

Alternative crops detection to be cultivated in Hidalgo through the analysis of satellite images

Identificación de productos agrícolas alternativos para ser cultivados en Hidalgo a través del análisis de imágenes satelitales

Roberto Samuel, HERNÁNDEZ SORIANO^{1*}, German, CUAYA SIMBRO², Rodrigo Alexander, CASTRO CAMPOS¹

¹*Universidad Autónoma Metropolitana Unidad Azcapotzalco, UAM. Av San Pablo Xalpa 180, San Martín Xochinahuac, Azcapotzalco, 02128 Ciudad de México, CDMX (0000-0002-8971-1460; 0000-0003-2275-5511)*

²*Instituto Tecnológico Superior del Oriente del Estado de Hidalgo, ITESA. Carretera Apan-Tepeapulco Km 3.5, Colonia Las Peñitas, Apan Hidalgo, México (0000-0001-6303-154X)*

Sent date: 05/September/2022 Acceptance date: 24/February/2023

Abstract:

Precision agriculture has become increasingly popular as a technique for resources optimization in the agricultural processes by implementing IT to collect and analyze crops data. Among this technology, satellite imagery for crop management has performed outstandingly well. This work applies CRISP-DM methodology, to gather some satellite images and find potential alternative crops that could be planted in the state of Hidalgo, Mexico. This constitutes an effort to fight soil erosion and desertification by avoiding monoculture practices. We built an image classification model that is capable to associate a satellite image with a crop that could be grown in the photographed area. The best of all the deep learning trained models reached an accuracy of 83.37%, based on the EfficientNet B0 architecture.

Keywords: precision agriculture, data science, monoculture.

Resumen:

La agricultura de precisión se ha popularizado como técnica para la optimización de recursos en el desarrollo de los procesos agrícolas, utilizando tecnologías de la información para recolectar y analizar datos referentes a los cultivos. Dentro de estas tecnologías se destaca el uso de imágenes satelitales para la toma de decisiones en la gestión de cultivos. Este trabajo aplica la metodología CRISP-DM, a través de la cual se busca extraer imágenes satelitales, en una primera aproximación para encontrar cultivos alternativos potenciales que puedan ser sembrados en el estado de Hidalgo, México. Esto se hace para combatir el problema de los monocultivos que resultan en el deterioro y finalmente en la desertificación de los suelos. Se generó un modelo de clasificación de imágenes, cuyo objetivo es asociar una imagen satelital con un cultivo que pueda ser plantado en la zona

representada. Se construyeron modelos con diferentes arquitecturas deep learning, obteniendo la mejor exactitud de 83.37% con el modelo implementado utilizando la arquitectura EfficientNet B0.

Palabras clave: agricultura de precisión, ciencia de datos, monocultivos.

* Corresponding author. E-mail: rshs@azc.uam.mx
Tel. 773-179-83-74

1. Introducción

Derivado del crecimiento demográfico y de la sobredemanda de productos alimenticios, el sector agrícola ha adoptado algunas prácticas no sostenibles como los monocultivos, provocando así el deterioro del suelo. En México, los monocultivos como los cereales crecieron 240% entre 1961 y 2017, y el algodón, que a 2013 se incrementó 162% (Secretaría de Medio Ambiente y Recursos Naturales, 2019). Sumando a estas prácticas la escasez de agua y los efectos del cambio climático, como el incremento de la temperatura o lluvias torrenciales, el desarrollo de los cultivos y la calidad de la comida se ven amenazados peligrosamente al estar sometidos a un estrés excesivo (Agrimonti et al., 2021). Ante esta problemática, han surgido algunas propuestas de solución como la agricultura de precisión, que es la técnica mediante la que se incluyen tecnologías de la información para adecuar el manejo de suelos y cultivos de acuerdo con la variabilidad natural de los mismos en una determinada extensión cultivable. De manera general, mediante el uso de tecnología GPS y algunos dispositivos electrónicos, se recopilan y analizan datos en tiempo real sobre algún cultivo.

Una de las tecnologías que también se ha aprovechado últimamente para el manejo de los cultivos, son las imágenes satelitales multiespectrales, las cuales pueden proporcionar información de la zona en la que se encuentra un cultivo, a fin de tomar medidas preventivas o correctivas para la gestión de éste. Las imágenes satelitales multiespectrales permiten representar datos sobre una región, en un determinado periodo, a través de una gama de colores que se corresponden con el rango de valores que el indicador deseado puede tomar. Dentro de estos indicadores, destaca el NDVI (Normalized Difference Vegetation Index) por su relevancia para diversas aplicaciones en el monitoreo ambiental, el sector agrícola, la macroecología, entre otros, debido a su facilidad de cálculo, puesto que solo requiere las mediciones de la radiación roja y cercana al infrarrojo (Pettorelli, 2013). Dichas mediciones son realizadas desde 2008 por el centro Earth Resources Observation and Science (EROS) del servicio geológico de Estados Unidos (USGS) con los sensores de la Visible Infrared Imaging Radiometer Suite (VIIRS), los cuales proporcionan una lectura de 22 bandas espectrales (Benedict et al., 2021).

Por otro lado, la ciencia de datos es un proceso a través del cual se busca explotar los datos disponibles para extraer información o patrones que soporten la toma de decisiones para la

optimización de procesos o mejora de resultados. Existen pocos trabajos (Demattê et al., 2018; Tantalaki et al., 2019) que reportan el uso de una metodología formal de ciencia de datos para un proceso de agricultura de precisión basado en el análisis de imágenes satelitales. Y más aún, un proceso de agricultura de precisión enfocado a proporcionar recomendaciones de cultivos alternativos para evitar las consecuencias de los monocultivos, lo cual es el objetivo de este trabajo de investigación.

Destaca también que los problemas planteados y, por ende, los algoritmos aplicados en los estudios que conjuntan agricultura de precisión y ciencia de datos empleando imágenes satelitales, están interesados típicamente en desarrollar una tarea de predicción, pues buscan determinar el valor de algún parámetro importante para el proceso agrícola (Fathololoumi et al., 2021), o bien, estimar la producción de un producto en una región (Nguyen et al., 2022; Novelli et al., 2019). Sin embargo, el interés de este trabajo se centra proponer productos agrícolas alternativos a los desarrollados en una determinada región, que concretamente resulta en un problema de clasificación.

2. Materiales y Métodos

La metodología de ciencia de datos aplicada para realizar un proceso de agricultura de precisión fue la Cross Industry Standard Process for Data Mining (CRISP-DM), que consta de seis etapas iterativas (Figura 1) y que ha sido ampliamente utilizada en proyectos de ciencia de datos (Saltz, 2021), cuyas etapas se describen a continuación, mientras que la quinta se presenta hasta la siguiente sección.

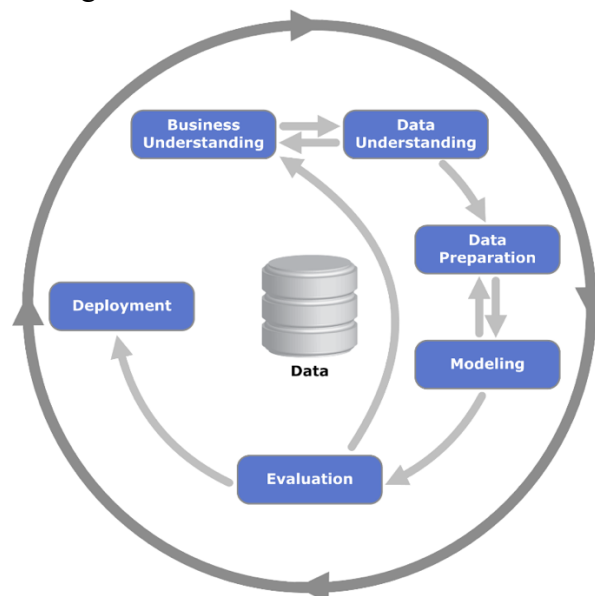


Figura 1. Diagrama de la metodología CRISP-DM.
Fuente: extraído de Saltz, 2021.

2.1. Entendimiento del problema

Como se ha comentado, los procesos agrícolas enfrentan nuevas y cada vez más críticas situaciones que someten a los cultivos a un mayor estrés. Además, la práctica extendida de los monocultivos ha deteriorado los suelos al reducir la diversidad y extensión de la población de las comunidades microbianas en las tierras cultivables, las cuales juegan un rol fundamental en el desarrollo de las plantas (Mallano et al., 2021).

Cabe mencionar que se tiene un interés particular en el estado de Hidalgo, por ser la región en la que se desarrolla esta investigación. Es por ello que los datos empleados para la generación del modelo serán referente a esta entidad. En todo caso, el proceso descrito puede ser replicado con información de otras entidades federativas.

Ahora bien, como lo menciona Tapan (2016), los monocultivos tienen un impacto ambiental fuerte, resultando en fallos para preservar la diversidad genética en la flora debido a la falta de biodiversidad, en la degradación del suelo y vulnerándolo ante la erosión. Así mismo, esta alteración en la vegetación provoca un desequilibrio en el ecosistema, tanto en la composición fisicoquímica del suelo al reducir el carbono en el suelo, nutrientes esenciales y el pH, como en una acumulación bioquímica menos diversa de las comunidades microbiológicas en el suelo. Esto a su vez resulta en la lixiviación de algunos nutrientes, provocando la contaminación del suelo y la reducción de la calidad del agua en el entorno. Por su parte, la acidificación del suelo reduce la calidad del fruto y propicia la propagación de hongos más que de bacterias, generando una reacción en cadena adversa contra las dinámicas y los organismos de niveles más altos en las redes tróficas (Pervaiz et al. 2020).

Las prácticas de monocultivos se tienen un trasfondo social y económico profundo, pues se busca satisfacer la sobredemanda de los productos agrícolas con los que se practican. Al existir sobredemanda, el beneficio económico suele ser mayor, además de representar un riesgo de la misma naturaleza mucho menor. Sin embargo, el desconocimiento de la viabilidad de la rotación de cultivos con productos similares que satisfagan las mismas necesidades también es parte del problema (Tapan, 2016).

2.2. Entendimiento de los datos

Como se mencionó, la información analizada para proponer una solución al problema fueron las imágenes satelitales multiespectrales, a través de las que se representan algunos indicadores propios de determinada región. Por otra parte, también se requiere saber el tipo de cultivo sembrado y cosechado en cada una de las regiones analizadas.

Para ello se buscaron conjuntos de datos (imágenes satelitales) que proporcionaran información temporal, debido a la variabilidad de los cultivos predominantes en una misma

región, además de que los cultivos tienen ciclos de vida diferentes. Se consiguieron entonces imágenes satelitales semanales de la república mexicana por semana desde febrero del 2012 y hasta junio de 2022, en donde se agregaba además el indicador eVIIRS NDVI. Dichas imágenes se obtuvieron del portal Earth Observation de la Administración Nacional de Aeronáutica y el Espacio (NASA por sus siglas en inglés) de Estados Unidos (Figura 2). Cada una de las imágenes tiene una resolución de 1024×447 píxeles.

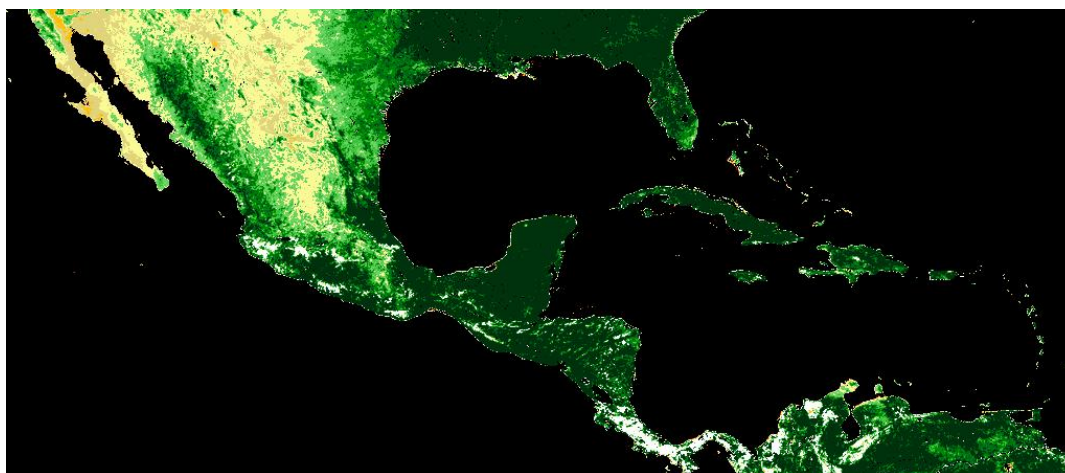


Figura 2. Imagen con el indicador eVIIRS NDVI de la zona de Centroamérica y el Caribe en el periodo del 21 de junio de 2022 al 30 de junio de 2022.

Fuente: extraído del portal de USGS Earth Explorer.

A su vez, se obtuvieron las estadísticas del avance mensual en la producción agrícola desde enero de 2018 hasta junio de 2022 en la república mexicana en el portal del Servicio de Información Agroalimentaria y Pesquera (SIAP) del gobierno de México. Estos datos son abiertos, pueden consultarse a nivel nacional, estatal, distrital o municipal y consisten en las hectáreas sembradas, cosechadas y siniestradas acumuladas hasta el mes y año agrícola correspondiente, como se observa en la Figura 3.

Entidad		Distrito	Superficie (ha)			Producción	Rendimiento (udm/ha)
			Sembrada	Cosechada	Siniestrada		
1	Hidalgo	Huejutla	2,467.66	0.00	0.00	0	

Figura 3. Información provista por el SIAP sobre el avance de siembras y cosechas de maíz grano del año agrícola 2022 al 31 de octubre de 2021. A esta fecha se tenían registradas 2,467.66

2.3. Preparación de los datos

Se realizaron varios tratamientos a los datos. Primeramente, se modificó el formato en el que se guardaron las estadísticas de cultivo, para facilitar su lectura y manipulación con los códigos desarrollados, cambiando de CSV a JSON. Luego, como se mencionó, se depuraron las estadísticas acumulativas de los cultivos obtenidas a fin de obtener el total de las hectáreas sembradas activas por cultivo en cada mes, esto debido a que la mayoría de las hectáreas cultivadas serían la muestra más representativa dentro de la imagen satelital con la que se busca una correspondencia. Se entiende por hectáreas sembradas activas a aquellas que fueron sembradas y que aún no han sido cosechadas ni se han perdido (siniestradas). Así, para cada registro con el que se cuenta se calcula la expresión:

$$\text{Sembradas Activas} = \text{Sembrada} - \text{Cosechada} - \text{Siniestrada}$$

Por ejemplo, en la Figura 5, el total de hectáreas activas de los cultivos de “Maíz grano” serían de 26,864.12, 14,880.05, 34,363.84, 15,205.11, 43,664.86 y 21,537.32 para los distritos de Huejutla, Zacualtipán, Tulancingo, Pachuca, Mixquiahuala y Huichapan respectivamente, esto para el mes de octubre de 2021 según los registros del año agrícola del 2021. Sin embargo, es importante considerar también el hecho de que puede haber registros de años agrícolas distintos que contribuyan a la estadística de un mismo mes. De hecho, esto sucede para el caso de ejemplo presentado, pues existen registros de hectáreas sembradas con maíz grano solo en el distrito de Huejutla, pero cuyo registro se realizó como producción del año agrícola 2022 (Figura 3). Así, las hectáreas sembradas activas totales del distrito de Huejutla para octubre del 2021 cambiarían a 29,331.78, considerando todos los registros del avance de siembras y cosechas.

	Entidad	Distrito	Superficie (ha)			Producción	Rendimiento (udm/ha)
			Sembrada	Cosechada	Siniestrada		
1	Hidalgo	Huejutla	42,041.02	15,176.90	0.00	19,563.34	1.29
2	Hidalgo	Zacualtipán	24,707.10	9,827.05	0.00	16,458.54	1.67
3	Hidalgo	Tulancingo	42,790.20	2,088.16	6,338.20	2,556.54	1.22
4	Hidalgo	Pachuca	19,909.84	2,657.73	2,047.00	6,186.00	2.33
5	Hidalgo	Mixquiahuala	52,681.91	8,512.00	505.05	84,370.42	9.91
6	Hidalgo	Huichapan	28,213.48	6,676.16	0.00	29,646.40	4.44

Figura 5. Información provista por el SIAP sobre el avance de siembras y cosechas de maíz grano del año agrícola 2021, al 31 de octubre de 2021.

Fuente: extraído del portal de avance de siembras y cosechas del SIAP.

Una vez que se contó con el cultivo más representativo por distrito por mes, se seleccionó mediante un análisis modal el cultivo que más hectáreas sembradas activas tenía a lo largo de los meses, durante los años en los que se contaba con estadísticas. Dicho cultivo predominante se utilizó como etiqueta para todas las imágenes representativas de ese distrito. Cabe mencionar que para cinco de los seis distritos que constituyen el estado de Hidalgo, se observó la predominancia indiscutible de un tipo de cultivo en cada uno de los distritos, reforzando la percepción del problema de los monocultivos planteado.

Ahora bien, las imágenes satelitales provistas por la NASA corresponden a la zona de Centroamérica y la región del Caribe, por lo que se realizó una interpolación para calcular los píxeles que corresponden a cada una de las muestras planteadas para los distritos requeridos. La interpolación se realizó con base en las coordenadas de los municipios que abarcan cada uno de los distritos de Hidalgo y que se obtuvieron del Instituto Nacional de Estadística Geografía (INEGI) (Figura 4) y con las coordenadas de las imágenes de la NASA que en latitud eran de 7° 0' 17" N a 33° 0' 0" N, y en longitud eran de 59° 0' 26.13" O a 118° 30' 0" O (Tabla 1). Así, se generaron las imágenes a partir de los píxeles calculados de la figura 6.

Tabla 1. Coordenadas de las esquinas opuestas del cuadrilátero en donde se delimita la muestra de los distritos de desarrollo rural hidalguenses y sus píxeles correspondientes.

Distrito de Desarrollo Rural	Coordenadas esquina superior izquierda	Coordenadas esquina inferior derecha	Píxeles [x0:xf, y0:yf]
Huejutla	21° 7' 59" N, 98° 32' 18" O	20° 46' 36" N, 98° 17' 12" O	[682:676, 244:237]
Zacualtipán	21° 0' 50" N, 98° 50' 37" O	20° 31' 57" N, 98° 35' 0" O	[688:682, 241:233]
Mixquiahuala	20° 33' 6" N, 99° 20' 58" O	20° 0' 27" N, 98° 57' 50" O	[694:687, 234:222]
Huichapan	21° 5' 57" N, 99° 22' 59" O	20° 33' 6" N, 99° 7' 23" O	[700:692, 244:228]
Pachuca	19° 49' 4" N, 98° 32' 50" O	19° 39' 25" N, 98° 24' 12" O	[686:678, 228:218]
Tulancingo	20° 20' 58" N, 98° 31' 2" O	20° 2' 7" N, 98° 12' 11" O	[680:672, 231:224]



Figura 6. Imágenes generadas a partir de la imagen satelital de la Figura 2, con base en los pixeles de la Tabla 1, para cada uno de los DDR de Hidalgo: Huejutla (a), Zacualtipán (b), Mixquiahuala (c), Huichapan (d), Pachuca (e) y Tulancingo (f).

Fuente: elaboración propia.

Es importante mencionar que todos los procesos realizados, tanto el cálculo de las hectáreas sembradas activas como la interpolación de los pixeles se realizaron a través de programas desarrollados en Node JS, debido a la previa experiencia de los autores con este. Por su parte, la generación de las imágenes distritales se realizó mediante la biblioteca libre de visión artificial OpenCV en Python.

2.4. Modelado

Para el modelado, se utilizaron los recursos provistos por Google en Google Cloud a través de las interfaces de programación de la AI Platform Training & Prediction y Compute Engine para la clasificación de imágenes, los cuales están basados en la infraestructura provista por TensorFlow.

Disponiendo de los algoritmos de entrenamiento de la plataforma, se realizaron las primeras pruebas de entrenamiento (Tabla 2), generando así diversos modelos de clasificación de imágenes basada en redes neuronales basados en las redes Efficient Net B0, B2 y B7, y Resnet 50 y Resnet 101, todos implementados a través del API Keras, de TensorFlow.

Tabla 2. Parámetros de la primera ronda de entrenamiento de los modelos de clasificación de imágenes.

Prueba	Algoritmo	Iteraciones	Tamaño de la muestra
1	EfficientNet B0	10,000	128
2	EfficientNet B2	10,000	64
3	EfficientNet B7	10,000	8
4	ResNet 50	10,000	128
5	ResNet 101	10,000	128

3. Resultados y Discusión

3.1. Creación del conjunto de datos

Contando con seis distritos en el estado de Hidalgo, que es la zona de interés, y un total de 750 imágenes satelitales de la serie temporal, se obtuvieron en total 4500 imágenes desde 3×2 píxeles y hasta 5×9 píxeles, asociadas a dos tipos de cultivos, puesto que uno de los productos agrícolas predominantes era el mismo para dos distritos (Tabla 3). De acuerdo entonces con el conjunto de datos generado, se tiene un problema multiclase con dos clases correspondientes a los cultivos “Maíz grano” y “Tuna”.

Tabla 3. Etiquetas asignadas a las imágenes obtenidas de cada distrito.

Distrito de Desarrollo Rural	Cultivo asociado
Huejutla	Maíz grano
Huichapan	Maíz grano
Mixquiahuala	Maíz grano
Pachuca	Tuna
Tulancingo	Maíz grano
Zacualtipán	Maíz grano

3.2. Creación y evaluación de modelos

Del conjunto de 4500 imágenes, se utilizaron 3378 para el entrenamiento de los modelos, 900 para la evaluación de estos y se reservaron además 222 para su validación. Los resultados de la primera ronda de entrenamientos se muestran en la Tabla 4. La métrica “exactitud” indicada como parte de los resultados se calculó considerando, de las 900 imágenes que forman parte del conjunto de evaluación, aquellas que fueron clasificadas correctamente con su etiqueta.

Tabla 4. Resultados de la primera ronda de entrenamiento de los modelos de clasificación de imágenes.

Prueba	Algoritmo	Horas de ejecución	Exactitud
1	EfficientNet B0	38	83.37%
2	EfficientNet B2	41	83.37%
3	EfficientNet B7	146	82.87%

4	ResNet 50	94	83.37%
5	ResNet 101	149	83.37%

De los experimentos que obtuvieron un mejor desempeño considerando la relación exactitud-horas de ejecución que tomó el entrenamiento, se realizaron otras pruebas de entrenamiento que retomaran el modelo entrenado hasta ese punto para mejorar la métrica mencionada. Los resultados de estas pruebas se muestran en la Tabla 5. Cabe mencionar que las horas de ejecución varían dependiendo de la estructura misma del algoritmo de entrenamiento, y que todos los entrenamientos de las redes se realizaron empleando la misma capacidad de hardware: la máquina heredada large_model de Google AI Platform Training.

Tabla 5. Parámetros y resultados de la segunda ronda de entrenamiento de los modelos.

Prueba	Algoritmo	Iteraciones	Tamaño de la muestra	Horas de ejecución	Exactitud
1	EfficientNet B0	10,000	128	47	83.37%
2	EfficientNet B2	10,000	128	40	83.37%

Se observa que la medida de exactitud para la mayoría de los modelos es de 83.37%, que corresponde con la proporción de imágenes del conjunto de evaluación que debían ser clasificadas como la etiqueta “Maíz grano”. Por su parte, se observa una mayor variabilidad en los resultados de salida para el modelo entrenado con el algoritmo EfficientNet B7, cuyo valor de exactitud alcanzó 82.87% con el repositorio de las imágenes para evaluación del modelo.

3.3. *Discusión de resultados*

Se sabe que mientras mayor sea la resolución espacial en una imagen satelital, mayor será la exactitud en los resultados y dictámenes emitidos por los modelos generados (Skakun et al., 2021). En ese trabajo, se realizaron simulaciones con imágenes de 3, 10, 20 y 30 m de resolución espacial y se explicó respectivamente el 100%, 86% 72% y 59% de los casos de la variabilidad del rendimiento dentro del campo. A pesar de las diferencias en la aplicación final del trabajo citado y el presente, se considera que se obtuvo una modelo con una buena exactitud, puesto que se trabajó con una resolución de 6.47 km por pixel.

Por otro lado, se observa que los algoritmos de EfficientNet tuvieron un mejor desempeño que los del tipo ResNet, lo cual era esperado debido al método compuesto de escalamiento que utilizan los primeros (Tan et al., 2019).

Además, cabe señalar que esta es una primera aproximación, la cual da indicios de la viabilidad del uso del tipo de datos para abordar el problema y que se puede utilizar información más detallada para mejorar el rendimiento de los modelos.


Con la metodología presentada se pueden obtener imágenes satelitales de otras regiones cuyos suelos y condiciones climatológicas sean similares y se pueden presentar dos situaciones: si se conoce el cultivo desarrollado en esa zona, el modelo indicará la similitud con alguno de los cultivos predominantes en el estado de Hidalgo, determinando así un cultivo alternativo para la región. Por otro lado, si no se conoce el cultivo en la zona o no es una región en la que se desarrolle la agricultura, la respuesta del modelo será el producto agrícola que podría considerarse para cultivar.

Proponiendo, por ejemplo, el DDR de Celaya en el estado de Guanajuato con condiciones similares a las de Hidalgo, y el DDR de Tapachula en Chiapas con condiciones un poco más contrastantes, se profundiza primero en el análisis modal de las estadísticas de producción agrícola, observando que, los cultivos predominantes en el mes de marzo fueron “cebada grano” y “café cereza” para Celaya y Tapachula respectivamente. Posteriormente, se determinan las coordenadas del distrito a partir de las coordenadas de los municipios que lo integran (Tabla 6), esto para generar las imágenes satelitales distritales que estarán asociadas a esta etiqueta (Tabla 7). La imagen satelital general empleada para la generación de las imágenes de estos distritos fue la correspondiente al periodo del 6 al 15 de marzo de 2020.

Tabla 6. Coordenadas de las esquinas opuestas del cuadrilátero en donde se delimita la muestra de los distritos de desarrollo rural hidalguenses y sus píxeles correspondientes.

Distrito de Desarrollo Rural	Coordenadas esquina superior izquierda	Coordenadas esquina inferior derecha	Píxeles [x0:xf, y0:yf]
Celaya, Guanajuato	20° 43' 12" N, 100° 48' 48" O	20° 11' 55" N, 100° 30' 37" O	[304:310, 211:220]
Tapachula, Chiapas	15° 20' 27" N, 92° 39' 30" O	14° 50' 14" N, 92° 10' 5" O	[445:453, 304:312]

Tabla 7. Etiquetas e imágenes obtenidas de cada distrito propuesto

Distrito de Desarrollo Rural	Cultivo asociado	Imagen Satelital
Celaya, Guanajuato	Cebada grano	

Tapachula, Chiapas

Café cereza



Se procede entonces con la evaluación del modelo reportado, entrenado con el algoritmo EfficientNet B0, proporcionando como entrada las imágenes satelitales de la Tabla 7, y recibiendo a la salida alguno de los cultivos predominantes en el estado de Hidalgo. Para ambos casos, la etiqueta asignada fue “Maíz grano”.

Como se mencionó, debido a que se conoce el cultivo asociado a la imagen satelital, se puede decir que tanto la cebada grano como el café cereza son cultivos alternativos viables al maíz en el estado de Hidalgo. De hecho, profundizando en el análisis modal realizado sobre la estadística, se observa que destaca la producción de café cereza en Hidalgo, lo que respalda la viabilidad del desarrollo de este producto en la zona.

Ahora, el otro enfoque del resultado obtenido es que, si se desconociera el cultivo asociado a las imágenes satelitales probadas, el modelo sugiere que tanto en Celaya como en Tapachula es viable el cultivo de maíz. Y nuevamente, en la estadística mensual calculada es claro que efectivamente el distrito de Celaya cuenta con una producción importante de “Maíz grano”, corroborando el resultado obtenido.

4. Conclusiones

Las propuestas de solución a los retos y problemas complejos y no lineales planteados por la agricultura de precisión suelen englobar datos de diversos orígenes y naturalezas, derivado de la amplia variedad de factores interrelacionados que afectan el desarrollo agrícola. Por ello, resulta conveniente analizarlos a través de algoritmos de aprendizaje profundo que facilitan el trabajo con datos heterogéneos de diversas fuentes.

Se integró el uso de las imágenes satelitales multispectrales en un proceso de ciencia de datos para realizar agricultura de precisión, generando un modelo de clasificación de imágenes satelitales que asocia esa entrada con un cultivo potencial a desarrollar, en este caso en el estado de Hidalgo.

Como trabajo futuro, se puede buscar mejorar la efectividad de los modelos. Esto podría lograrse, por un lado acrecentando el conjunto de entrenamiento, incluyendo imágenes de los distritos de desarrollo rural en otras entidades federativas que primero desarrollen los mismos cultivos y que después agreguen otro tipo de cultivos como etiquetas de clasificación; por otro lado, a través de la ampliación en la resolución espacial de las imágenes satelitales empleadas, lo que resultaría en un conjunto de entrenamiento más vasto, nítido y por tanto de mayor calidad de las regiones de interés y a su vez podría abrir la posibilidad de realizar

una clasificación a nivel municipal y no distrital, acotando el error debido a la generalización de un cultivo principal en una zona más extensa.

Finalmente, también como trabajo futuro, el modelo podría emplearse en apoyo a la política agraria desde dos perspectivas. En primera instancia puede realizarse una comparativa como la presentada en la última parte de la sección “3.3 Discusión de resultados” mucho más profunda, que incluyera otras regiones y meses en el año que permitan integrar un catálogo de cultivos alternativos mucho más amplio para su posterior difusión. Por otra parte, puede publicarse el modelo con una interfaz que permita la selección de una región para la clasificación por parte del modelo y que muestre la etiqueta que sugiera el cultivo a desarrollar en la zona seleccionada.

5. Referencias

- Secretaría de Medio Ambiente y Recursos Naturales. (2019). Monocultivos agrícolas y ganaderos, incompatibles con el entorno social. Available at: <https://www.gob.mx/semarnat/articulos/monocultivos-agricolas-y-ganaderos-incompatibles-con-el-entorno-social>. Accessed: Septiembre 2, 2022.
- Agrimonti, C., Lauro, M., & Visioli, G. (2021). Smart agriculture for food quality: facing climate change in the 21st century. *Critical Reviews in Food Science and Nutrition*, 61(6), 971–981. <https://doi.uam.elogim.com/10.1080/10408398.2020.1749555>
- Pettorelli, N. (2013). *The Normalized Difference Vegetation Index*. Editorial OUP Oxford, Reino Unido.
- Benedict, T. D., Brown, J. F., Boyte, S. P., Howard, D. M., Fuchs, B. A., Wardlow, B. D., Tadesse, T., Evenson, K. A., & ZHIZHIN, M. (2021). Exploring VIIRS Continuity with MODIS in an Expedited Capability for Monitoring Drought-Related Vegetation Conditions. *Remote Sensing*, 13(6), 1210. <https://doi.org/10.3390/rs13061210>
- Demattê, J. A. M., Fongaro, C. T., Rizzo, R., & Safanelli, J. L. (2018). Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. *Remote Sensing of Environment*, 212, 161–175. <https://doi.uam.elogim.com/10.1016/j.rse.2018.04.047>
- Tantalaki, N., Souravlas, S., & Roumeliotis, M. (2019). Data-Driven Decision Making in Precision Agriculture: The Rise of Big Data in Agricultural Systems. *Journal of Agricultural & Food Information*, 20(4), 344–380. <https://doi.uam.elogim.com/10.1080/10496505.2019.1638264>
- Fatholouloumi, S., Vaezi, A. R., Alavipanah, S. K., Ghorbani, A., Saurette, D., & Biswas, A. (2021). Effect of multi-temporal satellite images on soil moisture prediction using a digital soil mapping approach. *Geoderma*, 385. <https://doi.uam.elogim.com/10.1016/j.geoderma.2020.114901>

- Nguyen, L. H., Robinson, S., & Galpern, P. (2022). Medium-resolution multispectral satellite imagery in precision agriculture: mapping precision canola (*Brassica napus* L.) yield using Sentinel-2 time series. *Precision Agriculture*, 23(3), 1051–1071. <https://doi.uam.elogim.com/10.1007/s11119-022-09874-7>
- Novelli, F., & Vuolo, F. (2019). Assimilation of sentinel-2 leaf area index data into a physically-based crop growth model for yield estimation. *Agronomy*, 9(5). <https://doi.uam.elogim.com/10.3390/agronomy9050255>
- Saltz, J. S. (2021). CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps. *2021 IEEE International Conference on Big Data (Big Data), Big Data (Big Data), 2021 IEEE International Conference On*, 2337–2344. <https://doi.uam.elogim.com/10.1109/BigData52589.2021.9671634>
- Mallano, A. I., Zhao, X., Sun, Y., Jiang, G., & Chao, H. (2021). Continuous monocropping highly affect the composition and diversity of microbial communities in peanut (*Arachis hypogaea* L.). *NOTULAE BOTANICAE HORTI AGROBOTANICI CLUJ-NAPOCA*, 49(4), 12532. <https://doi.uam.elogim.com/10.15835/nbha49412532>
- Tapan Kumar Nath. (2016). *Monoculture Farming: Global Practices, Ecological Impact and Benefits/Drawbacks*. Nova.
- Pervaiz, Z. H., Iqbal, J., Zhang, Q., Chen, D., Wei, H., & Saleem, M. (2020). Continuous Cropping Alters Multiple Biotic and Abiotic Indicators of Soil Health. *Soil Systems*, 4(4), 59. <https://doi.org/10.3390/soilsystems4040059>
- Cámara de Diputados del H. Congreso de la Unión. (2021). Ley de Desarrollo Rural Sustentable. Available at: https://www.diputados.gob.mx/LeyesBiblio/pdf/235_030621.pdf. Accessed: Septiembre 2, 2022.
- Tan, Mingxing and Le, Quoc V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *International conference on machine learning*. <https://doi.org/10.48550/arxiv.1905.11946>
- Skakun, S., Kalecinski, N. I., Brown, M. G. L., Johnson, D. M., Vermote, E. F., Roger, J.-C., Franch, B., & Mondino, E. B. (2021). Assessing within-Field Corn and Soybean Yield Variability from WorldView-3, Planet, Sentinel-2, and Landsat 8 Satellite Imagery. *Remote Sensing*, 13(5), 872. <https://doi.org/10.3390/rs13050872>