

Machine learning for soil analysis

Uso del aprendizaje máquina para el análisis de suelos

Luis Enrique Partida-Aguilar¹, Daniel Fajardo-Delgado^{*1}, María Guadalupe Sánchez¹,
Raquel Ochoa-Ornelas¹, Himer Ávila-George², and Jesús Ezequiel Molinar-Solis¹

¹ *Tecnológico Nacional de México, Campus Cd. Guzmán,*
Departamento de Sistemas y Computación,
Av. Tecnológico 100, Cd. Guzmán, Jal., México, 49100.

kikes.1295@gmail.com, {daniel.fd, maria.sc1, raquel.oo, jesus.ms}@cdguzman.tecnm.mx

² *Universidad de Guadalajara, Centro Universitario de los Valles,*
Departamento de Ciencias Computacionales e Ingenierías
Carr. a Guadalajara km. 45.5, Ameca, Jal., México, 46600.
himer.avila@academicos.udg.mx

Abstract

The Pfeiffer circular chromatography (PCC) is a qualitative analysis technique that provides microbiological, mineral, and organic matter of the health status of a soil. This paper addresses the building of a data set conformed by PCC images from different types of soils, based on the following three variables: acidity, electrical conductivity, and soil texture. This work also explores the use of deep learning techniques to automatically extract the characteristics of such images and classify the soils based on their type. It is the first time that learning techniques are applied to this classification problem. Experimental results show an F1-score of 0.7889 in the soil texture classification suggesting a significant relationship between this variable with the PCC. On the other hand, results also show a low correlation of the PCC with acidity and electrical conductivity.

Keywords— Pfeiffer circular chromatography, machine learning, deep learning, soil analysis.

Resumen

La cromatografía circular de Pfeiffer (PCC, por sus siglas en inglés) es una técnica de análisis cualitativa que provee de información microbiológica, mineral y de materia orgánica del estado de salud de un suelo. El presente trabajo trata de la conformación de un conjunto de datos de imágenes PCC de diferentes tipos de

suelos con base en tres variables: acidez, conductividad eléctrica y textura del suelo. Este trabajo también explora el uso de técnicas de aprendizaje profundo para extraer automáticamente las características de las imágenes y clasificar los suelos con base en su tipo. Es la primera vez que se utilizan este tipo de técnicas de aprendizaje para este problema de clasificación. Los resultados experimentales muestran un puntaje-F1 de 0.7889 para la clasificación de la textura de los suelos, lo que sugiere una relación significativa de esta variable con el PCC. Por otra parte, los resultados también sugieren una baja correlación de los PCC con las variables de acidez y conductividad eléctrica de los suelos.

Palabras clave— Cromatografía circular de Pfeiffer, aprendizaje máquina, aprendizaje profundo, análisis de suelos.

I. Introducción

El análisis de suelos es una herramienta de diagnóstico que permite identificar problemas nutricionales y requerimientos de fertilización de plantas y cultivos [1]. Actualmente, este tipo de análisis se ha convertido en un paso importante en cualquier ciclo agrícola, donde es necesario evaluar el balance de nutrientes esenciales para el desarrollo adecuado de un cultivo. Uno de los métodos más empleados para el análisis de suelos es la cromatografía, un conjunto de técnicas basadas en el principio de retención selectiva para la caracterización de mezclas complejas [2]. Existen diferentes técnicas de cromatografía, las cuales se realizan generalmente en

* Autor de correspondencia

laboratorios comerciales cuyo proceso de análisis puede ser lento y costoso [3].

Recientemente, la PCC ha tomado auge como una alternativa económica y sustentable para el diagnóstico de la calidad de los suelos. La PCC provee de información microbiológica, mineral y de materia orgánica del estado de salud de un suelo [4]. Esta información se representa en diferentes círculos radiales sobre un papel filtro impregnado denominado cromatograma (ver Fig. 1). La imagen circular tiene cuatro zonas distintas, que proveen de información visual que puede interpretarse con base al tamaño, forma y colores de éstas [3, 5].

La interpretación biológica de los PCC requiere de experiencia en el reconocimiento de patrones, colores y formas que se presentan en la imagen, así como de conocimiento del dominio. De acuerdo con la interpretación de [2], la zona central del PCC es el lugar donde todas las sustancias presentes en la muestra de suelo circulan debido al fenómeno de capilaridad. Cuando esta zona no se manifiesta, indica la inexistencia de material orgánico. Si se manifiesta con un color blanco bien definido, revela que el suelo recibe dosis excesivas de abonos nitrogenados y altamente solubles, o se encuentra sometido a una constante aplicación de herbicidas. La situación ideal es cuando se tiene una coloración blanco-cremosa que se desvanece suavemente.

Por otro lado, la zona interna contiene información acerca de minerales tales como el nitrógeno, sodio, potasio y fósforo. El color pardo o negruzco en esta zona indica un suelo probablemente erosionado, muy mineralizado y sin actividad biológica. Sin embargo, una coloración que proviene de la zona central y se integra gradualmente hasta la zona media de forma armónica y sin interrupciones, es un indicador de plena actividad microbiológica integrada y diversa.

La zona intermedia revela información acerca de la presencia o ausencia de carbono y materia orgánicos. Finalmente, la zona externa trata acerca del humus, un proceso de descomposición orgánico propiciado por hongos y bacterias y que provee de nutrientes como nitrógeno, fósforo, potasio y magnesio a la tierra y las plantas.

En el presente trabajo, se explora el uso de técnicas de aprendizaje máquina para el análisis automático de suelos a través de la PCC. En particular, se utiliza el paradigma del aprendizaje profundo para extraer automáticamente las características de las imágenes y clasificar los suelos con base en su tipo. Para este fin, fue necesaria la creación de un conjunto de datos de imágenes PCC con tres variables de caracterización: acidez, conductividad eléctrica y textura del suelo, así como una variable adicional relacionada a la salud general del suelo. Estas imágenes se utilizan como instancias para el entrenamiento de los modelos de clasificación.

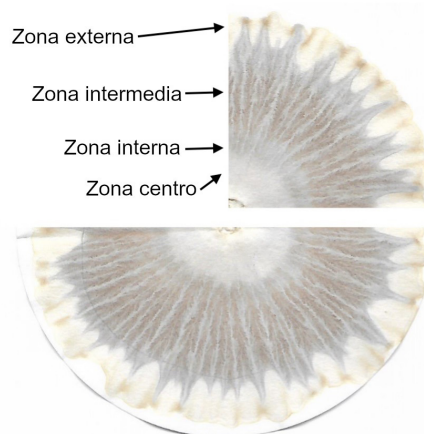


Figura 1: Cromatograma circular de Pfeiffer

II. Conjunto de imágenes PCC

Al inicio del desarrollo del presente trabajo no existía una base de datos pública con imágenes PCC, por lo que fue necesario construir una para el entrenamiento de los modelos de aprendizaje. Para ello, se contó con la colaboración y el apoyo de investigadores del Instituto Tecnológico de Tlajomulco (ITT), quien brindó la asesoría y permitió el uso de instalaciones del laboratorio de suelos y de reactivos químicos para la generación e interpretación de los PCC.

Los PCC se generaron utilizando la metodología descrita en [2], tanto para la recolección de muestras como para el trabajo de laboratorio. Se utilizaron 79 muestras de suelos de distintos terrenos y predios del sur de Jalisco, que fueron recolectados por agricultores y llevados para su análisis al laboratorio de suelos del ITT. Es el propio agricultor quien realiza un diagnóstico integral del predio de donde se obtienen las muestras, informando al laboratorio del tratamiento que se le ha brindado al terreno y de las tareas (químicas, mecánicas, biológicas) actuales y previas en el cultivo. A continuación, se describe el proceso que los agricultores siguieron en la recolección de las muestras de suelo.

II.1. Recolección de muestras

Para obtener las muestras del suelo, el agricultor debió tomar en cuenta el tipo de cultivo y la extensión de la propiedad, ya que a partir de estos datos se toma la decisión del método y de la cantidad de muestras requeridas. Los métodos más comunes para la obtención de muestras de suelos son dos: el muestreo de rejilla y el de zonas (ver [6] para más información). El método empleado por cada agricultor dependió de las condiciones del predio y de las recomendaciones hechas por los investigadores del ITT. La cantidad de suelo extraído fue de entre 0.5 kg a

1 kg. La profundidad varió dependiendo del propósito de cada cultivo; ej., en hortalizas se puede tomar de 10 a 40 cm, para pastos forrajeros de 20 a 50 cm, y para terrenos frutales hasta 1.5 m de profundidad.

Cada muestra se rotuló y marcó con datos relevantes (ej., el lugar de extracción y la profundidad de excavación). Luego, se procedió al secado de las muestras. Esto generalmente se hace a media sombra, evitando luz directa y la humedad. Con las muestras ya secas, se tomaron cantidades de 100 a 150 g haciéndolas pasar por un colador plástico, y se pasó a la molienda de la muestra. En este paso, se maceró la muestra en un recipiente de porcelana hasta obtener polvo fino.

II.2. Trabajo de laboratorio

Una vez que se recolectaron las muestras, se siguió la metodología propuesta en [2] para las tareas de laboratorio. Se utilizaron los siguientes materiales: tres cajas de papel filtro N° 1 de 15 cm (150 mm) de diámetro, que sirven de base principal donde se llevan a cabo las reacciones químicas y se visualizará la imagen PCC resultante; una balanza electrónica capaz de pesar de 1/2 g hasta 100 g; cajitas Petri de 5 cm de diámetro (pueden ser plásticas); 100 g de hidróxido de sodio también conocido como soda cáustica o sosa cáustica; y 5 g de nitrato de plata en cristales. A continuación, se describe la metodología para la generación de un PCC.

Primeramente, se prepara una solución de nitrato de plata (AgNO_3) diluido al 0.5 %, por lo cual se vierte en un recipiente de cristal 5 g de nitrato de plata en 1 l de agua destilada y con apoyo de un agitador magnético las sustancias se mezclan (ver Fig. 2). Después, se procede a la elaboración de una solución de hidróxido de sodio (NaOH) o soda cáustica al 1 % en agua destilada. Para este paso se requiere preparar 10 l de la solución, por lo que es necesario 100 g de hidróxido de sodio. Esta cantidad es suficiente para realizar las muestras de suelo del presente trabajo. De manera similar que la solución de AgNO_3 , los 100 g de hidróxido de sodio se vierten en un recipiente de cristal con 1 l de agua destilada y con apoyo del agitador magnético las sustancias se mezclan. Luego la solución resultante se diluye en 9 l más de agua destilada. Posteriormente la solución se almacena en una garrafa plástica.

Después, se preparan los papeles filtro para ser impregnados. Esto se logra haciendo una perforación en el centro del papel filtro y haciendo dos pequeñas marcas, con ayuda de la punta de un clip, a los 4 y 6 cm (desde el centro). También se recortan unos pequeños cuadros de papel filtro con medidas de 2×2 cm que se enrollan de manera que puedan colocarse en la perforación central del papel filtro. A estos rollos de papel filtro se les conoce comúnmente como pabilos. Acto seguido se realiza la impregnación del papel filtro con nitrato de plata, esto



Figura 2: Solución de nitrato de plata al 0.5 %

se logra colocando el papel filtro con su pabilo sobre una caja de Petri donde se vierte una cantidad suficiente de la solución para que el papel filtro se impregne hasta la de los 4 cm (ver Fig. 3). Una vez impregnados, se retira el pabilo y se resguardan entre hojas blancas por un día lejos de la luz del sol.



Figura 3: Impregnación de papel filtro

Posteriormente, se lleva a cabo la extracción de submuestras de cada una de las muestras. Para ello, se extraen 5 g de suelo, esto con el fin de obtener una disolución correcta con la solución de NaOH preparada. Cada submuestra se almacena en un matraz para una mejor manipulación (ver Fig. 4).



Figura 4: Solución de nitrato de plata al 0.5 %

Lo siguiente en las tareas de laboratorio, es mezclar las submuestras de suelo con la solución de NaOH al 1 %. A cada matraz con los 5 g de suelo se le vierten 50 ml de solución de NaOH . Inmediatamente comienza el proceso de agitado, donde a cada matraz se le implementa 7

agitaciones manuales a manera de giros hacia la derecha y 7 hacia la izquierda, repitiendo esta secuencia hasta completar 49 agitaciones. Las mezclas se dejan reposar por 15 minutos, cumplido el plazo se repite el proceso de agitado, luego se deja reposar por 1 hora más y, por última vez, se repite el proceso de agitado. Finalmente, se dejan reposar por 6 horas más.

Una vez transcurridas las 6 horas, se lleva a cabo la impregnación de las disoluciones de suelo en los papeles filtro impregnados anteriormente con la solución preparada de AgNO_3 . Para completar este paso primero se coloca un nuevo pabalo a cada papel filtro. Después, se vierte en las cajas Petri una cantidad suficiente de cada mezcla. Finalmente, el papel filtro con el pabalo se coloca sobre la caja Petri para ser impregnado con solución. Este proceso se repite para todas las muestras de suelo y tarda de entre 30 a 45 minutos para que el papel filtro sea impregnado hasta la marca de los 6 cm (ver Fig. 5). Este paso en particular, presentó diversas anomalías en el corrido de la imagen PCC, donde varias muestras no completaron el proceso.



Figura 5: Impregnación de PCC

Una vez que el corrido de la imagen PCC llega a la marca de los 6 cm, se retira de la caja Petri y se le extrae el pabalo. Después, la imagen PCC se cuelga de la orilla y se exponen a media luz (indirectamente de la luz solar) donde pasa un día más con el objetivo de que las coloraciones adquieran una tonalidad óptima para el análisis (ver Fig. 6).



Figura 6: Secado de PCC

Una vez que se completó el trabajo de laboratorio, se

consideraron tres variables para interpretación: el pH, la conductividad eléctrica (CE) y la textura del suelo. El pH se refiere a la acidez o basicidad de una solución que depende de la cantidad de iones de hidrogeno presentes en ella [7]; la CE del suelo se refiere a la capacidad de conducir la corriente eléctrica que depende de la cantidad de iones positivos y negativos que se encuentran en la solución del suelo, por eso la CE de la solución de suelo es un indicador del contenido de sales [8]; finalmente, la textura del suelo se define como la proporción (de las partículas menores a 2 mm de diámetro (arena, arcilla y limo) existentes en los horizontes del suelo.

III. Aprendizaje profundo en imágenes PCC

En este trabajo se utiliza el aprendizaje profundo, un paradigma del aprendizaje máquina, para la detección automática de características de las imágenes de cromatogramas circulares de Pfeiffer y la clasificación automática de las mismas con base en su tipo.

La implementación del aprendizaje profundo se llevó a cabo con el servicio de la nube de Google**, que permite generar máquinas virtuales con diversas herramientas de software adecuadas para el procesamiento masivo de datos (big data) y de aprendizaje máquina. Se utilizó el lenguaje de programación Python versión 3.7, el cual contiene una gran diversidad de librerías que implementan la mayoría de los métodos y algoritmos del estado del arte en el área de aprendizaje máquina. En particular se utilizó tensorflow, una plataforma de código abierto especializada en el aprendizaje profundo en sistemas de gran escala y en entornos heterogéneos [9]. Tensorflow utiliza un enfoque en entrenamiento e inferencia en redes neuronales profundas, que se procesan a través de grafos de flujo de datos para representar el cálculo, el estado compartido y las operaciones que mutan de un estado a otro en equipos multiprocesadores y distribuidos***.

Existe una gran variedad de diseños de redes neuronales profundas enfocadas al reconocimiento en imágenes, tales como AlexNet [10], VGGNet [11], GoogLeNet [12], ResNET [13], entre otras. En el presente trabajo se utilizó la red neuronal profunda MobileNetV2 [14], una arquitectura “ligera” diseñada principalmente para su uso en dispositivos móviles y en entornos con recursos computacionalmente limitados.

III.1. Clasificación de las imágenes PCC

El presente trabajo trata del problema de clasificación de imágenes PCC para cuatro casos específicos: acidez, conductividad eléctrica, textura, y salud general del suelo.

** Google Cloud Platform. <https://cloud.google.com/>

***Tensorflow, an end-to-end open source machine learning platform. <https://www.tensorflow.org>

Los tres primeros, respectivamente, representan las características del suelo provistas por el PCC. En la clasificación de la acidez, se consideró el pH del suelo etiquetado como: pH bueno, pH regular y pH malo. Respecto a la conductividad eléctrica del suelo, se crearon las etiquetas: menor o igual a 0.1, desde 0.2 hasta 0.3 y mayor o igual a 0.4. En relación a la textura del suelo, esta se dividió en dos categorías: franco-arcilloso y franco-arenoso. Por último, para la clasificación donde se evalúa la salud general del suelo, se dividió el conjunto de imágenes en tres categorías: verde, amarillo y rojo. Es importante enfatizar que estas características y sus etiquetas se basan en resultados químicos y de análisis de laboratorio interpretado por investigadores del Instituto Tecnológico de Tlajomulco. Para resolver este problema de clasificación se utiliza MobileNetV2, la red neuronal convolucional pequeña y eficiente mencionada anteriormente. Esta red neuronal ya está especializada en la detección de características y representaciones de objetos en imágenes digitales. Lo que aquí se propone es un proceso de reentrenamiento, que consiste en agregar nuevo conocimiento y experiencias a la red neuronal.

Para la clasificación de acidez, conductividad eléctrica, textura, y salud general del suelo de las imágenes PCC, se utiliza la red MobileNetV2 aplicada a diversos subconjuntos con base en la técnica de validación cruzada k -fold. En esta validación, se divide el conjunto de entrenamiento en k subconjuntos uniformes. Después, se prueba secuencialmente cada subconjunto utilizando el clasificador entrenado en los subconjuntos restantes ($k - 1$). Por lo tanto, cada instancia del conjunto de entrenamiento se predice una vez, por lo que la precisión de la validación cruzada es el porcentaje de datos que se clasifican correctamente. La decisión de elección de esta técnica de validación sobre otras (como hold-out o leave-one-out) se debe principalmente a la restricción en el tamaño del conjunto de datos y con el objetivo de disminuir la variabilidad en el entrenamiento de los modelos de clasificación.

En el presente trabajo, para la implementación de la validación cruzada se utiliza el método `StratifiedKFold`, donde $k = 15$. Se eligió este valor de k debido a que en algunos trabajos se ha demostrado empíricamente ser un valor adecuado para algunas aplicaciones (ej., [15, 16, 17]). Así, se consideran 14 subconjuntos para entrenar el modelo de aprendizaje y se deja el último subconjunto para validación; todo esto se realiza 15 veces, cada una de ellas por cada subconjunto del conjunto inicial.

III.2. Métricas de desempeño

Existe una variedad de métricas para la evaluación del desempeño de un modelo de aprendizaje. Generalmente estas métricas se obtienen a partir de una matriz de confusión, una tabla que describe el rendimiento del modelo

con cuatro términos de predicción: verdaderos positivos (TP, por sus siglas en inglés), verdaderos negativos (TN, por sus siglas en inglés), falsos positivos (FP, por sus siglas en inglés) y falsos negativos (FN, por sus siglas en inglés). En el presente trabajo, se utilizan las métricas de pérdida de entropía cruzada, exactitud, precisión, sensibilidad y el puntaje F1 para cada uno de los 15 entrenamientos. La pérdida de entropía cruzada (CELoss) mide el desempeño de un modelo de clasificación cuya salida es un valor de probabilidad (entre 0 y 1). Este valor incrementa conforme la probabilidad de predicción diverge de la etiqueta actual. Por otro lado, la métrica de exactitud (Acc) es la razón entre el número de predicciones correctas y el número total de muestras de entrada; i.e., $(TP+TN)/(TP+TN+FP+FN)$. La métrica de precisión (Prec) representa el número de verdaderos positivos que son realmente positivos en comparación con el número total de valores positivos predichos; i.e., $Prec=TP/(TP+FP)$. La métrica de exhaustividad/sensibilidad (Recall) muestra la cantidad de verdaderos positivos que el modelo clasifica en función del número total de valores positivos; i.e., $Recall = TP/(TP+FN)$. Finalmente, el puntaje F1 combina la precisión y la sensibilidad para calcular un puntaje que sirva de compromiso entre ambas métricas; i.e., $F1=2 (Prec \cdot Recall) / (Prec+Recall)$. El F1 tiene un rango de $[0, 1]$, que indica qué tan preciso es su clasificador (cuántas instancias clasifica correctamente), así como qué tan robusto es.

IV. Resultados experimentales

En esta sección se presentan los resultados experimentales al aplicar los modelos de aprendizaje profundo al conjunto de imágenes de cromatogramas circulares de Pfeiffer. Estos modelos se implementaron para resolver los siguientes cuatro problemas de clasificación: acidez (pH), conductividad eléctrica (CE), textura, y salud general del suelo.

Las Tablas 1-4 muestran los resultados de aplicar los modelos propuestos a los conjuntos de validación para cada problema de clasificación. Se utilizan las métricas CEloss, Acc, Prec, Recall y F1 de la Sección III.2 para los 15 pliegues (o folds) de la validación cruzada. La última fila de cada una de las tablas muestra el promedio (Avg) de esos pliegues. Se observa que los mejores resultados se obtienen en la clasificación de la textura de suelos con un F1 del 79 %, mientras que en las otras clasificaciones ninguna obtiene resultados mayores al 37 %. Esto se debe que en la clasificación de texturas (Tabla 3) se consideran únicamente dos clases franco-arcilloso y franco-arenoso, por lo que puede entenderse como un problema más sencillo de resolver. Por otro lado, los otros clasificadores constan de tres clases (ver Sección III.1), lo cual representa un reto mayor que el caso anterior.

Tabla 1: Resultados del modelo para la clasificación de acidez de suelos

Fold	CELoss	Acc	Prec	Recall	F1
1	1.4579	0.1667	0.2	0.1667	0.1818
2	1.0701	0.3333	0	0	0
3	1.1132	0.3333	0	0	0
4	1.3841	0.3333	0.2	0.1667	0.1818
5	1.055	0.3333	0	0	0
6	1.209	0.3333	0.5	0.3333	0.4
7	1.504	0.5	0.6	0.5	0.5455
8	1.237	0.3333	0.3333	0.1667	0.2222
9	1.1121	0.3333	0	0	0
10	1.1036	0.1667	0	0	0
11	1.1295	0.1667	0	0	0
12	1.6186	0.3333	0.4	0.3333	0.3636
13	1.3684	0.3333	0.3333	0.1667	0.2222
14	1.1011	0.3333	0	0	0
15	1.2681	0.3333	0.3333	0.1667	0.2222
Avg	1.2488	0.3111	0.1933	0.1333	0.156

Tabla 2: Resultados del modelo para la clasificación de conductividad eléctrica de suelos

Fold	CELoss	Acc	Prec	Recall	F1
1	1.2237	0.5	0	0	0
2	1.0907	0.3333	0.5	0.1667	0.25
3	1.2086	0.1667	0	0	0
4	2.324	0.3333	0.4	0.3333	0.3636
5	1.098	0.3333	0	0	0
6	2.0592	0.3333	0.3333	0.3333	0.3333
7	1.2002	0.3333	0.25	0.1667	0.2
8	1.0981	0.3333	0	0	0
9	1.1326	0.5	0.25	0.1667	0.2
10	1.1	0.3333	0	0	0
11	1.1088	0.5	0.6	0.5	0.5455
12	1.3774	0.3333	0	0	0
13	1.4023	0.3333	0.2	0.1667	0.1818
14	1.1011	0.5	0	0	0
15	1.3015	0.3333	0.4	0.3333	0.3636
Avg	1.3217	0.3667	0.1956	0.1444	0.1625

Adicionalmente, los resultados de las tablas también pueden reflejar un nivel de correlación entre las imágenes PCC y las características químicas que se buscan clasificar.

En particular, los resultados de la Tabla 1 concuerdan con lo presentado en [4], donde se estudió la correlación entre la evaluación de patrones PCC con mediciones de zonas del PCC, interpretaciones visuales y el análisis de texturas versus el análisis químico de suelos.

También allí se encontró una correlación no significativa del pH con imágenes PCC. Los resultados aquí presentados sugieren que esto también ocurre para la

Tabla 3: Resultados del modelo para la clasificación de textura de suelos

Fold	CELoss	Acc	Prec	Recall	F1
1	0.6011	0.8333	0.8333	0.8333	0.8333
2	0.6642	0.8333	0.8333	0.8333	0.8333
3	0.6032	0.8333	0.8333	0.8333	0.8333
4	1.1728	0.1667	0.1667	0.1667	0.1667
5	1.6862	0.8333	0.8333	0.8333	0.8333
6	1.167	0.8333	0.8333	0.8333	0.8333
7	0.508	0.8333	0.8333	0.8333	0.8333
8	0.5558	0.8333	0.8333	0.8333	0.8333
9	0.7985	0.8333	0.8333	0.8333	0.8333
10	0.4919	0.8333	0.8333	0.8333	0.8333
11	0.5251	0.8333	0.8333	0.8333	0.8333
12	0.5205	0.8333	0.8333	0.8333	0.8333
13	0.4278	0.8333	0.8333	0.8333	0.8333
14	0.6622	0.8333	0.8333	0.8333	0.8333
15	2.2533	0.8333	0.8333	0.8333	0.8333
Avg	0.8425	0.7889	0.7889	0.7889	0.7889

Tabla 4: Resultados del modelo para la clasificación de la salud general de suelos

Fold	CELoss	Acc	Prec	Recall	F1
1	1.8307	0.4	0.4	0.4	0.4
2	2.1815	0.5	0.5556	0.5	0.5263
3	1.8185	0.3	0.2222	0.2	0.2105
4	1.1594	0.3	0.5	0.3	0.375
5	3.4925	0.3	0.375	0.3	0.3333
6	0.901	0.6	0.6667	0.2	0.3077
7	2.1746	0.5	0.4444	0.4	0.4211
8	1.6483	0.3	0.3333	0.3	0.3158
9	1.0075	0.4	0.5	0.4	0.4444
10	1.3289	0.4	0.4444	0.4	0.4211
11	2.595	0.4	0.4444	0.4	0.4211
12	1.439	0.4	0.3333	0.3	0.3158
13	6.2355	0.2	0.2	0.2	0.2
14	2.5826	0.4	0.4	0.4	0.4
15	2.7986	0.4	0.4	0.4	0.4
Avg	2.2129	0.3867	0.4146	0.34	0.3661

conductividad eléctrica (ver Tabla 2).

Finalmente, los resultados de la Tabla 4 muestran resultados satisfactorios para el problema de multclasificación del estado de salud general de suelos. Como una primera aproximación al problema, el modelo es capaz de definir si la imagen PCC refleja un estado de salud malo, regular o bueno. Todos los scripts se ejecutaron en el sistema operativo GNU/Linux Ubuntu 18.04, con todas las herramientas y librerías de software libre.

V. Conclusiones

El presente trabajo trata de la implementación de modelos de aprendizaje profundo para la clasificación de tipos de suelos utilizando imágenes de PCC. Una de las aportaciones principales, con apoyo de los investigadores del Instituto Tecnológico de Tlajomulco, es la conformación de la primera base de datos pública de imágenes PCC.

Los problemas de clasificación que se trataron son cuatro: acidez (pH), conductividad eléctrica (CE), textura, y salud general del suelo. Los resultados son satisfactorios en general, y en particular el problema de acidez corrobora los resultados de una investigación reciente acerca de la correlación no significativa del pH con las imágenes PCC. Los resultados obtenidos dan certeza a la hipótesis: “el uso de técnicas de aprendizaje máquina, y en especial del aprendizaje profundo, permiten la clasificación de tipos de suelos para su análisis a través de imágenes PCC”. Cabe resaltar que esta investigación, es el primer trabajo académico que utiliza herramientas de aprendizaje máquina en imágenes PCC para el análisis de suelos.

Como trabajo futuro, queda el seguir robusteciendo la base de datos con nuevas imágenes PCC que originen nuevas investigaciones en el área. Por otro lado, también es interesante seguir trabajando con la especificación de otras características que sean relevantes para el análisis de suelos, y con ello propiciar el desarrollo de software especializado para la industria agropecuaria.

Agradecimientos

El Dr. Fajardo agradece al Tecnológico Nacional de México por su apoyo para la realización de su Año Sabático bajo el Programa de Investigación Científica, Tecnológica o Educativa (No. de autorización AS-2-027/2021). También, los autores agradecen a la Dra. María de Jesús Ramírez Ramírez y al M.C. Osvaldo Amador Camacho, ambos adscritos al Instituto Tecnológico de Tlajomulco, por su apoyo para el trabajo de laboratorio e interpretación de los PCC.

Referencias

- [1] J Benton Jones Jr. *Soil analysis handbook of reference methods*. CRC Press, 1999.
- [2] Jairo Restrepo Rivera y Sebastiao Pinheiro. *Cromatografía imágenes de vida y destrucción del suelo*. COAS Ediciones, 2011.

- [3] Deepak Khemani, Minu Mary Joseph y Saritha Variganti. «Case Based Interpretation of Soil Chromatograms». En: *Advances in Case-Based Reasoning*. Ed. por Klaus-Dieter Althoff y col. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, págs. 587-599. ISBN: 978-3-540-85502-6.
- [4] Maria Olga Kokornaczyk y col. «Analysis of soils by means of Pfeiffer's circular chromatography test and comparison to chemical analysis results». En: *Biological Agriculture & Horticulture* 33.3 (2017), págs. 143-157. DOI: 10.1080/01448765.2016.1214889.
- [5] Ehrenfried Pfeiffer. *Chromatography applied to quality testing*. SteinerBooks, 1984.
- [6] Martin R Carter y Edward Gerard Gregorich. *Soil sampling and methods of analysis*. CRC press, 2007.
- [7] Rüdiger Albin. «Determinación del pH en diversas especies de los renovales de la provincia de Valdivia». En: *Bosque* 1.1 (1975), págs. 3-5.
- [8] Matías Bosch Mayol y col. «Relación entre la variabilidad espacial de la conductividad eléctrica y el contenido de sodio del suelo.» En: *Ciencia del suelo* 30.2 (2012).
- [9] Martín Abadi y col. «TensorFlow: A System for Large-Scale Machine Learning». En: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. Savannah, GA: USENIX Association, nov. de 2016, págs. 265-283. ISBN: 978-1-931971-33-1. URL: <https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi>.
- [10] Alex Krizhevsky, Ilya Sutskever y Geoffrey E. Hinton. «ImageNet Classification with Deep Convolutional Neural Networks». En: *Commun. ACM* 60.6 (mayo de 2017), págs. 84-90. ISSN: 0001-0782. DOI: 10.1145/3065386. URL: <https://doi.org/10.1145/3065386>.
- [11] Karen Simonyan y Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. arXiv: 1409.1556 [cs.CV].
- [12] Christian Szegedy y col. «Going Deeper With Convolutions». En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Jun. de 2015.
- [13] Kaiming He y col. «Deep Residual Learning for Image Recognition». En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Jun. de 2016.
- [14] Mark Sandler y col. «MobileNetV2: Inverted Residuals and Linear Bottlenecks». En: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Jun. de 2018.

- [15] S. Franchini y col. «Evaluation of a Support Vector Machine Based Method for Crohn's Disease Classification». En: *Neural Approaches to Dynamics of Signal Exchanges*. Ed. por Anna Esposito y col. Singapore: Springer Singapore, 2020, págs. 313-327. ISBN: 978-981-13-8950-4. DOI: 10.1007/978-981-13-8950-4_29.
- [16] Gopi Battineni y col. «Comparative Machine-Learning Approach: A Follow-Up Study on Type 2 Diabetes Predictions by Cross-Validation Methods». En: *Machines* 7.4 (2019). ISSN: 2075-1702. URL: <https://www.mdpi.com/2075-1702/7/4/74>.
- [17] Probir Kumar Ojha y Kunal Roy. «Development of a robust and validated 2D-QSPR model for sweetness potency of diverse functional organic molecules». En: *Food and Chemical Toxicology* 112 (2018), págs. 551-562. ISSN: 0278-6915. DOI: 10.1016/j.fct.2017.03.043.