precisión	0.74	0.62	0.71
Recall	0.80	0.48	0.81
F1	0.77	0.54	0.75

La Figura 8 presenta el resultado obtenido por la aplicación de Transformers con el clasificador kNN en su variante Pesos uniformes. Esta imagen permite comparar la fotografía ingresada por el usuario con la imagen ya clasificada, donde se muestran porcentajes de 53.87% Fondo, 25.80% Enfermos y 20.33% Sanos. Estos valores representan la proporción total de píxeles pertenecientes a cada clase en la imagen.

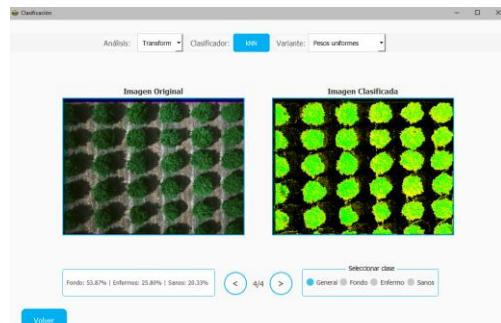


Fig. 8. Resultado de clasificación utilizando Transformers con kNN (Pesos uniformes), mostrando las proporciones de cada clase (Fondo, Enfermos y Sanos).

Además, la herramienta brinda la opción de pigmentar los píxeles de cada clase en la imagen original, como se observa en la Figura 9, donde se utilizó AC con EL (SV), resaltando en color verde los píxeles correspondientes a la clase Sanos en la Imagen Original.

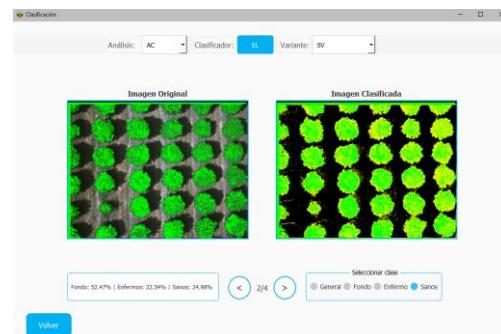


Fig. 9. Ejemplo de pigmentación de píxeles en la imagen original con la técnica AC y Ensemble Learning (Soft Voting), resaltando en verde las áreas correspondientes a la clase Sanos.

Por otro lado, en la Figura 10 se presentan los resultados obtenidos con Transformers SVM utilizando su kernel Gaussiano, donde se alcanzaron porcentajes de 56.85% Fondo, 22.19% Enfermos y 20.97% Sanos, pigmentando de color amarillo los píxeles de la clase Enfermos. Estos resultados permiten concluir que el clasificador obtiene salidas razonables, ya que detecta las tres clases, considerando que el árbol no está totalmente enfermo, pero también identificando áreas externas a los árboles, representadas en negro con ligeros pigmentos amarillos, lo que refleja su eficacia en la detección.

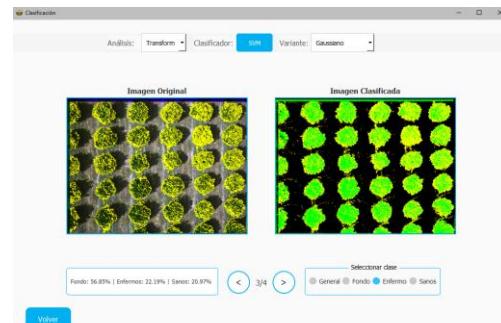


Fig. 10. Clasificación mediante Transformers SVM con kernel Gaussiano, resaltando en amarillo los píxeles de la clase Enfermos.

Finalmente, la Figura 11 muestra los resultados de clasificación obtenidos al aplicar Transformers NB Multinomial. En esta imagen se observa la clasificación de las tres clases con 54.18% Fondo, 18.05% Enfermos y 27.77% Sanos, donde los píxeles de la clase Fondo se pigmentaron de color negro. Esta clasificación se considera aceptable, ya que permite

interpretar de manera clara las regiones correspondientes a cada clase.

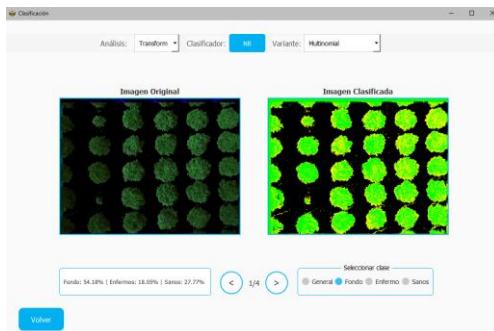


Fig. 11. Clasificación con Transformers NB Multinomial, mostrando la distribución porcentual de las clases Fondo, Enfermos y Sanos, con pigmentación en negro para la clase Fondo.

En general, los resultados se consideran satisfactorios, dado que las pruebas se realizaron en un campo donde los árboles sí presentaban la enfermedad, aunque no de forma avanzada, ya que los agricultores suelen aplicar técnicas para prolongar la calidad de vida de los cítricos. Por esta razón, en la mayoría de los árboles se observan zonas tanto sanas como enfermas; sin embargo, es importante destacar que el HLB es una enfermedad progresiva que consume al portador de manera paulatina.

6. CONCLUSIONES

El desarrollo de una herramienta para la detección de HLB en árboles de cítricos mediante técnicas de visión artificial es una tarea compleja, ya que para obtener los resultados presentados fue necesario adquirir conocimientos sobre la manipulación de imágenes multiespectrales, emplear repositorios adecuados de imágenes, contar con el equipo especializado para la captura fotográfica, y solicitar el apoyo de otras disciplinas como la agronomía para realizar las exploraciones de campo. Asimismo, se requirió capacitación en el manejo de vehículos aéreos no tripulados, la recolección de datos en distintos cultivos para conformar un repositorio de imágenes representativo, el ajuste y segmentación de regiones de interés, así como el desarrollo de la etapa de entrenamiento que incluyó la aplicación de índices de vegetación, los análisis de datos (AC, ACP, AE, AC-ACP, AC-AE, mRMR, Transformers) y la implementación de

modelos de clasificación con sus respectivas variantes: SVM (polinomial, gaussiano, lineal), kNN (pesos uniformes y pesos ponderados), NB (gaussiano, multinomial, Bernoulli) y EL (AdaBoost con DecisionTree, XGB y RandomForest con Soft Voting y Hard Voting).

Es fundamental mencionar que las fotografías utilizadas fueron adquiridas mediante una cámara multiespectral RedEdge-P, que al contar con los canales Red, Green, Blue y NIR permitió generar las imágenes RGB y NIR necesarias para el funcionamiento de la herramienta.

Finalmente, se obtuvieron resultados visuales que permiten interpretar de manera clara las tres clases de interés (Fondo, Enfermos y Sanos). Estos resultados demuestran que la visión artificial combinada con aprendizaje supervisado para la segmentación de píxeles constituye una alternativa eficaz, brindando una base sólida para la continuidad de futuros proyectos que busquen identificar zonas de interés en otros cultivos o detectar diferentes enfermedades a partir de la reflectancia de los índices de vegetación.

Esta herramienta representa una propuesta efectiva y práctica para los agricultores, al ofrecer un apoyo tecnológico aplicable en la producción de jugos, frutas, aceites, mermeladas y otros derivados de los cítricos.

7. TRABAJOS FUTUROS

En el futuro se espera que este trabajo cimente bases para la visión artificial empleando el aprendizaje supervisado ya sea aplicado en distintos tipos de cultivos para analizar las enfermedades o anomalías que pueden llegar a afectar su calidad de vida y la producción agrícola. Por lo cual, podría incluso utilizarse para identificar los distintos elementos que se encuentren en una imagen siempre y cuando se haya considerado para su detección en el entrenamiento previo.

8. LITERATURA CITADA

- Bové, J. M. (2006). Huanglongbing: A destructive, newly-emerging, century-old disease of citrus. *Journal of Plant Pathology*, 88(1), 7–37.

- Carmona, J., Mexicano, A., Bee, K., García, L., & Mexicano, L. (2025). Optimization and artificial vision: Innovative tool for detecting huanglongbing in citrus. *DSPA*, 4(1), 1–15.
- Dadras, F., Samadzadegan, F., Hossein, S., & Fazeli, H. (2019). UAV-based multispectral imagery for fast citrus greening detection. *Journal of Plant Diseases and Protection*, 126(1), 307–318.
- Deng, X., Lan, Y., Xing, X., Ma, H., Li, J., & Huang, Y. (2016). Detection of citrus huanglongbing based on image feature extraction and two-stage BPNN modeling. *International Journal of Agricultural and Biological Engineering*, 9(6), 20–26.
- García, F., Sankaran, S., Maja, J. M., Suk, W., Rasmussen, J., & Ehsani, R. (2013). Comparison of two aerial imaging platforms for identification of huanglongbing-infected citrus trees. *Computers and Electronics in Agriculture*, 91(1), 106–115.
- García, I., & Caranqui, V. (2015). La visión artificial y los campos de aplicación. *Tierra Infinita*, 1(1), 98–108. <https://doi.org/10.32645/26028131.76>
- Gómez, W., Garza, J., & Edmundo, S. (2022). A huanglongbing detection method for orange trees based on deep neural networks and transfer learning. *IEEE Access*.
- Graham, S. (1999). *Remote sensing*. Academic Press.
- Gupta, P., & Sehgal, N. (2021). *Introduction to machine learning in the cloud with Python*. Springer Cham. <https://doi.org/10.1007/978-3-030-71270-9>
- Halbert, S., & Manjunath, K. (2004). Asian citrus psyllids (Sternorrhyncha: Psyllidae) and greening disease of citrus: A literature review and assessment of risk in Florida. *Florida Entomologist*, 87(3), 330–353.
- Lan, Y., Huang, Z., Deng, X., Zhu, Z., Huang, H., Zheng, Z., Lian, B., Zeng, G., & Tong, Z. (2020). Comparison of machine learning methods for citrus greening detection on UAV multispectral images. *Computers and Electronics in Agriculture*, 171, 105234. <https://doi.org/10.1016/j.compag.2020.105234>
- Madhukar, N. (2021). Precision agriculture: Smart farming taken to the next level. *Natural Volatiles & Essential Oils*, 8(5), 4316–4324.
- Nazir, I. (2020). Polymerase chain reaction (PCR): A short review. Department of Plant Breeding and Genetics, University of Agriculture Faisalabad.
- Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8).
- Pourreza, A., Suk, W., Raveh, E., Hong, Y., & Kim, H. (2013). Identification of citrus greening disease using a visible band image analysis. *ASABE Annual International Meeting Paper*, 1(1).
- Pourreza, A., Lee, W., Raveh, E., Ehsani, R., & Etxeberria, E. (2014). Citrus huanglongbing detection using narrow-band imaging and polarized illumination. *American Society of Agricultural and Biological Engineers*, 57(1), 259–272.
- Pourreza, A., Suk, W., Ehsani, R., Schueller, K., & Raveh, E. (2015). An optimum method for real-time in-field detection of huanglongbing disease using a vision sensor. *Computers and Electronics in Agriculture*, 110(1), 221–232.
- Pourreza, A., Suk, W., Czarnecka, E., Verner, L., & Gurley, W. (2017). Feasibility of using the optical sensing techniques for early detection of huanglongbing in citrus seedlings. *Robotics*, 6(11).
- Sankaran, S., Mari, J., Buchanon, S., & Ehsani, R. (2013). Huanglongbing detection using visible, near-infrared and thermal imaging techniques. *Sensors*, 13(2), 2117–2130.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need.

- Wetterich, C., Kumar, R., Sankaran, S., Belasque, J., Ehsani, R., & Marcassa, L. (2013). A comparative study on application of computer vision and fluorescence imaging spectroscopy for detection of huanglongbing citrus disease in the USA and Brazil. *Journal of Spectroscopy*, 1(1).
- Xu, Q., Cai, J., Ma, L., Tan, B., Li, Z., & Sun, L. (2023). Custom-developed reflection-transmission integrated vision system for rapid detection of huanglongbing based on the features of blotchy mottled texture and starch accumulation in leaves. *Plants*, 12(616).

CLASIFICACIÓN, AGRUPAMIENTO Y DESCUBRIMIENTO DE TÓPICOS CON INTELIGENCIA ARTIFICIAL EN TEXTOS SOBRE LA ENFERMEDAD DE SJÖGREN

Ana Laura Lezama-Sánchez^{1,2}, Mireya Tovar-Vidal²

¹Secretaría de Ciencia, Humanidades, Tecnología e Innovación, Ciudad de México, México

²Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla, 72570 Puebla, México,

analaura.lezama@correo.buap.mx, mireya.tovar@correo.buap.mx

RESUMEN. En este artículo proponemos el uso de técnicas de *machine learning* para analizar textos automáticamente. Para ello, se partió de un conjunto de documentos sobre enfermedades autoinmunes, específicamente sobre el síndrome de *Sjögren*, que fue preprocesado con el fin de eliminar ruido y palabras irrelevantes. Posteriormente, se entrenó un modelo de clasificación supervisada mediante *random forest* para categorizar los textos. Después, se aplicó *K-Means* para agrupar documentos similares en *clusters*. Finalmente, aplicamos un modelo de tópicos que nos permitió identificar los principales temas presentes en el corpus. Este pipeline ilustra cómo la inteligencia artificial puede contribuir a descubrir patrones, organizar información y extraer conocimiento de grandes volúmenes de texto.

PALABRAS CLAVE: Aprendizaje Automático, Enfermedades Sistémicas, Enfermedades Órgano Específicas

ABSTRACT. In this article, we propose using machine learning techniques to automatically analyze texts. We started with a set of documents on autoimmune diseases, specifically Sjögren's syndrome, and preprocessed them to remove noise and irrelevant words. Then we trained a supervised classification model using random forests to categorize the texts. After that, we applied K-Means to group similar documents into clusters. Finally, we applied a topic model to identify the main themes in the corpus. This pipeline illustrates how artificial intelligence can help uncover patterns, organize information, and extract knowledge from large volumes of text.

KEYWORDS: Machine Learning, Systemic Diseases, Organ-Specific Diseases

1. INTRODUCCIÓN

Las enfermedades autoinmunes son patologías en las que el sistema inmunológico encargado de defender al organismo frente a agentes externos pierde la capacidad de reconocer lo propio y ataca por error células, tejidos u órganos del cuerpo. Esto provoca inflamación y daño en distintas partes del organismo.

Estas enfermedades se clasifican en dos tipos.

- Órgano-específicas: afectan principalmente a un solo órgano.
- Sistémicas: cuando comprometen múltiples órganos o sistemas.

Aunque su causa exacta aún no ha sido descubierta, se reconoce que intervienen factores genéticos, ambientales y alteraciones en el funcionamiento del sistema inmunológico (Jadue y González, 2012).

Dentro de este grupo se encuentra el síndrome de *Sjögren*, caracterizado por la infiltración de células inmunitarias en las glándulas exocrinas,

lo que provoca sequedad ocular y bucal. También puede presentar manifestaciones sistémicas y afectar diferentes órganos. Su diagnóstico suele ser complejo debido a la variabilidad de síntomas y a la superposición con otras enfermedades autoinmunes (Aulla et al., 2024).

La diversidad de manifestaciones clínicas y la abundancia de información médica generada (historias clínicas, informes, literatura científica, etc.) hacen que el manejo de estas enfermedades sea un reto tanto para la medicina clínica como para la investigación. En este contexto, surge la necesidad de utilizar herramientas computacionales que apoyen la organización y análisis de grandes volúmenes de información.

En este artículo, presentamos la aplicación de métodos de aprendizaje automático para el análisis de textos médicos vinculados al síndrome de *Sjögren*. A pesar de que el conjunto de datos ya está categorizado (por ejemplo, en

función de la existencia o ausencia de la enfermedad), nuestro objetivo es integrar varios métodos que faciliten su estudio. Esto incluye: la clasificación supervisada para anticipar etiquetas en nuevos textos, el agrupamiento para hallar patrones encubiertos y el análisis de tópicos para detectar temas comunes. Esta perspectiva intenta demostrar cómo la inteligencia artificial puede ayudar a estructurar, interpretar y producir conocimiento a partir de datos médicos complejos.

2. MARCO TEÓRICO

En esta sección se exponen los conceptos empleados en el desarrollo de este trabajo.

En términos generales, el aprendizaje automático (*machine learning*) es un área de la inteligencia artificial que facilita a las computadoras el análisis de datos y el aprendizaje de patrones, sin que estén específicamente programadas para cada función específica. En el ámbito médico, estos métodos se utilizan para obtener información de grandes cantidades de datos y respaldar decisiones clínicas, lo que es especialmente útil en enfermedades complejas como las autoinmunes. En el ámbito del aprendizaje automático, hay tres métodos que son particularmente pertinentes para este análisis:

Clasificación supervisada: Este método facilita el entrenamiento de modelos con datos etiquetados previamente, con la finalidad de clasificar nueva información en categorías preestablecidas. En este trabajo, se utiliza para clasificar automáticamente entre textos que hacen referencia al síndrome de *Sjögren* y aquellos que no, u otras categorías clínicas de interés (Cardoso, Ferneda, y Botega, 2023).

Agrupamiento (Clustering): El agrupamiento es un método de aprendizaje no supervisado que identifica grupos de datos con atributos parecidos, evitando la necesidad de etiquetas previas. Este procedimiento puede ser beneficioso para identificar patrones encubiertos, como subconjuntos de documentos que tienen temas o perspectivas parecidas, lo cual es útil en contextos de alta variabilidad clínica o textual (Ting et al., 2025).

Descubrimiento de tópicos: Método empleado para descubrir temas recurrentes en extensas colecciones de texto. Algoritmos como Latent Dirichlet Allocation (LDA) facilitan la

identificación de grupos de palabras que a menudo se presentan en conjunto y su organización en tópicos, lo que simplifica la exploración semántica del corpus. Esto posibilita identificar si en los documentos se mencionan con mayor frecuencia síntomas, tratamientos, causas u otros elementos relacionados (Lezama-Sánchez et al., 2025).

3. ESTADO DEL ARTE

En esta sección se presentan algunos trabajos existentes que analizan datos relacionados con la enfermedad de *Sjögren* mediante el uso de técnicas de inteligencia artificial. Estos estudios aplican diferentes procedimientos y algoritmos para procesar, clasificar o predecir información clínica.

Las investigaciones comienzan con (Gan et al., 2025), quienes exploraron el uso de algoritmos de aprendizaje automático para diferenciar entre el síndrome de *Sjögren* con poliartritis (SS-PA) y el síndrome de superposición artritis reumatoide-*Sjögren* (RA-SS), dos entidades con manifestaciones clínicas similares, pero con implicaciones terapéuticas distintas. El estudio lo realizaron con 241 pacientes, y utilizaron técnicas como la regresión *LASSO* para la selección de variables, así como la construcción de modelos diagnósticos empleando cuatro algoritmos de *machine learning*. El algoritmo que mostró el mejor desempeño fue *Random Forest*, con un *AUC* de 0.854 (IC 95%: 0.747–0.944), superando incluso a la regresión logística tradicional. Además, el análisis SHAP identificó como principales variables discriminantes los niveles de anti-CCP, factor reumatoide (FR), número de articulaciones erosivas, anticuerpos anti-SSA/Ro60 y proteína C reactiva (PCR). Estos hallazgos permitieron concluir a los autores que el aprendizaje automático podría mejorar significativamente la diferenciación diagnóstica entre RA-SS y SS-PA, ofreciendo herramientas de apoyo clínico más precisas y personalizadas.

Por otro lado, (Módis et al., 2025) propusieron un enfoque que integra aprendizaje automático para evaluar el peso predictivo de variables biológicas, psicológicas y sociales en el síndrome de *Sjögren* primario (pSS). A partir de una cohorte de 117 pacientes, los autores construyeron un conjunto de datos que incluye medidas objetivas (autoanticuerpos SSA/SSB) y

subjetivas (el índice autorreportado ESSPRI) de la carga de la enfermedad. Para modelar la influencia relativa de los distintos dominios biopsicosociales, entrenaron algoritmos de aprendizaje automático utilizando cada grupo de variables por separado y luego de forma combinada. La importancia de las características predictoras se estimó mediante permutaciones. Los resultados revelaron que, mientras los marcadores biológicos fueron más útiles para predecir los autoanticuerpos, el modelo que mejor estimó la carga subjetiva (ESSPRI) fue el que integró todas las dimensiones. Entre los predictores más relevantes se identificaron tanto variables clínicas (IgG, RF, edad, conteo plaquetario) como psicológicas, destacando la ansiedad como rasgo de personalidad, que fue un predictor negativo de la carga biológica. Por otro lado, el estado de ansiedad y el rasgo temperamental de fatigabilidad se asociaron positivamente con la carga percibida. El rasgo de personalidad conciencia se asoció negativamente con el ESSPRI, lo que sugiere un posible efecto modulador del perfil de personalidad sobre la experiencia subjetiva de la enfermedad.

(Kise et al., 2025) propusieron un modelo predictivo para el diagnóstico del síndrome de Sjögren (SS) basado en técnicas de radiómicas y aprendizaje automático aplicadas a imágenes de tomografía computarizada (TC) de las glándulas parótidas. En su estudio, analizaron 132 glándulas parótidas de 66 sujetos (33 con SS y 33 controles), extrayendo un total de 129 características radiómicas incluyendo variables de primer orden, forma y textura a partir de segmentaciones manuales realizadas en 3D *Slicer*. El modelo fue generado automáticamente mediante la plataforma *Prediction One*, logrando un desempeño notable en la validación interna (AUC = 0.92; precisión = 0.90; *recall* = 0.85; F1 = 0.88) y una eficacia aún mayor en la validación temporal posterior (AUC = 0.96; precisión = 0.85; *recall* = 0.92). Estos resultados evidencian el potencial del enfoque *radiomics-machine learning* para diferenciar de forma no invasiva las glándulas afectadas por SS, con un flujo de trabajo eficiente y reproducible. Además, la implementación automatizada mediante *Prediction One* optimizó el análisis de datos y la generación del modelo, reduciendo considerablemente el tiempo

requerido, lo cual sugiere una futura aplicación. (Xue. et al., 2025) utilizaron una metodología fundamentada en la metabolómica fecal no dirigida y algoritmos de aprendizaje automático para crear una estrategia de diagnóstico innovadora para pSS (El síndrome de Sjögren primario). Mediante la espectrometría de masas combinada con cromatografía líquida (LC-MS), obtuvieron información de 93 pacientes con pSS y 42 controles saludables. Mediante técnicas de selección de características como SHAP (*Shapley Additive Explanations*) y métodos de decisión multicriterio (MCDM), identificaron 10 metabolitos clave de un total de 151. Posteriormente, implementaron un modelo de ensamble con seis algoritmos de *machine learning*, logrando una precisión diagnóstica destacada (AUC de 0.98, *recall* de 0.97, especificidad de 0.90). Además, desarrollaron un sistema visual amigable para facilitar su aplicación clínica. Este análisis evidenció la capacidad de fusionar la inteligencia artificial con la metabolómica para enfrentar los desafíos diagnósticos en patologías autoinmunes complejas como el pSS.

El síndrome de Sjögren primario (pSS) y la tiroiditis autoinmune (AIT) tienen varios rasgos genéticos e inmunológicos en común, lo que indica una potencial conexión patogénica entre ambos. En este marco, (Wu et al., 2025) llevaron a cabo una investigación con el propósito de anticipar la existencia de autoanticuerpos tiroideos específicos (TPOAb y TgAb) en pacientes con pSS a través de algoritmos de aprendizaje automático. Los escritores examinaron información clínica y de laboratorio de 96 pacientes, segmentándola en base a la positividad de estos autoanticuerpos. Adicionalmente, utilizaron cuatro algoritmos de aprendizaje automático, de los cuales el clasificador Random Forest registró el rendimiento más alto (AUC = 0.755). Este modelo detectó las siguientes variables predictivas: la edad, los niveles de IgG, el componente del complemento C4 y la existencia de sequedad en la boca. Los resultados sugirieron que el uso de modelos de aprendizaje automático permite anticipar el desarrollo de AIT en pacientes con pSS, lo que representa un avance potencial para la detección temprana y la personalización del manejo clínico en este grupo de pacientes.

(Li et al., 2023) propusieron un modelo llamado AFKF (*ALBERT-based Fusion Kalman Filter*) para la clasificación de textos médicos, combinando el modelo de lenguaje ALBERT con un filtro de Kalman para reducir el ruido en los registros médicos electrónicos. Mediante una estrategia de ventana deslizante y un bloque de fusión de características, el modelo logró mejorar la precisión entre un 13 % y un 20 % frente a los métodos tradicionales consultados por los autores durante el desarrollo de su trabajo. Además, su enfoque incrementó el rendimiento de clasificadores clásicos como SVM y regresión logística, evidenciando su eficacia en la tarea de clasificación de texto clínico.

Wang et al. (2023) propusieron un modelo de clasificación de textos basado en la combinación de redes convolucionales de grafos (GCN) y modelado de temas mediante LDA. El enfoque construye un grafo heterogéneo utilizando relaciones documento-palabra y cocurrencias de palabras (calculadas con Point-Wise Mutual Information (PMI) y TF-IDF), transformando la tarea en un problema de clasificación de nodos. Los resultados experimentales mostraron que el modelo supera a métodos existentes en cinco conjuntos de datos de referencia, destacando su capacidad para capturar dependencias semánticas complejas en el texto.

En este trabajo a pesar de que el conjunto de datos seleccionado para el desarrollo de la aproximación propuesta está previamente categorizado, distintas investigaciones han demostrado el valor de aplicar enfoques complementarios de aprendizaje automático para extraer capas adicionales de conocimiento (Steyvers y Griffiths, 2007; Singh et al., 2020). En este contexto, nuestro análisis combina:

- Clasificación supervisada, que permite entrenar modelos para predecir categorías clínicas en textos no etiquetados;
- Técnicas de agrupamiento (*clustering*), utilizadas para detectar estructuras o grupos semánticos dentro de los datos;
- Modelos de tópicos, como LDA, que permiten descomponer grandes volúmenes de texto en temáticas subyacentes de forma no supervisada.

Esta estrategia integrada permite no solo validar las etiquetas existentes, sino también descubrir relaciones temáticas, similitudes semánticas y posibles sesgos en la información médica textual.

4. TRABAJO PROPUESTO

En esta sección, se describe un flujo de análisis de textos médicos mediante técnicas de aprendizaje automático. El objetivo es demostrar cómo distintos métodos pueden combinarse para clasificar documentos, agruparlos según similitudes y extraer los temas predominantes.

Para ello, se empleó un conjunto de textos relacionados al síndrome de Sjögren. El proceso seguido incluye las siguientes etapas:

Carga de datos: Se emplearon dos conjuntos de datos en formato CSV, uno para entrenamiento (80%) y otro para prueba (20%). Cada archivo contenía textos clínicos junto con su respectiva clase diagnóstica. Se validó que los datos estuvieran correctamente etiquetados y sin registros duplicados.

Condiciones de experimentación: Todos los experimentos se llevaron a cabo usando Python 3.10. Las bibliotecas utilizadas incluyen *scikit-learn*, *gensim*, y *nltk*.

Preprocesamiento del texto: Se diseñaron tres rutinas específicas para distintas tareas:

- **Clasificación:** El texto fue convertido a minúsculas, se eliminaron caracteres no alfabéticos, se tokenizó y se retiraron *stopwords*. El resultado fue una versión limpia para el vectorizador *TF-IDF*.
- **Clustering:** Se aplicó una limpieza similar a la anterior, pero sin eliminar *stopwords*, para preservar la densidad semántica del texto.
- **Modelado de temas (Topic Modeling):** Se generaron listas de tokens a partir de los textos, también filtrados por *stopwords*, para su posterior representación en forma de bolsa de palabras.

Entrenamiento supervisado: Se entrenó un modelo de *Random Forest* utilizando una representación vectorial *TF-IDF* del texto preprocesado. El modelo fue evaluado sobre el conjunto de prueba, y su desempeño se midió mediante métricas estándar como accuracy, precisión, recall y F1. Estos valores se obtuvieron utilizando validación cruzada con 5 particiones (*k-fold cross-validation*), lo que contribuyó a evaluar la confiabilidad y estabilidad del modelo.

Controles aplicados: Para evitar sobreajuste, se aplicaron técnicas como validación cruzada y ajuste de hiperparámetros.

Análisis no supervisado (Clustering): Se aplicó el algoritmo *k-means* sobre vectores *TF-IDF* para agrupar los textos en 5 clusters. La calidad del agrupamiento se evaluó mediante la métrica de *Silhouette Score*, que permitió validar la coherencia de los grupos formados.

Modelado de temas (LDA): A partir de los tokens generados, se construyó un diccionario y un corpus compatible con el modelo LDA. Además, se extrajeron cinco temas latentes representados por los términos más relevantes. La coherencia semántica de los tópicos fue validada usando la métrica de coherencia de tópicos.

Por lo tanto, este procedimiento permite realizar un análisis supervisado y no supervisado de notas clínicas, con potencial para la clasificación automatizada, el agrupamiento exploratorio y la extracción de temas latentes en el contexto del diagnóstico de enfermedades autoinmunes.

Por lo que, se puede concluir que el contenido textual de las notas clínicas contiene información estructurada útil para apoyar tareas de predicción y análisis de tópicos. La combinación de técnicas de validación cruzada, métricas de desempeño y pruebas de coherencia proporciona un respaldo metodológico a los resultados obtenidos.

La Figura 1 muestra la arquitectura general del pipeline desarrollado, el cual integra las etapas de preprocesamiento, representación vectorial y análisis supervisado y no

supervisado de los textos clínicos.

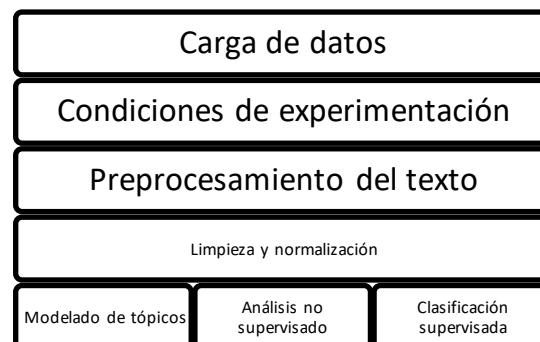


Figura 1 Pipeline metodológico del estudio expuesto

5. RESULTADOS

En esta sección se presentan y analizan los resultados obtenidos a partir de la implementación de tres enfoques principales sobre las notas clínicas: clasificación supervisada, agrupamiento (clustering) y descubrimiento de tópicos.

El conjunto de datos utilizado incluyó un total de 1,850 sentencias clínicas, de las cuales 1,480 se emplearon para el entrenamiento y 370 para las pruebas. Está compuesto por notas clínicas en inglés, clasificadas en cuatro categorías semánticas relacionadas con el diagnóstico de la enfermedad de Sjögren. Cada documento contiene un fragmento de texto clínico y su correspondiente clase anotada, perteneciente a una de las siguientes categorías: *Signs and Symptoms*, *Diagnostic Tests*, *Associated Conditions* y *Risk Factors*. La distribución de ejemplos entre los conjuntos de entrenamiento y prueba se presenta en la Tabla 1. La categoría con mayor número de instancias corresponde a *Signs and Symptoms*, seguida de *Diagnostic Tests*, mientras que *Risk Factors* presenta la menor cantidad de ejemplos.

En la tarea de clasificación, se utilizó un modelo de *Random Forest* entrenado con representaciones *TF-IDF* del texto preprocesado. El modelo alcanzó un desempeño general con una precisión de 0.7645.

Tabla 1 Descripción del conjunto de datos seleccionado

Clase	Entrenamiento	Prueba
Associated Conditions	362	78
Diagnostic Tests	450	104
Risk Factors	69	25
Signs and Symptoms	605	163

Para la clasificación supervisada, se entrenó un modelo de bosque aleatorio (*Random Forest*) sobre los textos preprocesados utilizando la técnica *TF-IDF* (*Term Frequency Inverse Document Frequency*), que permite representar el texto en función de la importancia relativa de cada término tanto dentro de un documento como en el conjunto total del corpus. Esta representación permite que el algoritmo de clasificación identifique patrones relevantes y tome decisiones basadas en el contenido textual. Como resultado, el modelo obtuvo los desempeños globales que se presentan en la Tabla 2, los cuales fueron calculados sobre el conjunto de prueba. Estos reflejan el rendimiento general sin distinguir entre clases.

Tabla 2 Resultados globales del modelo *Random Forest*

Métrica	Resultado
Precisión	0.7645
Recall	0.7622
F1	0.7608
Accuracy	0.7622

A nivel de clases, la Tabla 3 detalla las métricas de rendimiento obtenidas para cada categoría dentro del conjunto de prueba. Se observa que la clase *Risk Factors* obtuvo un desempeño inferior en comparación con los demás, lo cual puede estar relacionado con la menor cantidad de ejemplos disponibles para su entrenamiento.

Tabla 3 Resultados por clase del modelo *Random Forest*

Clase	Precisión	Recall	F1
Associated Conditions	0.73	0.81	0.77
Diagnostic Tests	0.76	0.86	0.81
Risk Factors	0.52	0.48	0.50
Signs and Symptoms	0.82	0.72	0.77

Los resultados globales del modelo *Random Forest* (ver Tabla 2) muestran un desempeño moderado, con un F1 general de 0.7608. Este rendimiento es comparable con lo reportado por Li et al. (2023) y Wang et al. (2023), quienes emplearon modelos de clasificación de textos clínicos y obtuvieron un F1 entre 0.72 y 0.80. Al analizar los resultados por clase (ver Tabla 3), se observa una disminución notable en la clase *Risk Factors*, posiblemente debido al desbalance de clases y a la complejidad semántica de dicha categoría, fenómeno también discutido por Gan et al. (2025) y (Módis et al., 2025). Por lo tanto, futuras mejoras en esta investigación podrían explorar técnicas de balanceo de clases o representaciones contextuales como ALBERT para mejorar este desempeño.

En la Tabla 4 se presentan algunos ejemplos de predicciones realizadas por el modelo. Endonde se puede observar que logra identificar correctamente la mayoría de las clases, aunque en ciertos casos puede confundirse cuando los textos contienen términos que pertenecen a múltiples categorías clínicas.

Aunque el corpus está anotado con cuatro clases para la tarea de clasificación, en el análisis no supervisado se consideró la posibilidad de una estructura temática más compleja. Por ello, se optó por obtener cinco grupos y cinco tópicos, lo que permitió una segmentación semánticamente más completa del contenido textual. Esta elección se justificó al observar que los grupos generados mostraban temas distintos entre sí y con contenido interno consistente, lo que sugiere que cinco clústeres y cinco tópicos reflejan una segmentación significativa del corpus.

Tabla 4 Ejemplos de predicciones generadas por el modelo de clasificación

Texto original	Clase original	Clase predicha
It can remain asymptomatic or manifest with metabolic abnormalities including hypokalemia paralysis, hypocalcaemia	Signs and Symptoms	Associated Conditions
There are few reported cases of concordance in monozygotic twins, and it is common for patients with	Associated Conditions	Associated Conditions
From this perspective, salivary proteomics has appeared as a promising tool considering that salivary proteins may	Diagnostic Tests	Diagnostic Tests
Furthermore, high IHC scores of membranous EpEX were independent risk factors for pSS, even for the pSS patients at early	Risk Factors	Risk Factors

En el análisis no supervisado, se aplicó *k-means* con 5 grupos y se evaluó con la métrica de *Silhouette Score*, que arrojó un valor de 0.61, lo cual indica una segmentación razonable con buena separación entre grupos. Esto respalda la validez del número de clústeres seleccionados. En la Tabla 5, se exponen ejemplos representativos de cada grupo. Los clústeres muestran una coherencia del tópico, agrupando, por ejemplo, textos relacionados con pruebas serológicas, comorbilidades, y síntomas pediátricos. Esto concuerda con lo reportado por (Luque Guzmán, 2020), donde usaron TF-IDF combinando con *k-means* permitió identificar dominios clínicos diferenciados en textos médicos.

Por otro lado, en el descubrimiento de tópicos (LDA), se identificaron cinco tópicos principales. El modelo alcanzó una coherencia del tópico de 0.55, medida con el índice de coherencia *C_V*. Lo cual indica una calidad aceptable en la

representación de los tópicos, permitiendo interpretar los tópicos latentes con base en las palabras clave obtenidas (ver Tabla 6).

Con base en los tópicos obtenidos por ejemplo *fatigue* y *arthritis* se observa que existe una estrecha relación con definiciones clínicas relevantes para el diagnóstico del síndrome de Sjögren, como se ha evidenciado también en estudios previos (Srockel et al., 2014), donde LDA logró extraer conceptos clínicos latentes consistentes con clasificaciones médicas establecidas.

Sin embargo, es importante destacar algunas limitaciones: El corpus presenta desbalance de clases, lo que puede haber afectado el rendimiento del modelo supervisado. Además, aunque los grupos y tópicos son interpretables, la validación humana (por expertos clínicos) podría fortalecer la evaluación semántica.

Por lo que, en general, con base en los resultados obtenidos deducimos que el contenido textual de las notas clínicas contiene patrones útiles para tareas de clasificación y segmentación temática, con posibles aplicaciones en sistemas de diagnóstico. Sin embargo, consideramos que replicar los experimentos con otros corpus y validar contra corpus como SNOMED CT o UMLS puede fortalecer la generalización del enfoque propuesto.

Tabla 5 Resultados del proceso de agrupamiento en 5 grupos o clusters

Grupo	Resultado
0	Casos relacionados con linfoma y enfermedades autoinmunes en familiares.
1	Diagnóstico clínico y procedimientos diagnósticos.
2	Confirmación de síndrome de Sjögren mediante pruebas serológicas.
3	Diagnósticos diferenciales y comorbilidades.
4	Presentaciones pediátricas y síntomas como parotiditis y fiebre.

El agrupamiento mediante *k-means* mostró una segmentación semántica coherente, con un *Silhouette Score* de 0.61, valor que indica una separación razonable entre los grupos. Estos resultados son comparables a los reportados en estudios que aplican *k-means* y modelos de tópicos a textos clínicos (Steyvers y Griffiths, 2007; Singh et al., 2020), donde los valores

típicos de separación oscilan entre 0.55 y 0.65. La coherencia interna de los clústeres sugiere que el lenguaje clínico presenta patrones temáticos recurrentes que pueden ser capturados incluso mediante representaciones no contextuales. Esto refuerza la pertinencia del enfoque para análisis exploratorio y construcción de perfiles clínicos preliminares.

Tabla 6 Resultados de los cinco tópicos identificados

Grupo	Resultado
0	gland, salivary, biopsy, lymphoma, diagnosis
1	dry, symptoms, eye, fatigue, mouth
2	patients, manifestations, sicca, fatigue, arthritis
3	syndrome, sjögren, systemic, autoimmune
4	parotid, lupus, erythematosus, anti, gland

El modelo de tópicos basado en LDA permitió identificar estructuras semánticas latentes asociadas a síntomas, pruebas diagnósticas y condiciones comórbidas (Tabla 6), lo cual concuerda con hallazgos previos en la literatura sobre síndrome de Sjögren (Srockel et al., 2014). Aunque la coherencia del tópico alcanzada (0.55) se considera moderada, se ubica dentro del rango esperado para textos clínicos con vocabulario altamente especializado (0.45–0.60). Esto indica que los tópicos son interpretables y clínicamente relevantes, aunque su validación podría fortalecerse mediante retroalimentación de especialistas médicos.

En conjunto, el modelo demuestra que el contenido textual de las notas clínicas contiene patrones informativos relevantes para tareas de clasificación y segmentación temática, lo cual puede ser aprovechado en sistemas de apoyo al diagnóstico. Sin embargo, se identifican dos limitaciones principales: (1) el desbalance de clases afecta el desempeño del modelo supervisado, y (2) la interpretación semántica de los grupos y tópicos requiere validación clínica adicional. En futuros trabajos se considera evaluar representaciones contextuales como *BioBERT* o *ClinicalBERT*, así como ampliar el corpus e integrar ontologías biomédicas como *SNOMED CT* o *UMLS* para mejorar la generalización.

6. CONCLUSIONES

En este trabajo se desarrolló un pipeline de procesamiento de lenguaje natural (PLN) sobre textos clínicos relacionados con el diagnóstico del síndrome de Sjögren, integrando enfoques supervisados (*Random Forest* + *TF-IDF*) y no supervisados (*K-Means* y *LDA*) para la extracción de conocimiento.

En base a los resultados obtenidos, podemos concluir que el modelo supervisado alcanzó un desempeño sólido en la clasificación de notas clínicas, mientras que los métodos no supervisados permitieron identificar clústeres y tópicos semánticos relevantes, revelando patrones temáticos emergentes que no se reflejaban directamente en las etiquetas originales. Lo que nos permitió identificar que los enfoques híbridos pueden complementar las técnicas tradicionales, ofreciendo una visión más completa del contenido clínico y apoyando tareas como la predicción diagnóstica y la segmentación temática.

El principal aporte metodológico de este trabajo es que un pipeline híbrido de PLN puede estructurar semánticamente textos clínicos de manera correcta. La integración de métodos supervisados y no supervisados permite un balance entre precisión en clasificación y descubrimiento de conocimiento, lo que constituye una ventaja frente a enfoques que utilizan únicamente un tipo de técnica.

Finalmente, se propone como trabajos futuros: incorporar modelos de lenguaje preentrenados en el dominio biomédico (por ejemplo, *BioBERT*), integrar ontologías médicas como *SNOMED CT* o *UMLS* para enriquecer la representación semántica, y evaluar el pipeline en tareas clínicas concretas, tales como detección de condiciones subdiagnosticadas, extracción de relaciones médico-paciente-fármaco y generación automática de resúmenes clínicos. Estos avances permitirían consolidar el pipeline como una herramienta útil en aplicaciones de salud basadas en texto.

7. LITERATURA CITADA

- Aulla C. J. C., Ortiz C. G. A., Orozco O. E. L., Cisneros M. L. F., Velasco V. S. A., y Cárdenas C. M. C. (2024). *El impacto de las enfermedades autoinmunes en la calidad de vida: un análisis de los síntomas, tratamientos y estrategias de afrontamiento*. Revista Sanitaria de Investigación, 5(8), 492.
- Cardoso, F. E., Ferneda, E., y Botega, L. C. (2023). *Clasificación de textos: un enfoque con uso de machine learning*. Revista EDICIC, 3(3), 1–17.
- Gan, M., Peng, Y., Ying, Y., Zhang, K., y Chen, Y. (2025). *Machine learning differentiation of rheumatoid arthritis–Sjögren's syndrome overlap from Sjögren's syndrome with polyarthritis*. Frontiers in Immunology, 16, 1614631.
- Jadue A. y González I. (2012). *Inmunopatogenia de las enfermedades autoinmunes*. Revista Médica Clínica Las Condes, 23(4), 464–472.
- Kise, Y., Fukuda, M., Shibata, T., Funakoshi, T., Ariji, Y., y Ariji, E. (2025). *Preliminary approach to creation of a prediction model for diagnosis of Sjögren's syndrome using radiomics and machine learning techniques on computed tomography images of the parotid glands*. Imaging Science in Dentistry, 55(2), 189–197.
- Lezama-Sánchez, A. L., Tovar Vidal, M., y Reyes-Ortiz José A. (2025). *Integration of Text Classification with Semantic Relationships Embedding Models in topic discovery aimed at opinion mining*. En *Intelligent Sustainable Systems: Selected Papers of WorldS4 2024*, (Vol. 1, pp. 1179, 325–334) Springer
- Luque Guzmán, M. (2020). *Text Mining y Medicina: Una aproximación a la detección temprana de enfermedades* (Tesis de maestría). Universidad de Sevilla.
- Módis V., Matuz A., Aradi Z., Horváth I. F., Szántó A., y Bugán, A. (2025). *Unveiling psychobiological correlates in primary Sjögren's syndrome: A machine learning approach to determinants of disease burden*. Frontiers in Psychiatry, 16, 1549756.
- Steyvers, M., y Griffiths, T. (2007). *Probabilistic topic models*. En T. Landauer, D. McNamara, S. Dennis, y W. Kintsch (Eds.), *Latent Semantic Analysis: A Road to Meaning* (pp. 427–448). Lawrence Erlbaum Associates.
- Singh, T. D., Divyansha, D., Singh, A. V., y Khilji, A. F. U. R. (2020, July). *A Hybrid Classification Approach using Topic Modeling and Graph Convolution Networks*. In *2020 International Conference on Computational Performance Evaluation (ComPE)* (285–289). IEEE.
- Srockel D. J. J., Fontalvo M. T., y Araque C. (2014). *Ánálisis de un modelo de clases latentes en las escalas de clasificación diagnóstica del síndrome coronario agudo*. Universitas Médica, 55(2), 138–151.
- Ting C., Nan W., Lili S., Jing L., y Zhenyu L. (2025). *Application of metabolomics and MCDM approach in developing a novel strategy for disease diagnosis: A case study in primary Sjögren's syndrome*. Journal of Pharmaceutical and Biomedical Analysis, 117080.
- Xue C., Minhyun L., Mohammad N. U., Xuange Z., y Vincent G. Z. (2025). *Analyzing different household energy use patterns using clustering and machine learning*. Renewable and Sustainable Energy Reviews, 212, 115335.
- Wang, X., Li, Y., Zhang, J., y Zhao, H. (2023). *Text classification via graph convolutional networks and topic modeling*. Expert Systems with Applications, 222, 119847



Tecnológico Nacional de México Instituto Tecnológico de Cd. Victoria

División de Estudios de Posgrado e Investigación

DOCTORADO EN CIENCIAS EN BIOLÓGÍA

SISTEMA NACIONAL DE POSGRADOS (SECIHTI - 003351)

Becas disponibles

Líneas de investigación:

- 1. Manejo y Conservación de Recursos Naturales**
- 2. Biodiversidad y Ecología**



Recepción de solicitudes: febrero-mayo de 2025

LÍNEAS DE INVESTIGACIÓN

- Biodiversidad y Ecología
- Manejo y Conservación de Recursos Naturales

Requisitos y antecedentes académicos de ingreso de los candidatos

- Contar con grado de Maestría (indispensable estar titulado) en un programa experimental o de investigación en el área de las Ciencias Biológicas.
- Promedio igual o superior a 8 (80 de 100) en estudios de maestría.
- Disponer de tiempo completo para cumplir con el programa doctoral.
- Aprobar el examen de conocimientos que aplica el programa o acreditar con al menos un 75% en conocimientos básicos y un 60% en habilidades de investigación en el EXANI-III del CENEVAL.
- Acreditar el examen de Inglés TOEFL, al ingresar al programa, mínimo 450 puntos. O bien acreditarlo este examen antes de egresar del programa, ya que este es un requisito para sustentar examen de grado y poder titularse.
- Presentar dos cartas académicas de recomendación expedidas por profesionistas reconocidos.
- Carta de exposición de motivos para el ingreso al doctorado, no mayor de una cuartilla, con fecha y firma.
- Visto bueno en entrevista con miembros del Claustro Doctoral.
- Presentar por escrito protocolo de investigación (3-5 cuartillas) para evaluar aptitudes y habilidades de experiencia previa, en el área de ciencias naturales.
- Carta de aceptación de uno de los miembros del Claustro Doctoral.

PLANTA DOCENTE

Almaguer Sierra Pedro, Dr. UANL. Aguas, Suelos, Agrometeorología e Hidroponia.

Azuara Domínguez Ausencio. Dr. Colegio de Posgraduados. Manejo Integrado de Plagas.

Barrientos Lozano Ludivina, Ph.D. Universidad de Gales. Entomología Aplicada. Ecología y Sistemática de Orthoptera.

Flores Gracia Juan, Dr. UANL. Genética y Biotecnología.

García Jiménez Jesús. Dr. UANL. Ciencias Forestales y Micología.

González Gaona Othón Javier. Dr. ITESM. Toxicología.

Guevara Guerrero Gonzalo, Dr. UANL. Biotecnología y Micología.

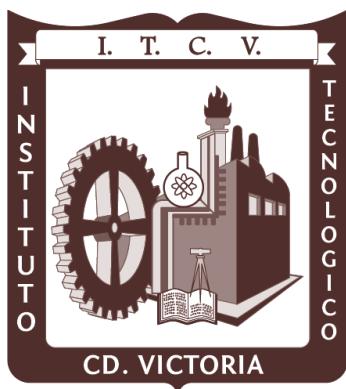
Horta Vega Jorge V., Dr. CINVESTAV-IPN Neurociencias y Entomología.

Rangel Lucio José Antonio. Dr. Colegio de Posgraduados. Edafología.

Rodríguez-Castro Jorge Homero, Dr. UAT. Acuacultura y Ecología Marina.

Venegas Barrera Crystian Sadiel. Dr. CIBNOR. Manejo y Preservación de Recursos Naturales (Ecología).

INFORMES



**TECNOLÓGICO NACIONAL DE MÉXICO
INSTITUTO TECNOLÓGICO DE CD.
VICTORIA.**

**División de Estudios de Posgrado e
Investigación.**

**Avenida Tecnológico N° 1301.
Cd. Victoria, Tam. C.P. 87010.
Tels. (834) 153 – 2000. Ext. 362**

Sitio web:

<http://www.itvictoria.edu.mx>

<https://www.itvictoria.edu.mx/oferta/doctora do.html>

Email:

Dra. Ludivina Barrientos Lozano
(doctorado@cdvictoria.tecnm.mx).

Coordinadora del Doctorado en Ciencias en Biología.

Dra. Adriana Mexicano Santoyo
(dposgrado@cdvictoria.tecnm.mx). Jefa de la División de Posgrado.



Tecnológico Nacional de México Instituto Tecnológico de Cd. Victoria

División de Estudios de Posgrado e Investigación

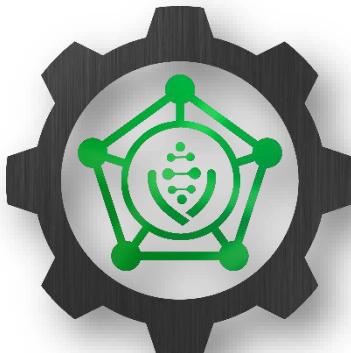
MAESTRÍA EN CIENCIAS EN BIOLOGÍA

SISTEMA NACIONAL DE POSGRADOS (SECIHTI - 002408)

Becas disponibles

Líneas de investigación:

- 1. Biodiversidad y Ecología**
- 2. Manejo de Agroecosistemas**



Recepción de solicitudes: febrero-mayo de 2026

PERFIL

El programa está diseñado para egresados de la carrera de biología o afines como médicos veterinarios, ingenieros agrónomos, ingenieros ambientales e ingenieros forestales. Podrán participar egresados de otras carreras con la aprobación del consejo de posgrado.

REQUISITOS DE INGRESO Y DOCUMENTACIÓN

- Carta de exposición de motivos indicando porque desea cursar una maestría y porque desea ingresar a este programa, Maestría en Ciencias en Biología-ITCV.
- Copia (s) de título profesional, certificado de calificaciones, diploma (s) y constancias de otros estudios.
- Constancia de promedio mínimo de 8 (ocho) en estudios de licenciatura.
- Currículum vitae con documentos probatorios adjuntos.
- Comprender el idioma inglés y aprobar examen de inglés del programa de MCB-ITCV.
- Dos fotografías tamaño credencial.
- Aprobar examen de admisión.
- Carta compromiso indicando que terminará su programa de maestría en dos años.
- Disposición para desarrollar e integrarse en proyectos de investigación.
- Entrevista con el comité de posgrado.
- Ser estudiante de tiempo completo.

PLAN DE ESTUDIOS

El programa está diseñado para concluirse en dos años y consta de cinco materias básicas, seis optativas y presentación de tesis de grado.

Áreas disponibles actualmente para investigación y desarrollo de tesis:

Malacología, Entomología, Micología,

Mastozoología, Ciencias Forestales (Biodiversidad, Sistemática, Ecología y Fisiología).

PLANTA DOCENTE

Almaguer Sierra Pedro, Dr. UANL. Agua-Suelos, Agrometeorología e Hidroponía.

Azuara Domínguez Ausencio, Dr. Colegio de Posgraduados. Manejo Integrado de Plagas.

Barrientos Lozano Ludivina, Ph.D. Universidad de Gales. Entomología Aplicada. Ecología y Sistemática de Orthoptera.

Flores Gracia Juan, Dr. UANL. Genética y Biotecnología.

García Jiménez Jesús, Dr. UANL. Micología y Parasitología Forestal.

González Gaona Othón Javier, Dr. ITESM. Toxicología.

Guevara Guerrero Gonzalo, Dr. UANL. Biotecnología y Micología.

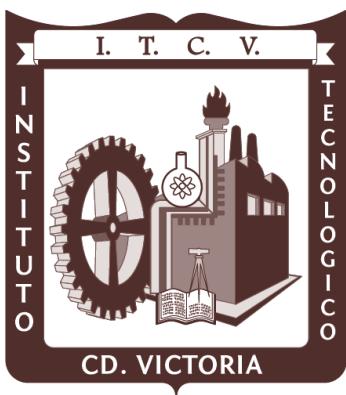
Horta Vega Jorge V., Dr. CINVESTAV-IPN Neurociencias y Entomología.

Rangel Lucio José Antonio, Dr. Colegio de Posgraduados. Edafología.

Rodríguez-Castro Jorge Homero, Dr. UAT. Acuacultura y Ecología Marina.

Venegas Barrera Crystian Sadiel, Dr. CIBNOR. Manejo y Preservación de Recursos Naturales (Ecología).

INFORMES



TECNOLÓGICO NACIONAL DE MÉXICO

**INSTITUTO TECNOLÓGICO DE CD.
VICTORIA.**

**División de Estudios de Posgrado e
Investigación.**

**Avenida Tecnológico N° 1301.
Cd. Victoria, Tam. C.P. 87010.
Tels. (834) 153 – 2000. Ext 326**

Sitio web:

<http://www.itvictoria.edu.mx>

<https://www.itvictoria.edu.mx/oferta/mbiologia.html>

Email:

Dra. Itzel Rubí Rodríguez de León
(coord_mcb@cdvictoria.tecm.mx o
posgrado.biologia@cdvictoria.tecm.mx).

Coordinadora de la Maestría en Biología.

Dra. Adriana Mexicano Santoyo
(dposgrado@cdvictoria.tecm.mx). Jefa de la
División de Posgrado.



Tecnológico Nacional de México Instituto Tecnológico de Cd. Victoria

División de Estudios de Posgrado e Investigación

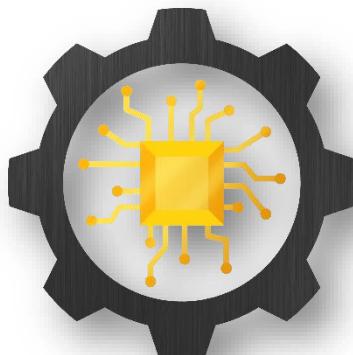
MAESTRÍA EN SISTEMAS COMPUTACIONALES

SISTEMA NACIONAL DE POSGRADOS (SECIHTI - 006805)

Becas disponibles

Líneas de investigación:

- 1. Ingeniería en Software**
- 2. Inteligencia Artificial**



Recepción de solicitudes: febrero-mayo de 2026

PERFIL

El programa está diseñado para egresados de carreras como Sistemas Computacionales, Informática, Tecnologías de la Información y Comunicaciones, así como áreas relacionadas con la ingeniería y la computación. Podrán participar egresados de otras disciplinas con la aprobación del consejo de posgrado, siempre que acrediten experiencia o formación vinculada al desarrollo de software, inteligencia artificial o tecnologías de la información.

REQUISITOS DE INGRESO Y DOCUMENTACIÓN

- Carta de exposición de motivos indicando porque desea cursar una maestría y porque desea ingresar a este programa, Maestría en Sistemas Computacionales.
- Copia (s) de título profesional, certificado de calificaciones, diploma (s) y constancias de otros estudios.
- Constancia de promedio mínimo de 8 (ocho) en estudios de licenciatura.
- Currículum vitae con documentos probatorios adjuntos.
- Comprender el idioma inglés y aprobar examen de inglés del programa de MSC-ITCV.
- Dos fotografías tamaño credencial.
- Aprobar examen de admisión.
- Carta compromiso indicando que terminará su programa de maestría en dos años.
- Disposición para desarrollar e integrarse en proyectos de investigación.
- Entrevista con el comité de posgrado.
- Ser estudiante de tiempo completo.

PLAN DE ESTUDIOS

Las líneas de trabajo seleccionadas para la Maestría Profesionalizante en Sistemas Computacionales, son: a) Ingeniería de Software, contempla el desarrollo de software

como un producto final tangible y que tendrá soporte en diferentes áreas de la computación como base de datos, sistemas distribuidos, lenguajes de programación, sistemas operativos, programación web, etc., que permitirán enriquecer y dar soluciones innovadoras y eficientes a problemas de manipulación de información en las diferentes empresas; b) Inteligencia Artificial, profundizando en la investigación y aplicación de soluciones en temas de biometría, control de acceso, monitoreo estratégico y otros relacionados a la toma de decisiones inteligentes.

PLANTA DOCENTE

Dra. Adriana Mexicano Santoyo, SNII Nivel 1, Doctorado en Ciencias de la Computación (CENIDET). Optimización combinatoria / Inteligencia Artificial / Ingeniería de Software

Dr. Jesús Carlos Carmona Frausto, SNII Nivel C, Doctorado en Ciencias en Computación (CINVESTAV-IPN). Metaheurísticas / Reconocimiento de Patrones / Optimización / Complejidad Computacional.

Dra. Lilia del Carmen García Mundo, Perfil PRODEP, Doctorado en Tecnologías Informáticas (UCLM). Ingeniería en Software / Modelos de Calidad para Evaluación de Software / Metodologías para el Desarrollo de Software.

Dr. Juan Antonio Vargas Enríquez, Doctorado en Tecnologías Informáticas (UCLM). Ingeniería de Software / Inteligencia Artificial

Dr. Mario Alberto Gómez Rodríguez, SNII Nivel Candidato, Doctorado en Ciencias en Computación (CINVESTAV-IPN). Inteligencia Artificial / Ingeniería de Software / Optimización.

M.S.C. Francisco Argüelles Granados, Maestría en Sistemas Computacionales

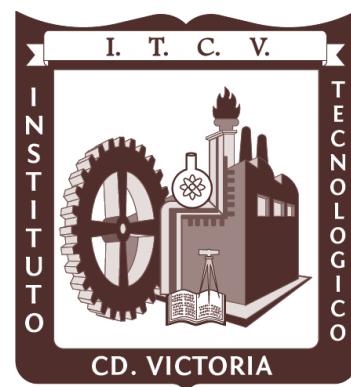
(TecNM – Campus Ciudad Victoria).
Ingeniería de Software

Dr. Salvador Cervantes Álvarez, SNII Nivel 1,
Doctorado en Ciencias de la Computación
(CENIDET). Inteligencia Artificial /
Ingeniería de Software.

Dr. Pascual Noradino Montes Dorantes, SNI
Nivel Candidato, Doctorado en Planeación
y Liderazgo Educativo (UAN - Campus
Saltillo). Optimización combinatoria /
Inteligencia Artificial / Logística

Dra. Nelva Nely Almanza Ortega, SNII Nivel 1,
Doctorado en Ciencias de la Computación
(CENIDET). Optimización combinatoria /
Inteligencia Artificial / Ingeniería de
Software.

INFORMES



TECNOLÓGICO NACIONAL DE MÉXICO
INSTITUTO TECNOLÓGICO DE CD.
VICTORIA.

División de Estudios de Posgrado e
Investigación.

Avenida Tecnológico N° 1301.
Cd. Victoria, Tam. C.P. 87010.
Tels. (834) 153 – 2000. Ext 327

Sitio web:

<http://www.itvictoria.edu.mx>
<https://www.itvictoria.edu.mx/oferta/msistema.html>

Email:

Dra. Lilia del Carmen García Mundo
(c.msc@cdvictoria.tecnm.mx). Coordinadora de
la Maestría Sistemas Computacionales.
Dra. Adriana Mexicano Santoyo
(dposgrado@cdvictoria.tecnm.mx). Jefa de la
División de Posgrado.



Tecnológico Nacional de México Instituto Tecnológico de Cd. Victoria

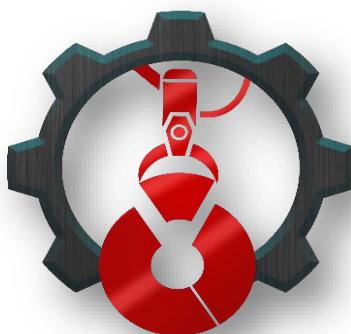
División de Estudios de Posgrado e Investigación

MAESTRÍA EN INGENIERIA INDUSTRIAL

SISTEMA NACIONAL DE POSGRADOS (SECIHTI - 005821)

Líneas de investigación:

- 1. Mejoramiento de los Sistemas de Calidad y
Productividad**
- 2. Optimización**



Recepción de solicitudes: febrero-mayo de 2026

PERFIL

Desarrollar el proceso de selección de los aspirantes a la generación 2023-2025 de la Maestría en Ingeniería Industrial con orientación profesional, la cual tiene como propósito formar posgraduados en Ingeniería Industrial que, de acuerdo a la problemática local, regional, nacional y global y a las tendencias de investigación de frontera en la disciplina, tengan capacidades de aplicar conocimientos para el desarrollo y la innovación, con sentido analítico y sistemático de manera que diseñen, planten, operen, controlen y mejoren sistemas de producción de bienes y servicios mediante el uso adecuado de los recursos; además, el egresado de ese programa será capaz de conocer y operar herramientas y recursos computacionales necesarios para desarrollar sus actividades y actuará como un agente de cambio en su disciplina, ejerciendo liderazgo y ética profesional en el desarrollo de sus actividades y estará consciente y comprometido con el cuidado y la preservación del medio ambiente.

REQUISITOS DE INGRESO Y DOCUMENTACIÓN

- Carta de exposición de motivos indicando porque desea cursar una maestría y porque desea ingresar a este programa, Maestría en Ingeniería Industrial-ITCV.
- Copia (s) de título profesional, certificado de calificaciones, diploma (s) y constancias de otros estudios.
- Constancia de promedio mínimo de 8 (ocho) en estudios de licenciatura.
- Currículum vitae con documentos probatorios adjuntos.
- Comprender el idioma inglés y aprobar examen de inglés del programa de MII-ITCV.
- Dos fotografías tamaño credencial.
- Aprobar examen de admisión.
- Carta compromiso indicando que

terminará su programa de maestría en dos años.

- Disposición para desarrollar e integrarse en proyectos de investigación.
- Entrevista con el comité de posgrado.
- Ser estudiante de tiempo completo.

OBJETIVOS PARTICULARES

- ❖ Formar recursos de alto nivel académico en el área de Ingeniería industrial, preparados para apoyar el desarrollo socioeconómico de nuestro país, capaces de aplicar conocimientos mediante el desarrollo de proyectos de aplicación científica y tecnológica y apta para mejorar las actividades productivas y académicas.
- ❖ Realizar investigación aplicada en el ámbito de la Ingeniería Industrial que permita el desarrollo de la enseñanza tecnológica y el aprovechamiento social de los recursos naturales materiales para traducir estos en nuevas fuentes de riqueza.
- ❖ Colaborar con los sectores públicos y privados en la consolidación del desarrollo tecnológico y social de la comunidad.
- ❖ Contribuir a la difusión científica y a la preservación de la cultura regional y nacional.

PLANTA DOCENTE

Dra. Adriana Mexicano Santoyo, SNII Nivel 1, Doctorado en Ciencias de la Computación (CENIDET). Optimización combinatoria / Inteligencia Artificial / Ingeniería de Software

Dr. Jesús Carlos Carmona Frausto, SNII Nivel C, Doctorado en Ciencias en Computación (CINVESTAV-IPN). Metaheurísticas / Reconocimiento de Patrones / Optimización / Complejidad Computacional.

Dr. Edgar Pérez Arriaga, Doctor en Medio Ambiente (Facultad de Ingeniería "Arturo Narro Siller"). Desarrollo Sustentable y Ambiental / Optimización / Sistemas de Calidad

Dr. Luis Antonio Vázquez Ochoa, SNII Nivel Candidato, Doctor en Ciencias en Biología (TecNM – Campus Ciudad Victoria). Gestión de Recursos Naturales / Cambio climático

Dra. Araceli Maldonado Reyes, SNII Nivel 1, Doctora en Ciencias en Ingeniería Metalúrgica y Cerámica (CINVESTAV-IPN Unidad Saltillo). Procesos de Manufactura Tradicional y Avanzado

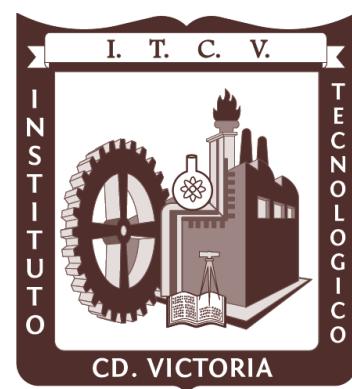
Dr. Ricardo Daniel López García, SNII Nivel Candidato, Doctor en Ciencias en Ingeniería Metalúrgica y Cerámica (CINVESTAV-IPN). Procesos de Manufactura

Dra. Martha Olivia Lázaro Dzul, SNII Nivel 1, Doctora en Ciencias Forestales y Entomología y Acarología (Colegio de Posgrados). Taxonomía y Diversidad de Escolitinos (Coleóptera) / Manejo de Insectos de Importancia Agrícola.

Dr. Jesús Leccinum García Morales, SNII Nivel 1, Doctor en Ciencias con especialidad en Gestión de Recursos Naturales (UANL). Gestión de los Recursos Naturales / Optimización e Innovación

M. C. Flora Alicia González Jiménez: Maestra en Ciencias en Ingeniería Industrial (TecNM – Campus Ciudad Juárez). Optimización / Sistema de Calidad / Estadística Industrial

INFORMES



TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE CD.
VICTORIA

División de Estudios de Posgrado e
Investigación.

Avenida Tecnológico No 1301
Cd. Victoria, Tam. C.P. 87010
Tels. (834) 153-2000. Ext 328

Sitio web:

<http://www.itvictoria.edu.mx>

<https://www.itvictoria.edu.mx/oferta/mindustrial.html>

Email:

M.C. Flora Alicia González Jiménez
(mindustrial@cdvictoria.tecnm.mx).

Coordinadora de la Maestría Ingeniería Industrial.

Dra. Adriana Mexicano Santoyo
(dposgrado@cdvictoria.tecnm.mx). Jefa de la
División de Posgrado.

CONVOCATORIA PARA PUBLICAR EN TecnoINTELECTO

TÍTULO CON MAYÚSCULAS, DEBIDAMENTE ACENTUADAS, EN NEGRITAS, CENTRADO, ARIAL 10, INTERLINEADO SENCILLO

Autor(es) Arial 10 puntos, itálica, centrado, interlineado sencillo; nombre (s) completo y apellidos completos, separados por un guión, sin grado académico, más de un autor separado por comas e indicador numérico para los datos siguientes: Institución(es) en 10 Arial, en itálica y centrado, interlineado sencillo, correo electrónico de los autores centrado, interlineado sencillo

RESUMEN. Deberá ser lo más general y significativo posible, de manera que en pocas palabras exprese la aportación más relevante del artículo. Letra tipo Arial de 10 puntos, interlineado sencillo y espaciado anterior de 8 puntos y posterior de 6, iniciando con la palabra **RESUMEN** en negritas. Texto con alineación ajustada en todo el artículo. Si el artículo está en español, adjuntar el resumen inglés.

PALABRAS CLAVE: Colocar las palabras (tres a cinco) más significativas en el artículo, no repetir palabras del título, fuente de 10 puntos, dejando un espacio entre el párrafo anterior.

ABSTRACT. The abstract shall be as general and substantial as possible, in such a way that provides in a few words a clear idea of the paper's contribution. Please use Arial font 10 points, single space, space above 8 points and below 6 points, begin text with the word **ABSTRACT** in bold face. All text through the paper must be aligned to fit page. If paper is in Spanish abstract shall be in English.

KEY WORDS: Please use the most (three to five) significant words, font of 10 points, leaving a space between the preceding paragraphs.

1. INTRODUCCIÓN

Los criterios para la revisión técnica son: importancia de la contribución a la divulgación científica, pertinencia de métodos empleados, correcta presentación de datos, soporte del manuscrito con literatura relevante y actualizada, discusión suficiente o necesaria. Además, figuras y tablas adecuadas. El manuscrito pasará al comité editorial, quien dictaminará si contiene el mínimo indispensable para ser publicado, lo cual se notificará vía electrónica en formato pdf.

2. CARACTERÍSTICAS

El cuerpo del artículo en dos columnas con 0.6 cm entre ellas y todos sus márgenes de 3 cm. Cada sección deberá contener un título numerado con formato de párrafo espaciado anterior de 12 y posterior de 6 puntos. La fuente de todo el manuscrito es Arial. En el cuerpo de 10 puntos, interlineado sencillo, con secciones numeradas con números arábigos.

2.1 Idioma: español o inglés.

2.2 Subsecciones

Las subsecciones en formato tipo título, negritas, interlineado sencillo y espaciado anterior y posterior de 6 puntos.

2.3. Las gráficas y tablas

Pueden ser a color o en **escala de grises** y se ajustarán de acuerdo con las características de ellas y al gusto del investigador. Deberán ser posicionadas de acuerdo con la necesidad del investigador y bajo su responsabilidad.

3. LINEAMIENTOS

Los artículos deberán ser inéditos. Cada trabajo deberá presentarse en un mínimo de 6 y un máximo de 12 páginas. De 6 páginas se considerarán artículos cortos y se publicarán a recomendación del comité editorial.

4. RESPONSABILIDADES

El investigador es responsable del contenido, la sintaxis y el envío de su artículo en Word a la coordinación editorial actual de TecnoINTELECTO:

tecnointelecto@cdvictoria.tecnm.mx.

El Instituto Tecnológico de Cd. Victoria será responsable de la revisión y aceptación o rechazo de los manuscritos, la edición de la

revista, el índice, la impresión y distribución, apoyándose en el Comité Editorial y otras instituciones, si lo considera pertinente.

Los artículos que no se ajusten a las normas editoriales serán rechazados para su adecuación.

El máximo número de autores y/o coautores por artículo es de 5.

5. FECHAS IMPORTANTES

Recepción de artículos todo el año.

Publicación julio-agosto y diciembre-enero.

6. LITERATURA CITADA

6.1 Referencias en texto

Sin numerar, solo citar apellido(s) según el caso y el año separado por una coma, si son más citas separar por punto y coma; dos autores se separan "y" y si son más de dos autores solo se pondrá el apellido(s) del primer autor seguido de "et al.,".

Al final, listar en orden alfabético sin numeración. Autor (es) iniciando con apellido (s) seguido por la inicial del nombre (s), si es el caso puede escribir los dos apellidos separados por un guion. Año. Título del artículo. Nombre de la Revista, Volumen y número de páginas, tipo Arial, 10 puntos, interlineado sencillo.

Artículo científico

Armenta, C. S., H. Bravo y R. Reyes. 1978. Estudios bioecológicos de *Epilachna*

varivestis Mulsant, bajo condiciones de laboratorio y campo. Agrociencia, 34: 133-146.

Ávila-Valdez, J., L. Barrientos-Lozano y P. García-Salazar. 2006. Manejo Integrado de la Langosta centroamericana (*Schistocerca piceifrons piceifrons* Walker) (Orthoptera: Acrididae) en el sur de Tamaulipas. Entomología Mexicana, 5: 636-641.

Libro o Tesis

Jaffe, K., J. Lattke y E. Pérez. 1993. *El mundo de las hormigas*. Equinoccio Ediciones. Universidad Simón Bolívar, Venezuela. 196 pp. En el caso de tesis señalar después del título si es profesional o de grado.

Capítulo de libro:

Navarrete-Heredia, J. L. y A. F. Newton. 1996. Staphylinidae (Coleoptera). Pp. 369-380. In: J. E. Llorente-Bousquets, A. N. García-Aldrete y E. González-Soriano (Eds.). Biodiversidad, Taxonomía y Biogeografía de Artrópodos de México: Hacia una Síntesis de su Conocimiento. Instituto de Biología, UNAM, México, D. F.

Tecnológico Nacional De México

Instituto Tecnológico de Cd. Victoria

División de Estudios de Posgrado e Investigación
Coordinación Editorial de TecnoINTELECTO

Dra. Adriana Mexicano Santoyo
tecnoinlecto@cdvictoria.tecnm.mx