

ARTÍCULO ACEPTADO

La inteligencia artificial como apoyo en la identificación de especies de plantas

Juan Augusto Campos-Leal, Inés F. Vega-López, Gerardo Beltrán-Gutiérrez, José Ramón López-Arellano y Arturo Yee-Rendón

Introducción

En los últimos años, la tecnología ha tomado un rol importante para la realización de las actividades de un individuo en su día a día. Usamos dispositivos que nos permiten llevar a cabo tareas simples y complejas o bien, tareas repetitivas y específicas. En busca de resolver tareas de forma automática y sencilla, hay una constante búsqueda por generar las herramientas que nos permitan realizarlo. La tecnología ha permitido que diversas áreas de la ciencia utilicen sus conocimientos para obtener estas herramientas inteligentes. La inteligencia artificial (IA) es el área de las ciencias de la computación que permite generarlas a través del procesamiento de datos.

Dichas herramientas han sido capaces de resolver tareas complejas para el ser humano, de forma eficiente y sencilla.

Una de las tareas en la que la inteligencia artificial ha apoyado es la identificación de especies de plantas. Este tipo de tarea es llevada a cabo por expertos bajo un proceso meticuloso y muy bien cuidado. En este proceso se identifican las características de un ejemplar para determinar a que grupo pertenece, estos grupos los determinan los expertos mediante características similares y puede ser a diferentes niveles, como familia, género o especie [1].

La identificación de especies de plantas a través de imágenes digitales no es un problema trivial para la IA, representa un reto extraordinario debido a las similitudes inter-especies y grandes variaciones intra-especie.

La idea detrás de este trabajo es mostrar que el uso de herramientas tecnológicas basadas en la IA, es una estrategia viable para auxiliar a expertos en el estudio y cuantificación de la biodiversidad de una región. El trabajo de registro y recolección de especies de plantas que se ha hecho durante años puede ser aprovechado por estas herramientas tecnológicas para entrenar modelos de identificación automática con altas tasas de aciertos.

Antecedentes

Durante muchos años, la tarea de identificación de especies de plantas se llevó a cabo de manera manual por taxónomos expertos, bajo proceso de laboratorio y observación. Este es un proceso totalmente biológico que no tenía nada que ver con la computación o con la inteligencia artificial. Sin embargo, Younis y sus colaboradores [2] mencionan que desde hace aproximadamente dos décadas fue propuesta la idea de generar herramientas inteligentes para la identificación de especies de plantas.

Las primeras herramientas inteligentes para este tipo de tarea se basaban en un proceso de aprendizaje automático, estas herramientas partían del aprendizaje supervisado, es decir, el experto había identificado previamente el ejemplar a caracterizar. Por lo tanto, conocía el dato de entrada y la respuesta que la herramienta de-

bería de proporcionar. La información de entrada eran las características del ejemplar y estas características se tomaban directamente del ejemplar o de imágenes digitales. En el caso de la extracción de características de las imágenes digitales se usaban algoritmos diseñados por un humano experto para el proceso de extracción. Sin embargo, este proceso de extracción estaba limitado por la capacidad humana. Estas técnicas son conocidas como técnicas convencionales de aprendizaje de máquina.

Nilsback y sus colaboradores [3] presentaron trabajos para la identificación de especies de plantas que se basaban en técnicas convencionales de aprendizaje de máquina. Además, algunos conjuntos de datos de especies de plantas fueron propuestos para auxiliar a la creación de este tipo de herramientas inteligentes. Algunos contenían imágenes de hojas como Swedish Leaf y Flavia, con imágenes de 15 y 32 especies, respectivamente. También algunos otros conjuntos de imágenes de flores como Oxford 17 Flower y Oxford 102 Flower. El rendimiento de cada herramienta propuesta era probado en alguno de estos conjuntos para tener un punto de referencia en cuanto a los resultados.

Hasta el año 2012 los resultados no eran muy alentadores y la aplicación de estas herramientas en el mundo real aún no era viable. Sin embargo, con la aparición de

las redes neuronales convolucionales [4] (CNNs, por sus siglas en inglés), un tipo de técnica basada en el aprendizaje profundo, las tareas de identificación de especies de plantas usando imágenes digitales se volvieron viables.

La principal diferencia entre el aprendizaje profundo y los métodos convencionales de aprendizaje de máquina es que el primero no tiene un proceso de extracción de características limitado por la capacidad humana.

En el aprendizaje profundo el proceso de extracción de características en una imagen digital es automático.

El proceso va mejorando en la extracción de características mediante el algoritmo de aprendizaje. Por lo tanto, una CNN aprende aquellas características que más le sirven para la identificación de especies de plantas. Sin embargo, para que el proceso de extracción de características pueda ser lo suficientemente bueno es necesario que durante el proceso de aprendizaje el modelo procese un gran número de imágenes. Cuando el número de imágenes en un conjunto de datos es limitado, se han aplicado técnicas de aprendizaje por transferencia para tratar de mejorar la tasa de aciertos de los modelos.

El año 2012 marcó un antes y un después con respecto a la identificación de imágenes. Fue la primera vez que las CNNs superaron a las técnicas convencionales de aprendizaje de máquina [4]. A partir de ese año han sido grandes los esfuerzos para la generación de herramientas inteligentes. En el ámbito de la identificación de especies de plantas ha quedado demostrado en trabajos como el de Sulc y colaboradores [5], donde a través de su propuesta basada en CNNs obtuvo el primer lugar en la edición 2018 del reto PlantCLEF, uno de los retos mas importantes a nivel mundial en este tipo de tarea.

En 2020, Little y sus colaboradores [6] presentaron un reto para fomentar el desarrollo de algoritmos para la identificación de especies de plantas de herbario. La competencia atrajo a 22 equipos a nivel mundial que ingresaron 254 modelos CNNs entrenados para la identificación de ejemplares de la familia *Melastomataceae*. Los cuatro mejores con una tasa de aciertos mayor a 88 %. En 2022, Campos-Leal y sus colaboradores [7] presentaron una propuesta basada en CNN para la identificación de 120 especies de plantas con flores nativas de México. En su trabajo presentan un modelo CNN simplificado que obtuvo un 91.92 % en tasa de aciertos y fue comparado con otros modelos basados en CNN. En 2023, Vega-López y colaboradores [8] realizaron un trabajo para la identificación de 202 especies de plantas de la flora mexicana usando cuatro órganos distintivos para la identificación, tallo, hoja, fruto y flor. Los autores usaron estrategias de aprendizaje por transferencia con lo que obtuvieron 86.97 % en tasa de aciertos.

Redes neuronales convolucionales

Las CNNs son un tipo de técnica del aprendizaje profundo capaces de extraer por sí mismas las característi-

cas visuales que ayudan a discriminar las clases que se le presentan. Esto lo realizan mediante el procesamiento de imágenes a través de filtros digitales, cuyos coeficientes se infieren durante un proceso de entrenamiento. Este tipo de redes se conforman por tres tipos de capas, convolución, agrupamiento y clasificación. Una arquitectura de CNNs es formada por un apilamiento de este tipo de capas. Generalmente, los primeros dos tipos de capas forman un bloque que es utilizado para la extracción de características. La salida de este bloque es un vector de entrada a un clasificador. Este clasificador es compuesto por capas densas. Las arquitecturas reportadas en la literatura científica, tales como VGG-16, Inceptionv3 y ResNet50, hacen uso de diferentes configuraciones de estos bloques. La Figura 1 muestra la estructura de la arquitectura ResNet50, que está formada por 50 capas agrupadas en bloques que se encargan de la extracción automática de características y clasificación. Esta arquitectura tiene 25.6 millones de parámetros.

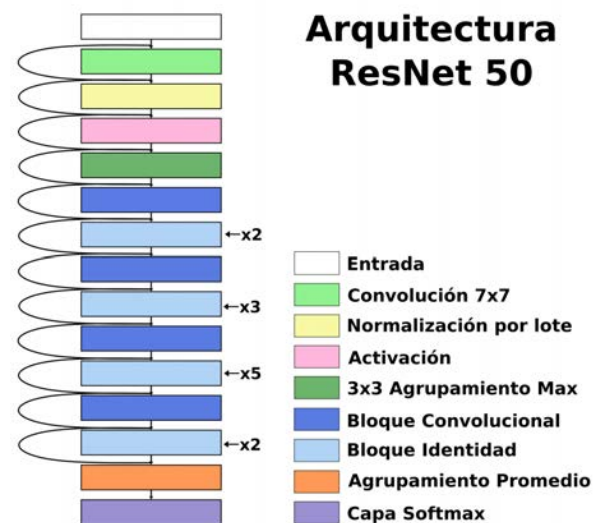


Figura 1. Diagrama de la arquitectura ResNet 50.

Las CNNs deben su nombre a su principal operación, la convolución. Una convolución es una operación matemática realizada por un filtro de tamaño $n \times n$ sobre

una región de una matriz de entrada. Cada elemento del filtro es multiplicado por el elemento correspondiente de la matriz. El filtro es desplazado por cada espacio disponible en la matriz, de izquierda a derecha y de arriba a abajo, a este proceso se le llama convolución. El número de celdas que avanza es determinado por un valor paramétrico llamado paso. Cuando el filtro se ha desplazado por toda la matriz de entrada se obtiene una matriz de activación, que será la entrada a procesar por una siguiente capa.

Las capas de agrupación tienen como propósito reducir el tamaño de las dimensiones *ancho* \times *alto* de las matrices de activación. Similar a las capas convolucionales, estas capas desplazan un filtro de tamaño $n \times n$ (generalmente de 2×2) por una matriz de entrada para aplicar la agrupación por regiones. Los tipos de agrupación son: por valor mínimo, máximo y promedio, tomando el valor más representativo según sea el caso. Cuando el filtro se ha desplazado por toda la matriz de entrada es obtenida una matriz que será la entrada a procesar por la siguiente capa. Las capas densas son las capas completamente conectadas tradicionales de las redes neuronales artificiales. Estas capas se encuentran normalmente al final de las CNNs y se utilizan para realizar el proceso de clasificación. Las capas densas normalmente reciben como entrada el vector de características obtenido durante el proceso de extracción.

Entrenamiento

El entrenamiento de una arquitectura de CNNs es un proceso donde el conjunto de datos es tratado iterativa-

mente para intentar descubrir características que permitan realizar la identificación. Durante este proceso se usa un algoritmo de aprendizaje llamado descenso de gradiente estocástico (SGD, por sus siglas en inglés). De esta manera, el problema de identificación es tratado como un problema de optimización que busca minimizar el error. Esto se hace mediante el ajuste de valores de los parámetros de la arquitectura. Durante el proceso de aprendizaje, la arquitectura procesa ejemplos de las categorías a identificar y ajusta los valores de los parámetros tratando de minimizar el error. Esto se hace mediante el algoritmo de retropropagación.

En las CNNs los valores de los parámetros se dividen en dos partes, filtros convolucionales y clasificador. Los primeros son los encargados de encontrar los patrones en las imágenes (características) y los segundos son los encargados de la clasificación. Para medir el comportamiento de la arquitectura durante el proceso de aprendizaje se mide la tasa de aciertos con respecto a las respuestas esperadas. Si la respuesta es la esperada contabilizamos un acierto, si no, entonces se utiliza una función de pérdida para medir que tan cerca o que tan lejos está la respuesta con respecto a la respuesta esperada. Para llevar a cabo el proceso de entrenamiento de las arquitecturas, el usuario proporciona un conjunto de valores que determinan el comportamiento de los algoritmos. Por ejemplo, controlan que tanto se modifica el valor de los parámetros, cuantas imágenes se procesan y cuantas iteraciones (o épocas) se observara por completo el conjunto de entrenamiento. El conjunto de estos valores son conocidos como hiperparámetros.

Durante el proceso de entrenamiento el conjunto de datos es utilizado y es dividido en tres subconjuntos, entrenamiento, validación y prueba.

El primer subconjunto contiene las imágenes que se procesarán y de las que el modelo aprenderá características, este conjunto debe ser el de mayor tamaño. El de validación es utilizado para medir el rendimiento del modelo cada vez que finaliza una iteración (o época), esto nos da una idea del comportamiento del modelo. El subconjunto de prueba es utilizado al final del entrenamiento, este lo utilizamos para obtener el rendimiento del modelo con un subconjunto de imágenes que no son vistas ni durante el entrenamiento ni la validación.

Aprendizaje por transferencia

El aprendizaje por transferencia (AT) es una técnica utilizada durante el entrenamiento para mejorar la tasa de aciertos de un modelo de clasificación. Esta técnica consiste en utilizar como punto de partida los valores de los parámetros de un modelo entrenando en un conjunto de datos a gran escala (generalmente con cientos de miles de

imágenes). Esta técnica se basa en el supuesto de que un modelo entrenado en una tarea A puede ser entrenado de nuevo para tratar una tarea B, ajustando solo un poco los valores de los parámetros durante el entrenamiento. Esto es debido a que el modelo A ya tiene la capacidad para extraer características y los valores de los parámetros no necesitan ser modificados por completo, solo necesitan ajustarse a las nuevas características. El punto de partida más utilizado en el aprendizaje profundo son los valores de los parámetros de modelos entrenados en el conjunto de datos de ImageNet 2012, esto es debido a que este conjunto es utilizado como punto de referencia para medir la tasa de aciertos de los modelos de clasificación.

Conjunto de datos

En este documento, para ilustrar como la IA puede ser aplicada a la identificación de especies de plantas utili-

zamos dos conjuntos de datos como experimento ilustrativo. Para este trabajo elegimos Oxford 102 Flower [3] como un conjunto de datos de especies de plantas en su hábitat natural y un subconjunto de PlantCLEF2020 [9] como un conjunto de especies de plantas en un ambiente controlado. El conjunto Oxford 102 Flower, es un conjunto que contiene 8,189 imágenes de flores que corresponden a 102 especies de plantas que pueden ser encontradas en Reino Unido. En la Figura 2 mostramos algunas imágenes de ejemplo en el conjunto. Estas imágenes fueron tomadas en condiciones de campo por lo cual pueden existir variaciones en la luz, fondo, perspectiva y escala, por mencionar algunas. Este conjunto contiene entre 40 y 258 imágenes por especie.

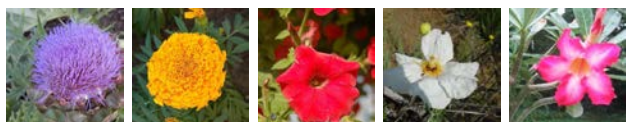


Figura 2. Imágenes del conjunto de Oxford 102 Flower.

El otro conjunto de datos utilizado es un subconjunto de PlantCLEF 2020, el cual consiste en 10 especies de plantas de herbario con mayor representación. Para cada especie se tomó una muestra de 50 imágenes. Cada imagen tiene dimensión rectangular con orientación vertical. La Figura 3 muestra imágenes de ejemplo para cada especie en el conjunto.



Figura 3. Imágenes del conjunto de PlantCLEF 2020.

Las CNNs reciben como entrada una imagen digital cuadrada de un tamaño específico. Cuando la imagen no cumple con estas especificaciones, sufre de algunas alteraciones que pueden llevar a distorsionar el objeto a identificar que está contenido en la imagen, o hay un redimensionamiento modificando la información de la imagen. Estas alteraciones pueden tener un impacto negativo en la tasa de aciertos del modelo, debido a que puede cambiar la naturaleza de los objetos a identificar. El preprocesamiento de las imágenes puede evitar las alteraciones en una imagen. De esta manera podemos enfocar la

atención del modelo y evitar distorsiones o alteraciones. El preprocesamiento consistió en la realización de dos recortes cuadrados por cada imagen. La Figura 4 muestra un ejemplo de este proceso. Cada recorte corresponde a una región de interés en la imagen. Para generar el primer recorte se tomó como referencia de inicio la esquina superior izquierda de la imagen sin procesar. Para el segundo recorte se tomó como referencia el centro de la imagen. De esta manera se obtuvieron dos regiones de interés cuadradas por cada imagen. En total formamos un conjunto de 1,000 imágenes con 100 imágenes para cada especie.

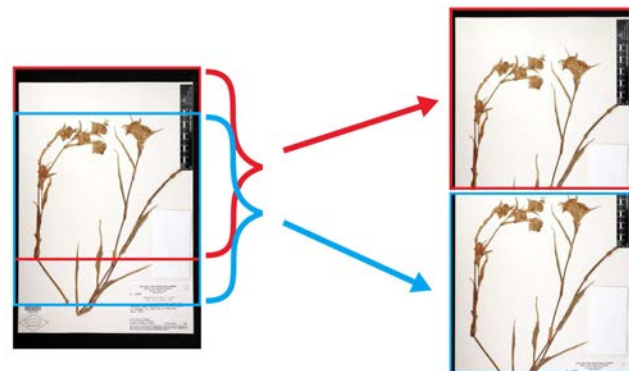


Figura 4. Imagen sin procesar y sus dos recortes.

Resultados

Utilizamos la arquitectura ResNet50 [10] para llevar a cabo nuestros experimentos. Entrenamos modelos de clasificación de la siguiente manera: modelos cuyos valores iniciales de sus parámetros fueron asignados aleatoriamente, conocido como entrenamiento desde cero, y modelos que fueron generados utilizando aprendizaje por transferencia del reto ImageNet 2012, es decir, modelos cuyos valores iniciales de parámetros fueron valores obtenidos por modelos previamente entrenados. Además, utilizamos una técnica de validación cruzada de 10 partes para tener una mayor certeza estadística en los resultados. Los valores de los hiperparámetros que utilizamos están descritos en la Tabla 1.

Tabla 1. Valores de los hiperparámetros.

Parámetros	Desde cero	Con AT
Optimizador	SGD	SGD
Tasa de aprendizaje	0.01	0.0001
Decaimiento	0.005	0
Impulso	0.9	0.9
Tamaño de lote	16	16
Número de épocas	200	40

En nuestros experimentos reportamos el desempeño de los modelos a través de la tasa de aciertos, conocido en inglés como *Top-K accuracy*. *Top-K* se refiere al porcentaje de respuestas correctas en el conjunto de las *K* respuestas mejor clasificadas por el modelo. Por ejemplo,

para un indicador *Top-1*, un acierto se contabiliza cuando la respuesta del modelo es igual a la respuesta esperada. Los resultados se muestran en la Tabla 2. Los renglones uno y dos representan a los modelos entrenados usando el conjunto de Oxford 102 Flower. El resto corresponde a los modelos usando el subconjunto de PlantCLEF 2020. De acuerdo a los resultados de la tabla, observamos que los modelos que utilizan estrategias de aprendizaje por transferencia obtiene mejores desempeños que aquellos modelos entrenados desde cero.

Tabla 2. Modelos ResNet50.

Modelo	Dataset	Top-1
ResNet50	Oxford 102 Flower	46.60 %
ResNet50 + AT	Oxford 102 Flower	94.87 %
ResNet50	PlantCLEF 2020 (10)	84.10 %
ResNet50 + AT	PlantCLEF 2020 (10)	93.50 %

En las Figuras 5 y 6, se presentan las curvas de aprendizaje de los modelos entrenados usando el conjunto de datos de Oxford 102 Flower. Las curvas de aprendizaje permiten conocer el comportamiento de los modelos durante el entrenamiento, para ello se utiliza el subconjunto de validación que nos permite conocer cómo está generalizando (aprendiendo) el modelo.

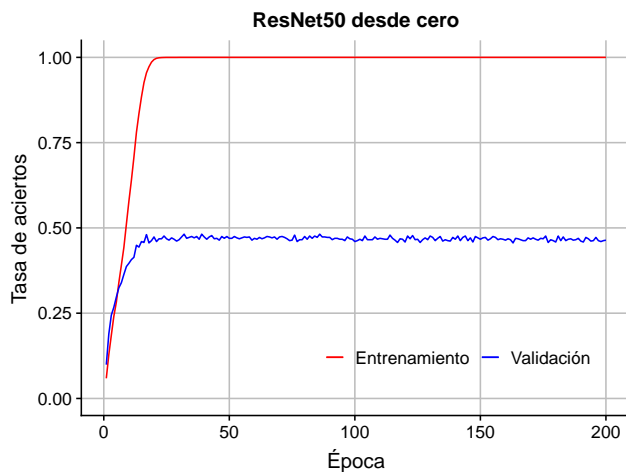


Figura 5. Modelo entrenado en Oxford 102 Flower.

La Figura 5 muestra la curva de aprendizaje de un modelo ResNet50 entrenado desde cero. El comportamiento presenta un amplio margen entre la tasa de aciertos del subconjunto de entrenamiento (línea de color rojo) y el subconjunto de validación (línea color azul). La diferencia entre las curvas es aproximadamente de 50 puntos porcentuales, lo que indica que el modelo obtiene buenos resultados durante el entrenamiento, pero no durante la validación, y por tanto, tampoco lo hará cuando se utilice con el conjunto de prueba. La Figura 6 muestra la curva de aprendizaje del modelo ResNet50 utilizando aprendizaje por transferencia. Observamos que el

margen entre las curvas de aprendizaje es mucho menor con respecto al margen del modelo entrenado desde cero. El análisis de las curvas de aprendizaje entre el modelo entrenado desde cero y el que usa aprendizaje por transferencia, nos permite validar y explicar los resultados de los modelos presentados en la Tabla 2.

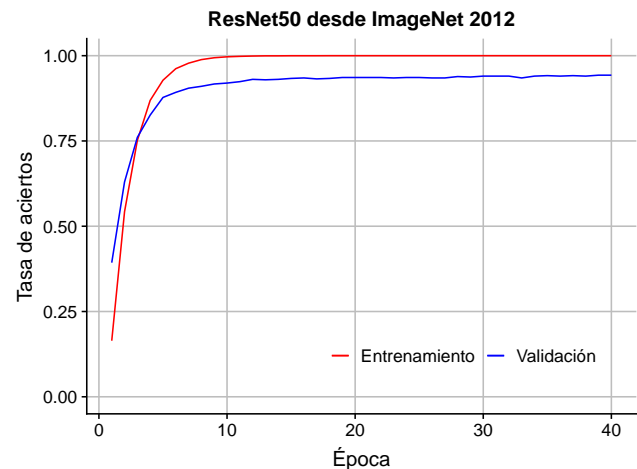


Figura 6. Modelo utilizando AT.

Conclusiones

La identificación de especies de plantas a través de herramientas tecnológicas basadas en IA ha mostrado buenos avances durante de la última década. Sin embargo, una considerable cantidad de datos de entrenamiento es requerida para obtener modelos de identificación automáticos con altas tasas de aciertos. En los últimos cinco años, los expertos en IA se han enfocado en utilizar las imágenes de especies de plantas de herbarios, pues la información recolectada a través de décadas está siendo digitalizada y es aprovechada para auxiliar a los expertos taxónomos en las tareas de identificación. En este trabajo llevamos a cabo experimentos ilustrativos para mostrar la utilidad de los modelos basados en IA utilizando dos conjuntos de datos públicos de imágenes de especies de plantas, imágenes de plantas en su hábitat natural e imágenes de plantas en herbario. Los resultados mostraron que los modelos de clasificación basados en IA son capaces de obtener altas tasas de aciertos en la identificación de especies de plantas mediante la utilización de técnicas de aprendizaje por transferencia. *

Agradecimientos

Los autores desean agradecer al Consejo Mexicano de Ciencia y Tecnología (CONACYT) por la beca otorgada al primer autor, así como por el financiamiento para el proyecto 291772 del fondo CONACYT-INEGI. De la

misma manera, los autores agradecen el apoyo económico brindado por la Universidad Autónoma de Sinaloa.

REFERENCIAS

1. Sulc, M. y J., Matas. (2014). Texture-Based Leaf Identification. En *Proc. Computer Vision-ECCV 2014 Workshops*, 185-200.
2. Younis, S., Weiland, C., et al. (2018). Taxon and Trait Recognition from Digitized Herbarium Specimens using Deep Convolutional Neural Networks. *Botany Letters*, 165(3-4), 377-383.
3. Nilsback, M. E. y Zisserman, A. (2008). Automated Flower Classification over a Large Number of Classes. En *Proc. 2008 Sixth Indian Conference on Computer Vision, Graphics and Image Processing*, 722-729.
4. Simonyan, K. y Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. En *Proc. of 3rd International Conference on Learning Representations*.
5. Sulc, M., Pícek, L. y Matas, J. (2018). Plant Recognition by Inception Networks with Test-time Class Prior Estimation. En *Proc. of the Conference and Labs of the Evaluation Forum*, 2125.
6. Little, D. P., Tulig, M., et al. (2020). An Algorithm Competition for Automatic Species Identification from Herbarium Specimens. *Applications in Plant Sciences*, 8(6), e11365.
7. Campos-Leal, J. A., Yee-Rendón, A. y Vega-López, I. F. (2022). Simplifying VGG-16 for Plant Species Identification. *IEEE Latin America Transactions*, 20(11), 2330-2338.
8. Vega-López, I. F., Vega-Aviña, R., Delgado-Vargas, F., et al. (2023). Identificación de especies de plantas de la flora mexicana utilizando aprendizaje por transferencia a través de Inception-v4. *RDE Revista Internacional de Estadística y Geografía*, 14(1), 22-37.
9. Goëau, H., Bonnet, P. y Joly, A. (2020). Overview of Lifecycle Plant Identification Task 2020. En *Proc. of the CLEF 2020-Conference and Labs of the Evaluation Forum*, 2696(140).
10. He, K., Zhang, X., et al. (2016). Deep Residual Learning for Image Recognition. En *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778.

SOBRE LOS AUTORES



Juan Augusto Campos Leal recibió su licenciatura en Sistemas Computacionales de la Universidad Autónoma de Occidente en 2008 y su maestría en Ciencias de la Computación en 2018 por la Universidad Autónoma de Sinaloa. Actualmente es estudiante de Doctorado en Ciencias de la Información en la misma universidad. Sus intereses de investigación actuales incluyen el aprendizaje profundo y las técnicas de aprendizaje automático.



Inés Fernando Vega-López es profesor de Ciencias de la Computación en la Universidad Autónoma de Sinaloa (Culiacán, México) desde 2004. Recibió su Doctorado en Ciencias de la Computación por la Universidad de Arizona en 2004. El Dr. Vega-López es miembro del Sistema Nacional de Investigadores Nivel I. Sus intereses de investigación actuales incluyen sistemas de bases de datos de alto rendimiento, análisis de datos a gran escala y aprendizaje automático.



Gerardo Beltrán-Gutiérrez es profesor de Ciencias de la Computación en la Universidad Autónoma de Sinaloa (Culiacán, México). Obtuvo su maestría en Ciencias de la Computación por la Universidad Autónoma de Sinaloa. Sus líneas de interés incluyen bases de datos, ciencia de datos y programación paralela y distribuida.



José Ramón López-Arellano es ingeniero en Sistemas de Información por el Instituto Tecnológico y de Estudios Superiores de Monterrey Campus Monterrey (ITESM). Obtuvo su maestría por EGADE Business School Monterrey y es Doctor en Gestión de las Organizaciones por las Universidades Autónoma de Nayarit, Autónoma de Sinaloa y Juárez del Estado de Durango. Actualmente, es Coordinador del Parque Científico Tecnológico (PCT) de la Universidad Autónoma de Sinaloa (UAS).



Arturo Yee-Rendón es profesor de Ciencias de la Computación en la Universidad Autónoma de Sinaloa. Obtuvo su maestría y doctorado en Ciencias de la Computación por el CINVESTAV-IPN en 2010 y 2015, respectivamente. El Dr. Yee-Rendón es miembro del Sistema Nacional de Investigadores Nivel I. Sus intereses de investigación actuales incluyen reconocimiento de patrones, técnicas de aprendizaje profundo, teoría de juegos y algoritmos de optimización (algoritmos genéticos).