

Predicción de propiedades petrofísicas básicas a partir de información de tomografía
usando un modelo de aprendizaje profundo

Erick Camilo Morris Duarte

Trabajo de Grado para Optar el Título de Ingeniero de petróleos

Director

Andrés Felipe Ortiz Meneses

M.Sc. en Ingeniería de Hidrocarburos

Universidad Industrial de Santander
Facultad de Ingenierías Físicoquímicas
Escuela de Ingeniería de petróleos
Bucaramanga

2021

Dedicatoria

Me permito empezar esta dedicatoria mencionando a mis padres, Verónica Duarte e Iduar Morris, quienes les debo mi vida y me han dado el amor, la fortaleza y el apoyo para continuar cuando he estado a punto de caer. A ustedes quienes han sabido formarme con buenos sentimientos, hábitos y valores, que me han permitido salir adelante.

A mis maestros, gracias por su tiempo, por su apoyo así como por la sabiduría que me transmitieron en el desarrollo de mi formación profesional, al ingeniero Andrés Ortiz, por haber guiado el desarrollo de este trabajo y llegar a la culminación de este.

Una dedicatoria especial a Camila Caicedo por brindarme su amor, comprensión y compañía, inspirándome a ser cada vez mejor persona y a esforzarme por mis metas y sueños.

Erick Camilo Morris Duarte

Agradecimientos

Al director del proyecto, Ingeniero Andrés Felipe Ortíz Meneses, por su apoyo, ayuda, motivación y disposición durante el desarrollo del proyecto, siendo un pilar importante para el éxito de este.

A mi familia y amigos, por su apoyo incondicional que de alguna u otra manera contribuyo a la realización de este proyecto.

Gracias Totales.

Tabla de Contenido

	Pág.
Introducción	11
1. Objetivos	12
1.1 Objetivo general	12
1.2 Objetivos específicos	12
2 Marco referencial	13
2.1 Tomografía computarizada	13
2.2 Aprendizaje automático	14
2.2.1 Entrenamiento	15
2.2.2 Pérdida	15
2.3 Aprendizaje profundo	16
2.3.1 Perceptrones multicapa	17
2.3.2 Redes Neuronales Convolucionales (RNCs)	20
2.4 Aprendizaje profundo y automático en la industrial de petróleo	22
2.5 Relaciones teóricas entre las mediciones de tomografía y las propiedades petrofísicas básicas	23
3. Diseño del modelo de aprendizaje profundo	27
3.1 Preprocesamiento de la información numérica	27
3.2 Preprocesamiento de las imágenes de tomografía computarizada	29
3.3 Diseño y parámetros para el modelo de aprendizaje profundo	30

4. Resultados	37
4.1 Análisis de métrica del modelo de porosidad	37
4.2 Análisis de métrica del modelo de permeabilidad	38
4.3 Predicción de propiedades petrofísicas	39
4.4 Discusión y análisis.....	41
5. Conclusiones	43
6. Recomendaciones	44
Referencias bibliográficas.....	45
Apéndice A	50
Apéndice B.....	51
Apéndice C script para el preprocesamiento de imágenes.....	52
Apéndice D script del preprocesamiento de datos numéricos	53
Apéndice E script del diseño del modelo de aprendizaje profundo.....	54
Apéndice F script para el entrenamiento y estudio de la métrica del modelo de ap.....	56

Lista de Tablas

	Pág.
Tabla 1 <i>Conjunto de datos numéricos para la red neuronal.</i>	28
Tabla 2 <i>Información numérica preprocesada para entrenamiento, prueba y validación del modelo</i>	29
Tabla 3 <i>Arquitectura completa de modelo</i>	34
Tabla 4 <i>Parámetros para el entrenamiento del modelo de aprendizaje profundo</i>	35
Tabla 5 <i>Conjunto de datos numéricos pertenecientes al segundo set de información.</i>	36

Lista de Figuras

	Pág.
Figura 1 <i>Pérdida alta en el modelo de la izquierda; pérdida baja en el modelo de la derecha.</i> .	16
Figura 2 <i>Propagación directa de la información a través de una neurona.</i>	17
Figura 3 <i>Esquema de un PMC</i>	19
Figura 4 <i>Esquema de una RNCs</i>	21
Figura 5 <i>Sección a escala de grises de una imagen para entrenamiento, validación y prueba del modelo</i>	29
Figura 6 <i>Estandarización de la densidad de píxel de la imagen</i>	30
Figura 7 <i>Arquitectura simplificada del modelo con los atributos de entrada y datos de salida.</i> ..	32
Figura 8 <i>Comportamiento de la RCN con la capa de maxPooling.</i>	33
Figura 9 <i>Métricas y gráfica de regresión lineal del modelo de porosidad</i>	37
Figura 10 <i>Métrica y gráfica de regresión lineal del modelo de permeabilidad</i>	38
Figura 11 <i>Predicciones de permeabilidad y porosidad para 4 secciones de núcleo del primer conjunto de datos.</i>	39
Figura 12 <i>Predicciones de permeabilidad y porosidad para 4 secciones de núcleo del segundo conjunto de datos.</i>	40

Lista de Apéndices

Apéndice A La canalización del aprendizaje profundo	50
Apéndice B Script para seccionar las imágenes para entrenamiento.....	51
Apéndice C Script para el preprocesamiento de imágenes	52
Apéndice D Script del preprocesamiento de datos numéricos.....	53
Apéndice E Script del diseño del modelo de aprendizaje profundo	54
Apéndice F script para el entrenamiento y estudio de la métrica del modelo de AP	56

Resumen

Título: Predicción de propiedades petrofísicas básicas a partir de información de tomografía usando un modelo de aprendizaje profundo*

Autor: Erick Camilo Morris Duarte**

Palabras Clave: TC, Imágenes de roca, Registros de pozo, Inteligencia artificial, Red neuronal convolucional.

Descripción: Se presenta un estudio de investigación destinado a probar la aplicabilidad de técnicas de aprendizaje automático y profundo para la predicción de propiedades petrofísicas básicas a partir de información e imágenes de tomografía. Se estableció la relación teórica entre estas propiedades petrofísicas y las imágenes de tomografía mediante un estudio literario. Se preparó un set de datos para entrenamiento, validación y prueba que contiene: Imágenes de tomografía de la sección transversal XZ y datos de tomografía como registros de densidad y factor fotoeléctrico, así como los datos de porosidad y permeabilidad medidos en laboratorio (Porosidad Boyle y permeabilidad al aire), tomado de los núcleos de perforación estudiados, y de la información de tomografía computarizada mencionada, para que las redes neuronales puedan predecir porosidad y permeabilidad a cada punto de profundidad, incluyendo los intervalos donde la información de laboratorio no fue tomada. Se utilizaron dos algoritmos principales de aprendizaje profundo, perceptrón multicapa y redes neuronales convolucionales 2D, como características para los datos de entrada. Se comparó el poder predictivo del modelo desarrollado con los datos medidos en laboratorio. Se estudió la métrica de los modelos desarrollados. Los resultados demuestran la aplicabilidad del aprendizaje automático y profundo en la predicción de propiedades petrofísicas básicas basada en imágenes y datos de tomografía, los modelos desarrollados son excepcionalmente buenos a una alta variedad de datos.

* Trabajo de Grado

** Facultad de Ingenierías fisicoquímicas. Escuela de Ingeniería de petróleos. Director: Andrés Felipe Ortíz Meneses. M.Sc Ingeniería de hidrocarburos.

Abstract

Title: Prediction of basic petrophysical properties from tomography information using a deep learning model*

Author: Erick Camilo Morris Duarte**

Key Words: CT, Artificial Intelligence, Rock Images, Well logs, Convolutional Neural Network.

Description: A research study is presented to test the applicability of deep and machine learning techniques for the prediction of basic petrophysical properties with information and tomography images. The theoretical relationship between these petrophysical properties and tomography images was established through a literary study. A data set was prepared for training, validation and testing that it contains. XZ cross-section tomography images and tomography data such as density and photoelectric factor records, as well as laboratory-measured porosity and permeability data (Boyle porosity and air permeability), taken from studied drill cores, and from the aforementioned computed tomography information, so that neural networks can predict porosity and permeability at each depth point, including the intervals where the laboratory information was not taken. Two main deep learning algorithms, multilayer perceptrons and 2D convolutional neural networks, were used as characteristics for the input data. The predictive power of the developed model was compared with the data measured in the laboratory. The metrics of the developed models were studied. The results show the applicability of machine and deep learning in the prediction of petrophysical properties based on images and tomography data, the models developed in exceptionally good to a high variety of data.

* Degree Work

** Faculty of Physicochemical Engineering. School of Petroleum Engineering. Director: Andrés Felipe Ortíz Meneses. M.Sc Hydrocarbon Engineering.

Introducción

La tomografía computarizada de rayos X es una herramienta útil en la caracterización de rocas y que permite observar el interior de la roca y generar versiones digitales en 3D de los núcleos de perforación incluso cuando aún se encuentran conservados.

La tomografía proporciona imágenes a una muy alta resolución de las propiedades petrofísicas básicas de la roca, sin embargo, hay que realizar una transformación de estas imágenes en datos cuantitativos que se puedan utilizar e integrar en información para la caracterización de yacimientos.

Uno de los enfoques más prometedores para manejar varios desafíos de ingeniería es la inteligencia artificial (IA) (Brunette et al., 2009), esta artificial ha sido aplicada para desarrollar varios problemas en la ingeniería de petróleo (Cuddy & Putnam, 1998).

En este documento, el objetivo es probar la aplicabilidad del aprendizaje profundo para proporcionar una evaluación rápida de cortes 2D de imágenes de rocas y de datos de densidad aparente (RHOB) y factor fotoeléctrico (PEF) que se calculan a partir de mediciones de tomografía computarizada (Ortiz Meneses et al., 2015; Siddiqui & Khamees, 2004), estimando propiedades correlacionadas como la permeabilidad y porosidad

1. Objetivos

1.1 Objetivo General

Predecir las propiedades petrofísicas básicas a partir de datos de tomografía utilizando un modelo de aprendizaje profundo.

1.2 Objetivos Específicos

1. Establecer relaciones teóricas entre las mediciones de tomografía y las propiedades petrofísicas básicas.

2. Diseñar un modelo de aprendizaje profundo y entrenarlo con la relación entre las mediciones petrofísicas básicas y las mediciones de tomografía.

3. Predecir valores de porosidad y permeabilidad mediante el modelo de aprendizaje profundo y de la información de tomografía.

4. Analizar la métrica de las predicciones realizadas por el modelo de aprendizaje profundo con respecto a los obtenidos en mediciones de laboratorio.

2 Marco Referencial

En esta sección se presenta una revisión literaria que respalda y soporta las relaciones teóricas entre la información e imágenes de tomografía y las propiedades petrofísicas básicas, además se exponen los principios y aplicabilidad de la inteligencia artificial en la industria del petróleo y en la ingeniería en general.

2.1 Tomografía computarizada

La tomografía computarizada es una técnica de imagen no destructiva que utiliza tecnología de rayos X y algoritmos de reconstrucción matemática para ver cortes transversales de un objeto (Siddiqui & Khamees, 2004). Su capacidad para inspeccionar la estructura interna de la roca de forma no intrusiva le permite verificar el estado de los núcleos de perforación dentro de su barril de aluminio, para así obtener una reconstrucción digital del estado inicial de las rocas que se conservarán con el tiempo (Ortiz et al., 2020; Withjack et al., 2003).

La tomografía computarizada aprovecha la diferencia que existe en la atenuación producida por diferentes materiales sobre los rayos-X que los atraviesa, de manera que las imágenes obtenidas de los cuerpos estudiados son representaciones gráficas de los diferentes coeficientes de atenuación de éstos (Honarpour et al., 1985; Ortiz Meneses et al., 2015). La transición de la inspección cualitativa a la determinación cuantitativa de datos es una propiedad aprovechable de la información a partir de la tomografía computarizada (A. Kantzas D.F. Marentette K.N. Jha, 1992).

Esta tecnología ha sido empleada en la caracterización de yacimientos, descripción de núcleos, características de la roca, métodos de recobro mejorado, estudios de daño a la formación entre otros (Siddiqui & Khamees, 2004), lo que demuestra su potencial en sus relaciones con las propiedades petrofísicas básicas como la porosidad y permeabilidad.

2.2 Aprendizaje automático

El aprendizaje profundo (AP) es un tipo específico de aprendizaje automático (AA). Para comprender bien el aprendizaje profundo, se debe tener una comprensión de los principios básicos del aprendizaje automático (Goodfellow et al., 2019).

Un algoritmo de AA es un algoritmo que puede aprender de los datos. Pero ¿Qué se entiende por aprendizaje automático? (Mitchel, 1997) Proporciona la definición “Se dice que un algoritmo de aprendizaje automático es un algoritmo que es capaz de mejorar el rendimiento de un programa de computadora en alguna tarea a través de la experiencia”.

Esta definición, aunque sencilla, es abstracta, para hacerla más concreta se presenta un corto ejemplo de un algoritmo de AA simple: regresión lineal. Como el nombre implica la regresión lineal soluciona un problema de regresión. En otras palabras el objetivo es construir un sistema que pueda tomar un dato de entrada x y predecir un dato de salida y (Goodfellow et al., 2019).

$$y = mx + b \quad (1)$$

Donde:

- y es el valor que se intenta predecir.
- m es la pendiente de la línea.
- x es el valor del atributo de entrada.
- b es la intersección en y .

Aquí la tarea es predecir un valor y a nuevos valores desconocidos de x mediante la experiencia de valores x , y ya otorgados. Según las convenciones del AA, la ecuación para un modelo de AA es la ecuación 2 (*Estudio detallado del AA: Regresión lineal, Google.*).

$$y' = w_0 + w_1 x_1 \quad (2)$$

Donde:

- y' es la etiqueta predicha (Resultado deseado).
- w_0 es la ordenada al origen (intersección en y).
- w_1 es la ponderación, también llamada peso del atributo 1. La ponderación es el mismo concepto de “pendiente” m , que se indicó anteriormente.
- x_1 es un atributo (Entrada conocida).

Para predecir un nuevo valor de y' en la ecuación 2 solo se agrega el valor de x_1 al modelo ya entrenado (*Estudio detallado del AA: Regresión lineal, Google.*).

Así se puede llegar a una definición concisa: Los sistemas de AA aprenden cómo combinar entradas para producir predicciones útiles sobre datos nunca vistos.

2.2.1 Entrenamiento

Entrenar un modelo simplemente significa aprender (determinar) valores correctos para todas las ponderaciones y las ordenadas al origen de los ejemplos etiquetados (valores de \mathbf{x} con valor de \mathbf{y} conocido). En un aprendizaje supervisado, el algoritmo de un AA construye un modelo al examinar varios ejemplos e intentar encontrar un modelo que minimice la pérdida. Este proceso se denomina minimización del riesgo empírico (*Estudio detallado del AA: Entrenamiento, Google.*).

2.2.2 Pérdida

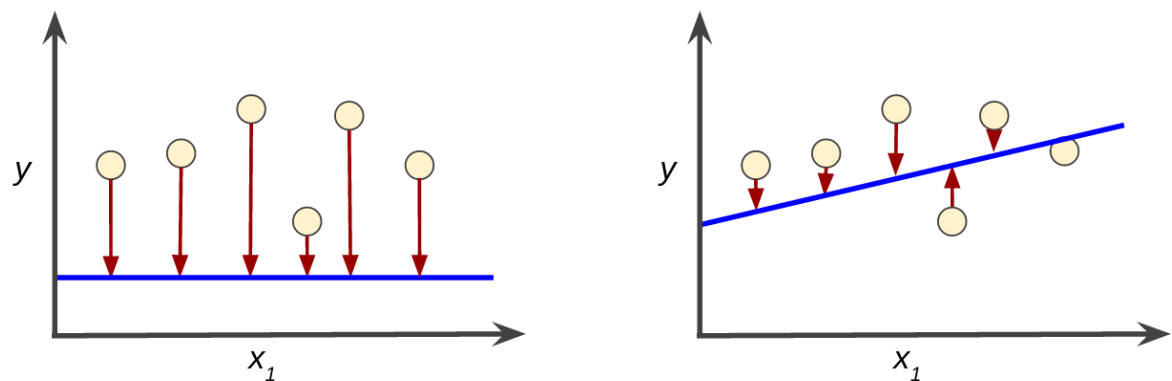
La pérdida es una penalidad por una predicción incorrecta. Esto quiere decir que la pérdida es un número que indica qué tan incorrecta fue la predicción del modelo en un solo ejemplo. Si la predicción del modelo es perfecta, la pérdida es cero; de lo contrario, la pérdida es mayor. El objetivo de entrenar un modelo es encontrar un conjunto de ponderaciones y ordenadas al origen

que, en promedio, tengan pérdidas bajas en todos los ejemplos (*Estudio detallado del AA: Perdida, Google.*). Una representación gráfica sobre la pérdida es expuesta en la figura 1.

Hay muchos estimadores de pérdida (métricas), entre ellos el coeficiente R^2 , error cuadrado medio, error absoluto medio, error máximo, etc. La elección de cual emplear varía de acuerdo con el problema a resolver mediante inteligencia artificial (IA), sin embargo, todos concuerdan en reducir la pérdida.

Figura 1

Representación gráfica de la pérdida.



*Nota. Las flechas rojas representan la pérdida, la línea azul representa las predicciones, se observa que las flechas rojas en la figura de la izquierda son mucho más largas que las de la derecha. Claramente, la línea azul en la figura de la derecha es un modelo de predicción mucho más acertado. Tomado en Enero 2021 de (*Estudio detallado del AA: Perdida, Google.*).*

2.3 Aprendizaje profundo

Los algoritmos de aprendizaje profundo intentan crear modelos jerárquicos profundos que pueden realizar tareas complejas, como identificar un objeto de otro, comprender el habla humana o crear nuevas pinturas basadas en múltiples imágenes de entrada (Pandey et al., 2020), estos

modelos tienen las mismas propiedades que los modelos de AA, sin embargo, aunque parecen ser más robustos y consistentes, estos requieren una mayor cantidad de datos para entrenarse.

Aunque existen varios algoritmos de aprendizaje profundo, para el estudio desarrollado en este documento se discutirán 2 tipos de algoritmos de aprendizaje profundo.

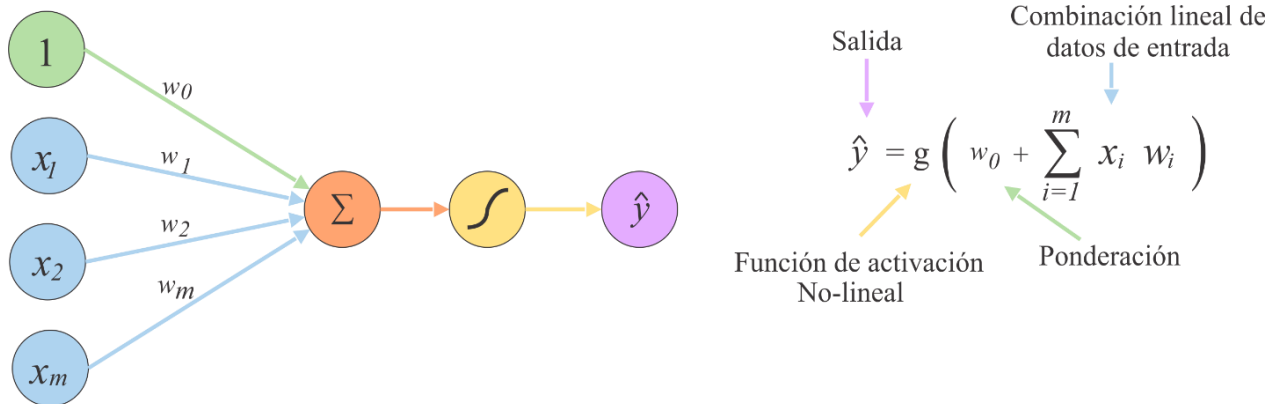
- Perceptrones multicapa (PMC)
- Redes neuronales convolucionales (RNC)

2.3.1 Perceptrones multicapa

Son los bloques fundamentales de una red neuronal, el ejemplo más simple de los PMC es una sola neurona, en una sola capa, el perceptrón; para entender el perceptrón, hay que establecer los parámetros que afectan la propagación directa de la información a través de una neurona, figura 2.

Figura 2

Propagación directa de la información a través de una neurona.



Entrada Ponderaciones Suma No-linealidad Salida

Nota. Modificado de (Amini, 2020) MIT Deep learning 6.S191

En la figura 2, se define un conjunto de datos de entrada de x_1 a x_m , y cada uno de estos datos de entrada tiene una correspondiente ponderación, de w_1 a w_m , luego con cada uno de estos

$$\hat{y} = g \left(w_0 + \sum_{i=1}^m x_i w_i \right)$$

Labels for the equation:

- Salida (purple arrow pointing to \hat{y})
- Combinación lineal de datos de entrada (blue arrow pointing to $\sum_{i=1}^m x_i w_i$)
- Función de activación No-lineal (yellow arrow pointing to g)
- Ponderación (green arrow pointing to w_i)

datos y ponderaciones se realizan una sumatoria de sus productos, a esto se le agrega la ordenada al origen w_0 , y luego se transforma a través de una función de activación no-lineal, la cual no es más que una función que arroja un resultado limitado entre dos valores, finalmente se obtiene el producto final \hat{y} que sería el dato de salida (Amini, 2020).

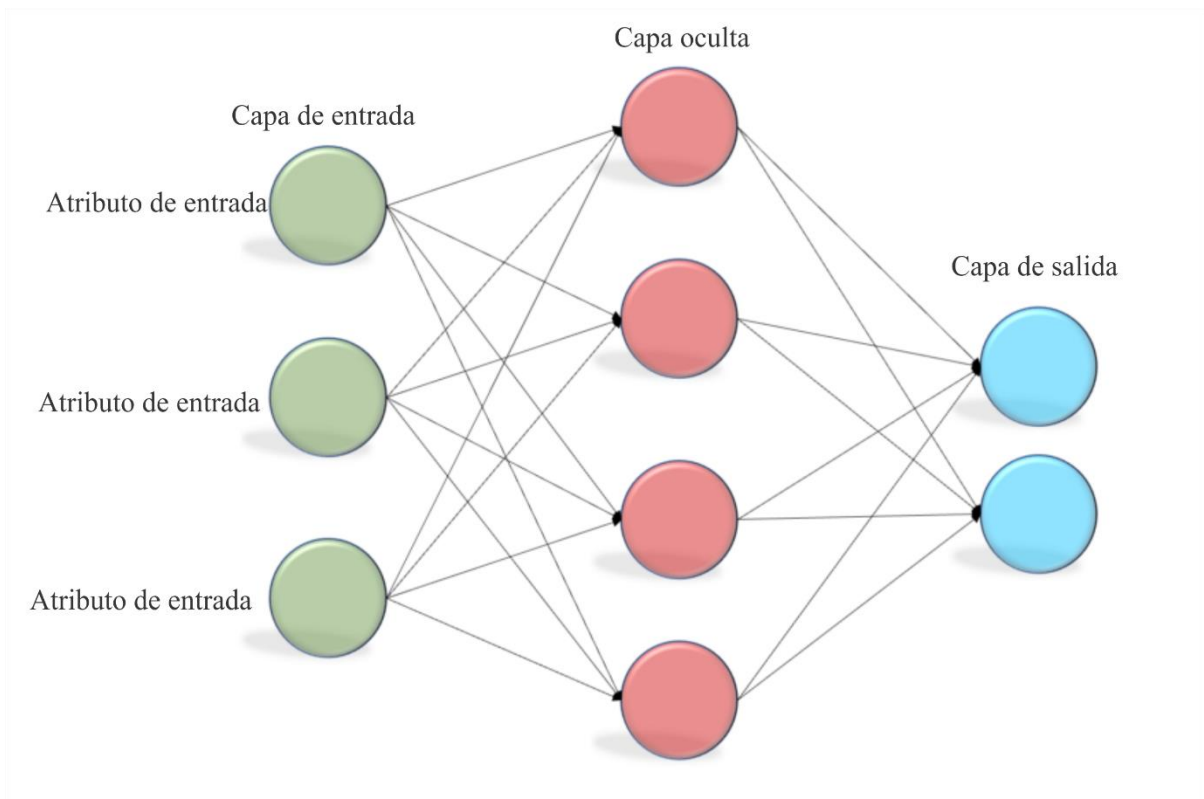
Note que este sumatoria de productos puede ser fácilmente reemplazada por un producto punto entre vectores.

Las transformaciones causadas por la función de activación no-lineal proveen al modelo la capacidad de manejar procesos altamente no lineales, lo cual ocurre con alta frecuencia natural e industrialmente (Pandey et al., 2020).

Un modelo de aprendizaje profundo generalmente está constituido de varias capas con varias neuronas, a esto se le conoce como perceptrones multicapa, una descripción esquemática de los PMC se muestra en la figura 3.

Figura 3

Esquema de un PMC



Nota. El PMC puede contener una cantidad n de capas ocultas, sin embargo, solo 1 capa de entrada y 1 de salida. Modificado de (Mohanty, 2021).

Además del perceptrón, un PMC consta de las capas que se ilustran en la figura 3.

- Capa de entrada: Esta capa contiene nodos, que facilitan la entrada de características a la red neuronal, aplican transformaciones no lineales.
- Capa oculta: Esta capa contienen neuronas o nodos con una función de activación no lineal. La capa oculta está representada por una matriz de ponderaciones (w_m) y un vector de la ordenada al origen (w_0). Las dimensiones de la matriz w y el vector w_0 dependen del número de nodos en las capas. Por ejemplo, la capa de entrada, que se conecta con la entrada m (matriz de datos de entrada x) y transmite valores a n nodos

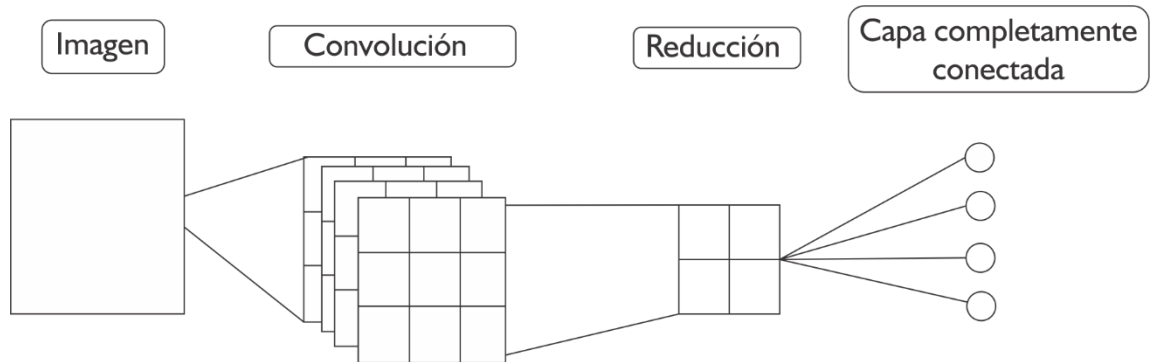
en la primera capa oculta, tiene una matriz de peso con dimensión $m \times n$. Además, el vector de la ordenada al origen w_0 tiene n elementos. La salida de esta capa es $g(wx + w_0)$, donde g es la función de activación (Pandey et al., 2020).

- Capa de salida: La capa de salida se conecta con la última capa oculta y asegura que la red neuronal proporcione una salida consistente con la formulación del problema. Por ejemplo, en un problema con salida de regresión no requiere función de activación.

2.3.2 Redes Neuronales Convolucionales (RNCs)

Las RNCs son un tipo especializado de redes neuronales para procesamiento de datos que tienen una topología conocida en forma de malla. Ejemplos incluyen información de históricos (datos de serie de tiempo), los cuales se pueden considerar como una malla 1D que toma muestras a intervalos de tiempo regulares, y datos de imágenes, que se consideran como una malla 2D de píxeles (Goodfellow et al., 2019). Las RNCs son redes neuronales que usan una operación matemática llamada convolución en lugar de la multiplicación general de matrices (Goodfellow et al., 2019).

Una descripción esquemática de las RNCs se encuentra en la figura 4

Figura 4*Esquema de una RNCs*

Nota. La imagen muestra un esquema básico de una RNCs. Autor

Se describe cada proceso de acuerdo con el esquema mostrado en la figura 4.

- **Imagen:** Es la imagen de entrada, los parámetros para tener en cuenta para la configuración de la capa de entrada en este caso no es el número de atributos de entrada, en su lugar es la forma de la imagen, es decir, el alto, ancho y cantidad de canales de color (Por ejemplo, una imagen RGB tendrá 3 canales de color).
- **Convolución:** En esta capa la imagen pasa a través de N filtros, estos filtros son matrices cuadradas de ponderaciones, lo que genera mapas de atributos; a su vez, estas capas se encargan de realizar la función de activación no lineal.
- **Reducción:** Consiste en reducir una matriz (o matrices) creada por una capa convolucional anterior a una matriz más pequeña. Por lo general, la reducción implica tomar el valor máximo o promedio en el área a ser reducida (Glosario sobre aprendizaje automatico, Google).
- **Capa completamente conectada:** Esta capa se encarga de transformar todos estos parámetros en una capa oculta, para posteriormente transformarse en una capa de salida.

2.4 Aprendizaje profundo y automático en la industrial de petróleo

La industria del petróleo ha usado el aprendizaje automático por décadas. Sin embargo, es posible que se le haya denominado inteligencia artificial (IA) o con otros nombres alternativos (Pandey et al., 2020).

La IA ha sido empleada para atacar varios problemas en la ingeniería de petróleo (Kohli & Arora, 2014). La meta de estas aplicaciones está enfocada en caracterización de yacimientos; como la predicción de permeabilidad a partir de registros de pozo (Kohli & Arora, 2014), no obstante, su aplicabilidad va mucho más allá.

En modelamiento geofísico las RNCs han demostrado ser útiles para la interpretación automática de fallas (Li et al., 2019), interpretación sísmica (Wu et al., 2019), entre otras aplicaciones.

El modelamiento geológico se ha visto beneficiado de la inteligencia artificial, Por ejemplo, en la clasificación de facies (Hall, 2016).

Los ingenieros de yacimientos tienen acceso a casi todos los datos utilizados por los geocientíficos. También tienen datos del historial operativo y de producción de hidrocarburos del campo del que son responsables. Utilizan una variedad diversa de datos para desarrollar un plan para la explotación económicamente óptima del petróleo y el gas del yacimiento y de todo el campo petrolífero en general (Pandey et al., 2020). Unos cuantos ejemplos del uso del AA y AP en la ingeniería de yacimientos van desde la planificación de desarrollo de campos (Bhattacharya & Nikolaou, 2011), pasando por ajuste histórico asistido (Shahkarami et al., 2014), hasta proyección de producción (Sun et al., 2018) y estimación de reservas (Mohaghegh et al., 2011).

En la caracterización de yacimientos se han desarrollado estudios que buscan predecir porosidades a partir de registros generados por información de tomografía (Ortiz et al., 2020),

distribución y tamaño de poro a partir de imágenes de microtomografía computacional (Alqahtani et al., 2018; Andrä et al., 2013; Sudakov et al., 2019), lo que ha demostrado su eficacia en este campo.

El sector operativo, perforación y completamiento de pozos, ha hecho uso de la inteligencia artificial en diversas ocasiones, mejorando la productividad y seguridad en estas operaciones; el estudio de (Unrau, 2017) demuestra que estos algoritmos pueden emplearse para detectar tempranamente eventos que pongan en peligro el control del pozo, como las patadas de pozo. Debido a las altas tarifas de alquiler de equipos de perforación, disminuir el tiempo no productivo es otra tarea constante (Pandey et al., 2020), avances en inteligencia artificial para predecir las pegas de tubería han ido desarrollándose (Elmousalami & Elaskary, 2020), los cuales permitirán reducir estos tiempos no productivos, entre otras aplicaciones de la inteligencia artificial.

Muchos autores usan técnicas de inteligencia artificial para asistir la ingeniería de producción, por ejemplo, en la optimización de la producción (Saputelli et al., 2005), identificación de candidatos para workover (Zangl et al., 2016), entre otros.

En general, la inteligencia artificial tiene la capacidad de proveer soluciones valiosas a problemas en la ingeniería de petróleo.

2.5 Relaciones teóricas entre las mediciones de tomografía y las propiedades petrofísicas básicas

Se han realizado muchos estudios utilizando tomografía computacional (TC) para análisis de núcleos y pruebas de flujo (Withjack et al., 2003).

La tomografía computarizada divide los cuerpos en una matriz de coeficientes donde cada elemento se llama vóxel (que es equivalente a un píxel 3D). Se estima un coeficiente de atenuación de rayos X único para cada vóxel utilizando la escala de Hounsfield, lo que significa que las

atenuaciones de vóxel se dan en números de TC. Los coeficientes de atenuación de rayos X estimados por la tomografía computarizada están relacionados con la densidad y el factor de número atómico efectivo (que está directamente relacionado con el PEF), por lo que RHOB y PEF se pueden estimar a partir de los números de TC de las imágenes (Ortiz Meneses et al., 2015; Siddiqui & Khamees, 2004).

Como los números de la tomografía computarizada y la densidad están directamente relacionados a la porosidad, varios métodos se han propuestos para estimar esta propiedad a partir de la información de tomografía. Un método para estimar la porosidad es utilizando el enfoque por tomografía computarizada de energía dual (Akin et al., 1996).

En la exploración por TC de energía dual, el objeto se explora dos veces en la misma ubicación, utilizando una configuración de alta y baja energía. Al seleccionar los ajustes de alta y baja energía, se pueden aprovechar las dos interacciones predominantes de los rayos X con la materia, a saber, la absorción fotoeléctrica (predominante a bajas energías) y la dispersión de Compton (predominante a altas energías). Estas, a su vez, tienen una dependencia diferente del número atómico y la densidad electrónica. La probabilidad de dispersión de Compton depende de la energía de los rayos X y de la densidad de electrones, mientras que la probabilidad de absorción fotoeléctrica aumenta rápidamente con el número atómico y disminuye rápidamente con el aumento de la energía de los fotones. Por lo tanto, a partir de las mediciones de la atenuación de los haces de rayos X de dos energías efectivas diferentes, es posible resolver la cantidad de dispersión de Compton y absorción fotoeléctrica en el material atravesado. Es posible calcular el número atómico efectivo (Z_{eff}) y la densidad de electrones de un objeto escaneándolo a dos energías de rayos X con suficiente separación de energía (Siddiqui & Khamees, 2004).

(Akin et al., 1996) estimaron la porosidad de la roca al escanearla dos veces, antes y después de cambiar el fluido que las saturaba.

Por otro lado, la permeabilidad debe correlacionarse con la porosidad estimada utilizando definiciones de tipo de roca que se basan en las relaciones permeabilidad-porosidad como la ecuación de Winland (Pittman, 1992).

En el estudio de (Sudakov et al., 2019) se ha demostrado que la permeabilidad puede llegar a ser evaluada usando la imagen del vóxel de tomografía de núcleos, y en el estudio de (Kohli & Arora, 2014) se formuló predecir la permeabilidad a partir de registros de pozo, lo que ha demostrado su relación tanto a registros como imágenes de tomografía.

La variedad de métodos utilizados en el procesamiento de imágenes de tomografía, debido a su alta resolución a menudo produce diferentes representaciones de la geometría y tonalidades del espacio poroso, que impactan muchas propiedades petrofísicas, especialmente aquellas que dependen directamente de la estructura de los poros, es decir, porosidad, factor de formación y permeabilidad absoluta (Leu et al., 2014).

Las imágenes de tomografía contienen matrices numéricas en donde se almacenan los coeficientes de atenuación en unidades de número CT, éstos pueden ser relacionados directamente con la densidad total como se muestra en la ecuación 3 (Ortiz Meneses et al., 2015).

$$RHOB = A * CTN_{low} + B * CTN_{high} + C \quad (3)$$

Donde RHOB es la densidad total, CTN_{low} es CT a baja energía, CTN_{high} CT a alta energía y A-C coeficientes de ajuste que se deben determinar para cada corte.

La relación entre la información de tomografía y la porosidad puede ser soportada con la relación entre la densidad y la porosidad, como se muestra en la ecuación 4. Si la densidad de grano cambia, la porosidad tiene que cambiar para mantener RHOB constante.

$$RHOB = \rho_g * (1 - \phi) + \rho_f * \phi \quad (4)$$

Donde ρ_g es la densidad del grano, ρ_f la densidad del fluido y ϕ la porosidad.

A su vez, la relación permeabilidad-porosidad fue expuesta en varias ocasiones, siendo una de ellas la ecuación de Winland (Pittman, 1992), siguiendo la ecuación 5.

$$\text{Log } r_{35} = 0.732 + 0.588 \text{ Log } K_{air} - 0.864 \text{ Log } \phi \quad (5)$$

Donde r_{35} es la apertura del radio de los poros correspondiente al percentil 35 (μm), K_{air} es la permeabilidad al aire (mD) y ϕ es la porosidad (%).

Gracias a la ecuación 3, 4 y 5, así como a la alta resolución de las imágenes de tomografía se puede establecer una relación entre la información de tomografía y la permeabilidad.

3. Diseño del modelo de aprendizaje profundo

Para diseñar cualquier modelo de AP es necesario conocer el conjunto de datos con el que se trabajará, cuáles son los datos de entrada, y los datos de salida. En este caso existen dos tipos de datos de entrada, numéricos e imágenes, y un dato de salida, la propiedad petrofísica a predecir.

Inicialmente un conjunto de datos de 36 imágenes de tomografía computarizada de secciones de núcleos de perforación de tres pies con sus respectivos datos densidad (RHOB) y factor fotoeléctrico (PEF) fueron proporcionados como entrada para el entrenamiento, las pruebas y la validación del modelo; tanto la información numérica como la de entrada de imágenes deben ser pre-procesada para entrenar a un modelo robusto.

3.1 Preprocesamiento de la información numérica

Se dispuso de 55 datos de porosidad y permeabilidad de RCAL en puntos dados de la sección del núcleo. Dado que la resolución de las imágenes de tomografía es 0,625 mm (resolución vertical de 0,002 pies), estos 55 datos se extrapolaron a 20 puntos de datos por encima y por debajo del punto de observación, con variaciones del 1% por encima y por debajo del valor de porosidad o permeabilidad, esto para otorgar valores matemáticamente diferentes pero estimativos iguales a los valores RHOB y PEF. Con este ajuste se logra tener una mayor densidad de información, pasando de tener 55 datos de porosidad y permeabilidad a un total de 2255 datos.

Después de este ajuste, el marco de datos numéricos es el que se muestra en la tabla 1.

Tabla 1*Conjunto de datos numéricos para la red neuronal.*

	Densidad	PEF	Zeff	Porosidad	Permeabilidad
Conteo	2255	2255	2255	2255	2255
Media	2.497	3.560	14.002	0.055	5.025
Std	0.340	1.612	1.526	0.036	23.073
Mín.	2.197	1.879	11.916	0.009	0.003
25%	2.413	2.519	12.925	0.028	0.017
50%	2.464	3.183	13.794	0.048	0.095
75%	2.501	3.943	14.638	0.076	0.651
Máx.	5.168	10.264	19.095	0.146	145.174

Del conjunto de datos mostrado en la tabla 1, los datos de entrada para el modelo serán la densidad, el factor fotoeléctrico y el número atómico efectivo (Este último es opcional, no tiene gran impacto debido a su dependencia con el factor fotoeléctrico).

Una vez la información a ser empleada ha sido seleccionada, una última práctica es estandarizar los datos, este es un requisito común para los modelos de aprendizaje automático, la idea detrás de la estandarización es transformar los datos para que su distribución tenga un valor de una media cercano a 0 y una desviación estándar cercana a 1, con el fin de garantizar que todos los parámetros de entrada tengan la misma ponderación al momento del entrenamiento. El valor estandarizado del conjunto de datos x se calcula utilizando la ecuación 6.

$$z = \frac{x - \mu}{s} \quad (6)$$

Dónde μ es la media, s la desviación estándar del conjunto de datos, y z es el valor estandarizado, obteniendo el conjunto de datos mostrado en la tabla 2.

Tabla 2*Información numérica pre-procesada para entrenamiento, prueba y validación del modelo*

Propiedad	Densidad	PEF	Zeff
Conteo	2255	2255	2255
Media	2.52E-17	4.29E-16	-1.17E-15
Std	1.00E+00	1.00E+00	1.00E+00
Mín.	-0.884	-1.04	-1.37
25%	-0.248	-0.646	-0.706
50%	-0.0988	-0.234	-0.136
75%	0.0109	0.237	0.417
Máx.	7.85	4.16	3.34

3.2 Preprocesamiento de las imágenes de tomografía computarizada

No todos los conjuntos de registros e imágenes tienen datos de porosidad y permeabilidad, por esta razón, es necesario seccionar las imágenes en los intervalos donde hay información de porosidad y permeabilidad, debido a que la cantidad de datos de tomografía es igual a la dimensión vertical del píxel de las imágenes del núcleo, este corte se puede realizar directamente tomando los píxeles donde se encuentra la información de porosidad y permeabilidad, obteniendo imágenes como la que se muestra en la figura 5 para cada sección.

Figura 5*Sección a escala de grises de una imagen para entrenamiento, validación y prueba del modelo*

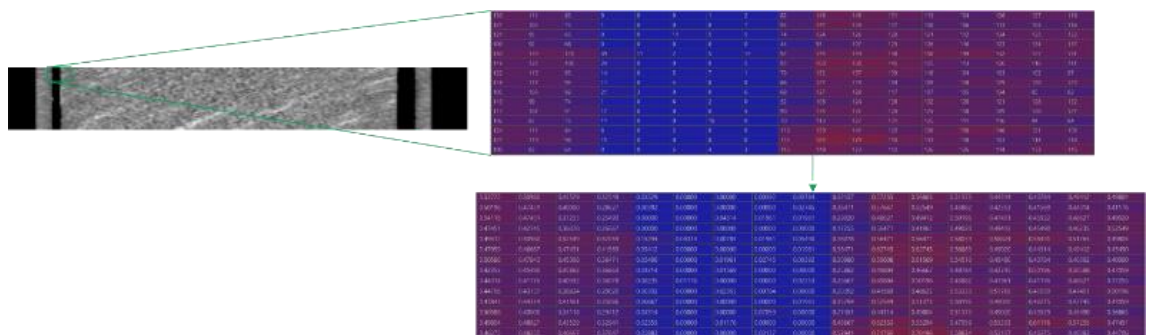
Para lograr un tiempo de entrenamiento de modelo más corto, el segundo paso en el preprocesamiento de la imagen es convertir la imagen a escala de grises como se muestra en la figura 5, ya que solo tiene un canal de color con una escala de 256 niveles. Las imágenes de

tomografía computacional sin procesar utilizan la escala de Hounsfield, que varía de -1000 (para aire) a 3000 o más para materiales más pesados como rocas.

Finalmente, se debe realizar una estandarización de los datos de la imagen. Como una imagen en sí es un tensor con alto, ancho y densidad de píxeles por canal de color (equivalente a una matriz de varias dimensiones), sus datos generalmente se escalan a valores entre 0 y 1. El valor máximo de píxeles que puede tomar una imagen es 255, luego, este procesamiento es sencillo ya que consiste en dividir el tensor de imagen en 255. La figura 6 muestra un ejemplo.

Figura 6

Estandarización de la densidad de píxel de la imagen



Nota. En la imagen se muestra de color rojizo los valores con mayor valor de densidad de pixel y en azul los valores cercanos a cero, note que la estandarización no altera la distribución de estos pixeles, solo los limita a valores entre 0 y 1.

3.3 Diseño y parámetros para el modelo de aprendizaje profundo

En el aprendizaje automático y profundo, hay 3 tipos principales de información:

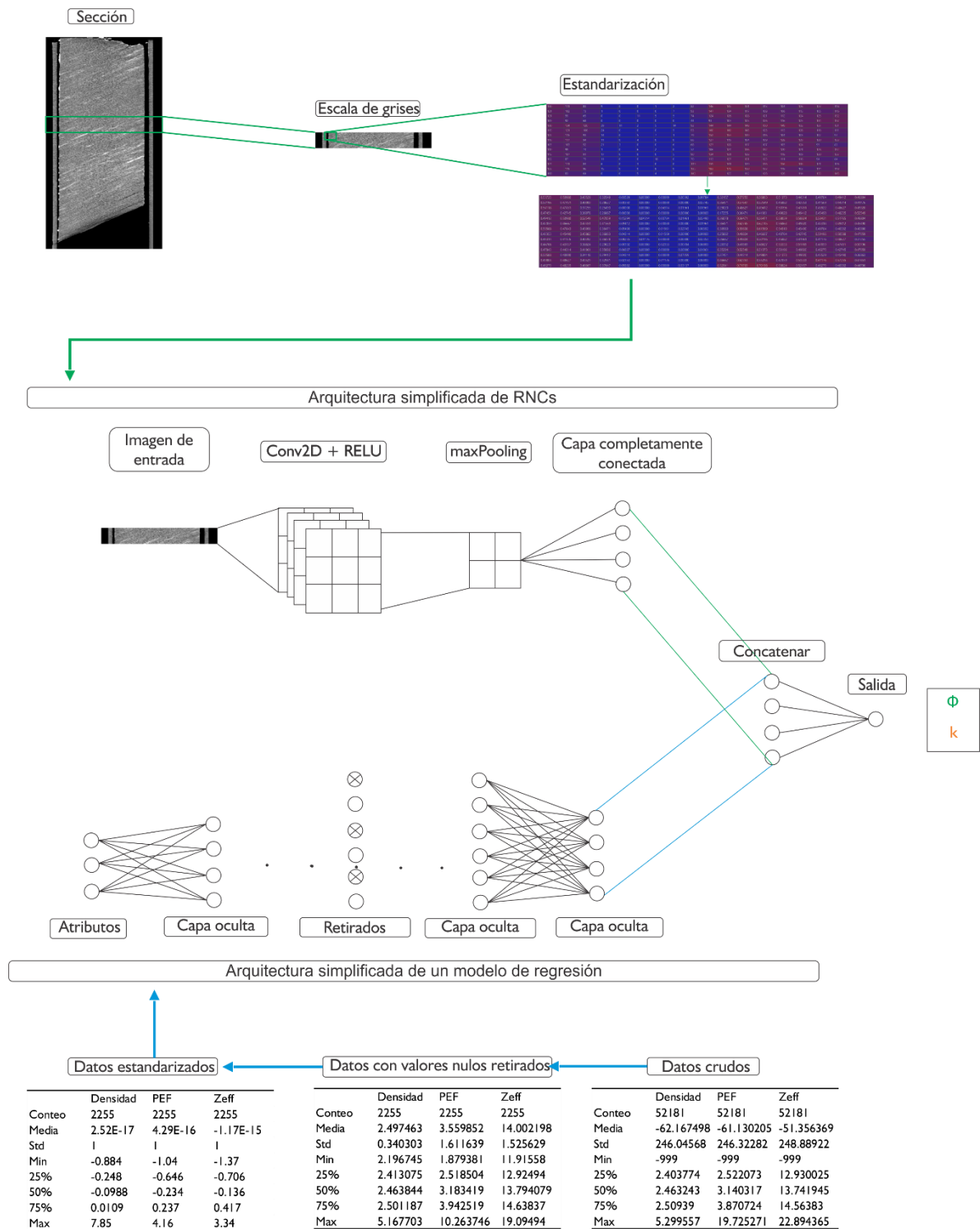
- Numérica/continua, como los registros.
- Categórica. Si/No o preguntas categóricas, como el tipo de roca.
- Imágenes, como las imágenes de tomografía computacional (vídeos también son incluidos).

En este caso, hay información numérica e imágenes de tomografía, esto es conocido como “información mezclada”, por esta razón, es necesario desarrollar un modelo de aprendizaje profundo capaz de manejar este tipo de información.

El modelo de aprendizaje profundo fue desarrollado en Python 3.7, usando diferentes librerías, principalmente la suite ofrecida por Google: Tensorflow-Keras, para aprendizaje profundo, esta suite otorga la posibilidad de crear varios modelos para cada tipo de información, y concatenarlos en un solo modelo; el modelo propuesto es mostrado en la figura 7.

Figura 7

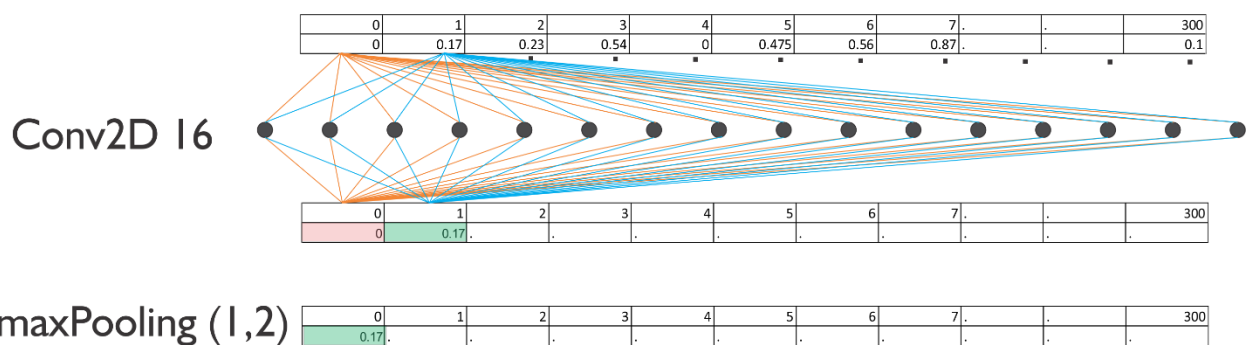
Arquitectura simplificada del modelo con los atributos de entrada y datos de salida.



El modelo en la parte superior de la figura 7 procesa las imágenes seccionadas y preprocesadas de entrada a través de RNCs con N filtros, con activación ReLU (Una función de activación), pasa a través de una capa de reducción maxPooling, la cual toma el valor mayor de una pareja ordenada (1,2) extrayendo las características más relevantes y teniendo un pequeño impacto en el modelo para evitar un sobreajuste sobre los datos de entrenamiento, como puede verse en la figura 8.

Figura 8

Comportamiento de la RCN con la capa de maxPooling.



Nota. A pesar de que la capa de maxPooling reduzca el vector de pixeles de la imagen a los atributos más relevantes, sigue manteniendo las dimensiones de la imagen de entrada.

Por el lado de la entrada de atributos numéricos (parte inferior del modelo de la figura 7), se observa que solo esta poblada con varias capas ocultas llena de varios neuronas de manera lineal, y solo una capa de retiro para prevenir el sobreajuste del modelo (es decir, que el modelo solo se ajuste a la información de entrenamiento y no de pruebas). Posteriormente el modelo es unido mediante una capa de concatenación que recibe exactamene 4 neuronas, una vez el modelo es concatenado, se reduce el número de neuronas en la última capa a 1, sin función de activación,

esta neurona es la neurona del dato de salida, el cual es la propiedad petrofísica a predecir, porosidad o permeabilidad.

El modelo completo empleado para este trabajo se encuentra resumido en la tabla 3.

Tabla 3

Arquitectura completa de modelo

Modelo de imágenes RNCs			Modelo de regresión lineal PMC		
Capas	Operación	Comentarios	Capas	Operación	Comentarios
3	2D Convolución	Tamaño de red= (3,3),	4	Ocultas	Neuronas = 32, 64, 128, 256, Activación = "ReLU"
		Segmento = (1,1),			
		Relleno = "mismo",			
		Filtros = 16, 32, 64,			
	Activación = "ReLU"	1	Retiro	Tasa= 0.2, Prevenir sobreajuste	
Normalización de bache			5	Ocultas	Neuronas = 128, 64, 32, 16, 4, Activación = "ReLU"
Max Pooling 2D	Pool size = (1,2)				
1	Retirar	Tasa= 0.2, retira 20% de las neuronas, para evitar sobreajuste.			
1	Aplanar	La capa aplanar, aplanar el tensor a 1D para ser tratado por capas ocultas.			
4	Ocultas	Neuronas = 1024, 512, 256, 4			
Concatenar los modelos					
Capas	Operación		Comentario		
1	Concatenar		Concatena el modelo		
1	Ocultas		Neuronas = 4,		
			Activación = "ReLU"		
1	Ocultas		1 salida (la propiedad petrofísica a predecir)		

Una vez el modelo es diseñado, los parámetros para el entrenamiento y estudio de la pérdida del modelo son establecidos en la tabla 4.

Tabla 4

Parámetros para el entrenamiento del modelo de aprendizaje profundo

Parámetro	Valor
Máximo número de épocas a entrenar	300
Porción para validación	20%
Porción para pruebas	20%
Métricas	R ² ECM y EAM
Numero de lotes	8

Nota. La métrica principal será el coeficiente R², sin embargo, se analizará el modelo también con el error cuadrado medio (ECM) y el error absoluto medio (EAM), el número de lotes hace referencia a la cantidad de ejemplos que hay en un lote, al decir que hay 8 lotes, se dice que se utilizan 8 ejemplos para una iteración, una iteración es una actualización de las ponderaciones de un modelo durante el entrenamiento.

Con estos parámetros se cuenta con 1353 datos para entrenamiento, 451 para validación y 451 para pruebas. Para evitar un sobre ajuste en el entrenamiento y un ajuste inadecuado en las pruebas, estos datos fueron asignados aleatoriamente en el entrenamiento, validación y prueba. Además, gracias al robusto esquema del modelo, es posible trabajar con lotes de información de tamaño 8.

Los modelos fueron reentrenados con más información del mismo pozo, añadiendo 1057 datos para entrenamiento, 353 para validación, y 353 para pruebas, con el mismo preprocesamiento, perteneciente a 53 datos extras de porosidad y permeabilidad.

El conjunto de datos para el reentrenamiento es mostrado en la tabla 5.

Tabla 5

Conjunto de datos numéricos pertenecientes al segundo set de información.

Propiedad	Densidad	PEF	Zeff	Porosidad	Permeabilidad
Conteo	1763	1763	1763	1763	1763
Media	2.269	2.338	12.567	0.221	2678.960
Std	0.161	0.741	0.922	0.081	4418.568
Mín.	1.986	1.735	11.655	0.075	0.025
25%	2.141	1.921	11.989	0.162	8.646
50%	2.234	2.028	12.170	0.250	214.408
75%	2.392	2.595	13.033	0.292	3333.349
Máx.	2.988	7.378	17.422	0.366	18157.47

Nota. En este caso la permeabilidad tiene valores mayores y no tan cercanos a cero a diferencia del primer conjunto de datos.

Tanto la información del primer conjunto de datos como del segundo conjunto de datos fueron preprocesados de la misma manera.

Los resultados del modelo resultante de este entrenamiento son concisos y son mostrados a continuación.

4. Resultados

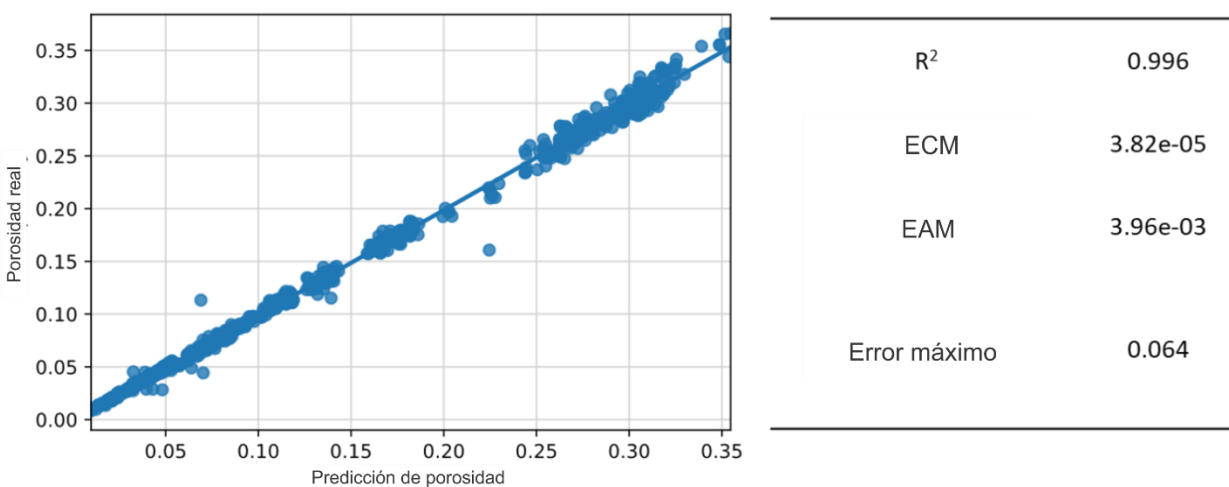
En esta sección se presentan los resultados de los modelos desarrollados, el análisis de sus métricas, y su desempeño prediciendo propiedades petrofísicas en zonas sin datos a partir de información e imágenes de tomografía de núcleo, además, se muestra una breve discusión y análisis de las predicciones obtenidas.

4.1 Análisis de métrica del modelo de porosidad

El comportamiento del modelo de porosidad se muestra en la figura 9, como se puede observar, los resultados obtenidos en el proceso de validación son excelentes, así como la pendiente de predicción y el coeficiente R^2 tiende a 1 (0,996), el error cuadrático medio (ECM) es $3.82e-5$ que tiende a cero, el error absoluto medio (EAM) es $3.96E-3$, también bastante bajo, y el error máximo de la muestra fue 0.064, por lo que se obtiene un modelo robusto.

Figura 9

Métricas y gráfica de regresión lineal del modelo de porosidad



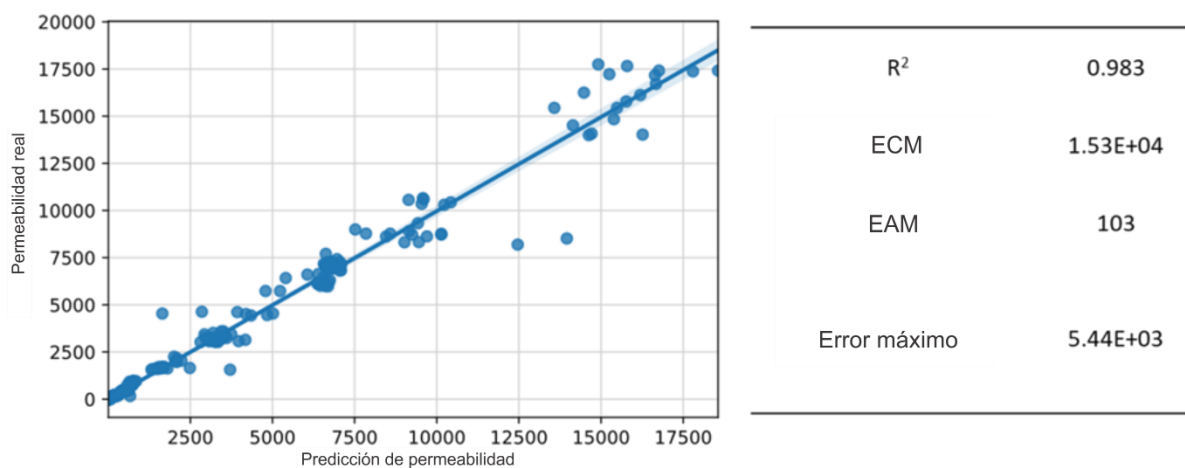
El modelo de porosidad presenta un ajuste excelente, capaz de predecir valores con una exactitud precisa.

4.2 Análisis de métrica del modelo de permeabilidad

El desempeño del modelo de permeabilidad se muestra en la figura 10. Los resultados obtenidos en la información de validación son buenos, así como la pendiente de predicción, el coeficiente R^2 tiende a 1 (0.983), sin embargo, dada la naturaleza logarítmica de la muestra el error cuadrático medio y el error absoluto medio muestran valores más altos, así como el error máximo alcanzado por una muestra en todo el conjunto de datos.

Figura 10

Métrica y gráfica de regresión lineal del modelo de permeabilidad



Nota. Valores desde 0 a 10000 mD de permeabilidad se ajustan mucho más, en general, el ECM, EAM y error máximo están altamente influenciados por estos valores tan altos de permeabilidad.

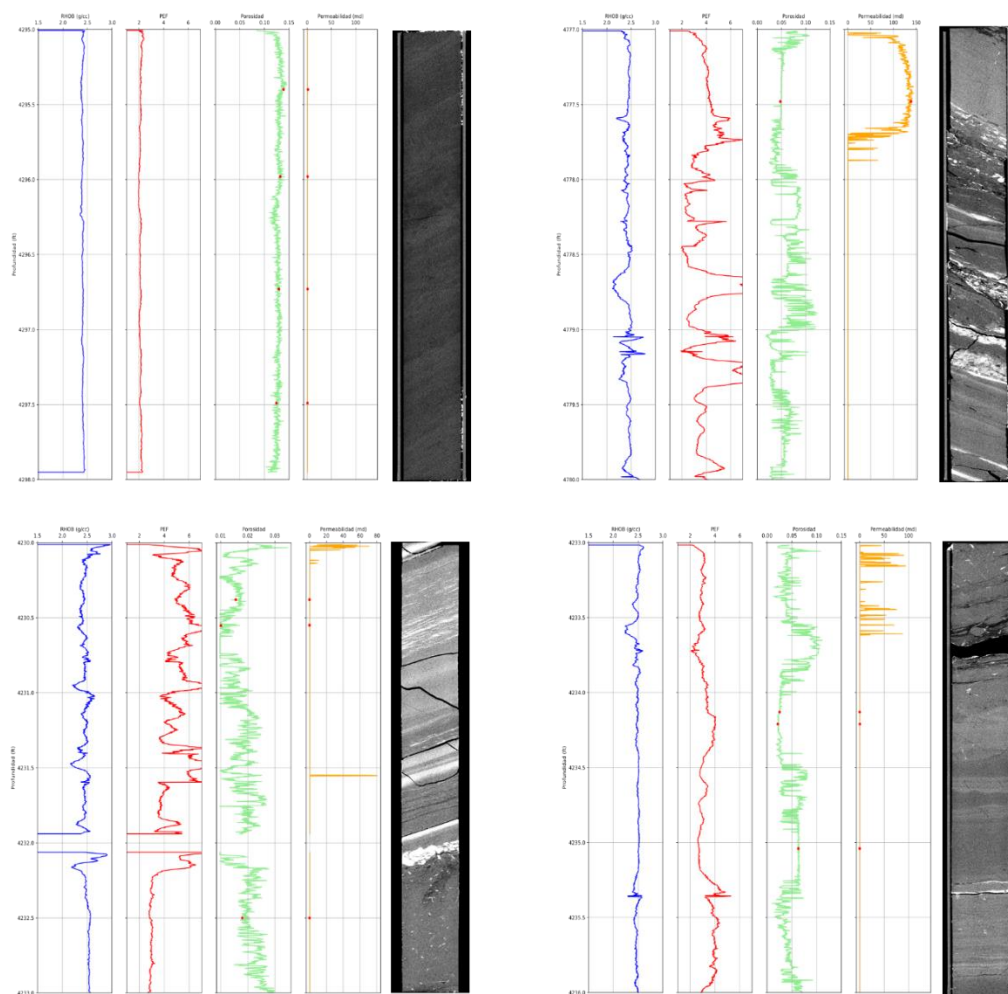
El modelo de permeabilidad presenta un ajuste muy bueno, capaz de predecir valores de 0 a 10000 mD con una gran exactitud, los valores de permeabilidad mayores a estos tendrán alguna desviación, sin embargo, se considera un excelente modelo.

4.3 Predicción de propiedades petrofísicas

Una vez los modelos ya entrenados, con las ponderaciones que más se ajustan, se procede a predecir estas propiedades para varios pies de roca y registro sin datos de porosidad y permeabilidad, como se muestra en las figuras 11 y 12.

Figura 11

Predicciones de permeabilidad y porosidad para 4 secciones de núcleo del primer conjunto de datos.

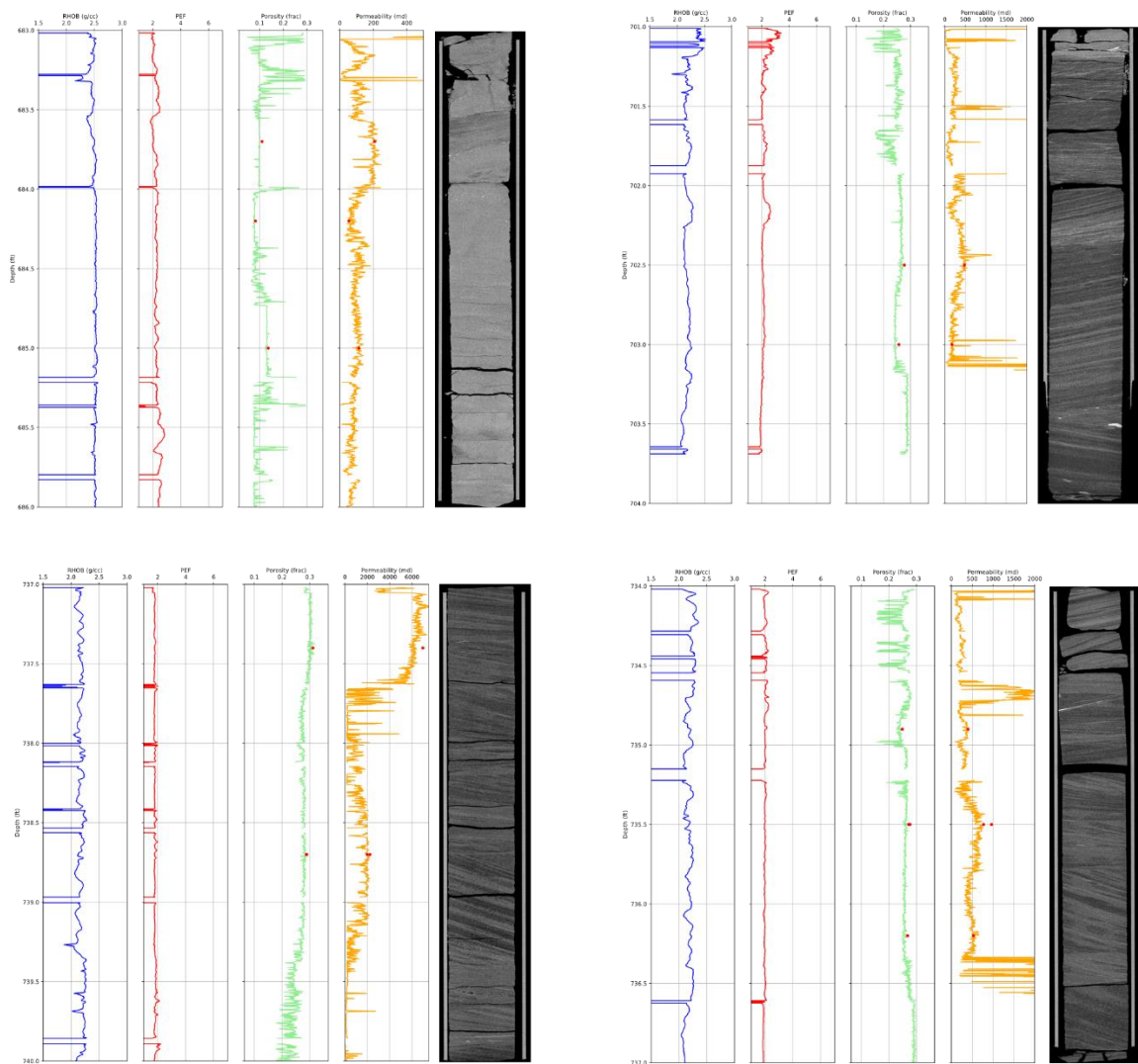


Nota. La curva de color azul en el track 1 es la densidad (RHOB), la curva de color rojo en el track 2 es factor fotoeléctrico (PEF), la curva de color verde en el track 3 son las

predicciones de porosidad a partir de la imagen de TC en el track 5 y los dos registros en el track 1 y 2, la curva de color naranja claro es la predicción de permeabilidad a partir de los mismos registros e imágenes, los puntos rojos en el track 3 y 4 son los valores reales de porosidad y permeabilidad.

Figura 12

Predicciones de permeabilidad y porosidad para 4 secciones de núcleo del segundo conjunto de datos.



Nota. La curva de color azul en el track 1 es la densidad (RHOB), la curva de color rojo en el track 2 es factor fotoeléctrico (PEF), la curva de color verde en el track 3 son las

predicciones de porosidad a partir de la imagen de TC en el track 5 y los dos registros en el track 1 y 2, la curva de color naranja claro es la predicción de permeabilidad a partir de los mismos registros e imágenes, los puntos rojos en el track 3 y 4 son los valores reales de porosidad y permeabilidad. Note que los valores de permeabilidad en este conjunto de datos son mucho más variados.

Como se puede observar en estas 8 predicciones de las figuras 11 y 12, los modelos se ajustan muy bien a los datos reales, permitiendo predecir porosidad y permeabilidad para varios pies de roca solo con los registros e imágenes de tomografía.

4.4 Discusión y análisis.

Los modelos desarrollados son un intento de utilizar el aprendizaje profundo y la inteligencia artificial para predecir la porosidad y la permeabilidad a partir de datos de densidad, factor fotoeléctrico, número atómico efectivo e imágenes tomografía computacional en escala de grises.

Los modelos expuestos presentan excelentes resultados, es decir, se ajustan a la teoría y a las postulaciones previas de (Kohli & Arora, 2014) y de (Sudakov et al., 2019), además de los estudios analíticos que relacionan la densidad con la porosidad, y por ende la relación porosidad-permeabilidad como por ejemplo la relación entre atenuación de rayos X, porosidad y densidad, o como la relación porosidad-permeabilidad radio de poro, o incluso como el PEF se relaciona con la litología lo que puede ser un indicativo de roca, que finalmente relaciona aún más la permeabilidad con la porosidad.

La RNC y la estimación de regresión lineal son buenas y pueden mejorarse proporcionando más información de los núcleos y los parámetros de entrenamiento para más épocas (Si se cuenta con un equipo de cómputo con mayor capacidad de tarjeta gráfica).

El entrenamiento del modelo se logra mediante el uso de la librería de aprendizaje automático y profundo “Tensorflow”. Toma aproximadamente 1 minuto por época gracias al aumento de GPU del equipo de cómputo. El entrenamiento total necesario para cada modelo es de aproximadamente 3 horas y media por modelo, más de 45 modelos de porosidad y permeabilidad fueron entrenados en este trabajo con el fin de encontrar la mejor combinación de capas para un entrenamiento óptimo.

Una vez entrenado el modelo, la estimación de las métricas de las instancias del conjunto de datos de prueba de tarda aproximadamente 5 segundos en completarse.

5. Conclusiones

Los resultados de este estudio demuestran el potencial significativo de la inteligencia artificial para modelos basados en datos e imágenes de tomografía computarizada en cuanto a la predicción de propiedades petrofísicas como la porosidad y permeabilidad. Gracias a las relaciones entre la porosidad y los coeficientes de atenuación de rayos X (Números de TC que lee el tomógrafo y se relacionan para la determinación de la relación con la densidad), las diferentes relaciones de permeabilidad dentro del pozo y la alta resolución de las imágenes de tomografía, la red neuronal permite encontrar la relación que mejor describe el comportamiento.

Se presenta un modelo predictivo que usa CNN y PMC en un modelo concatenado para estimar dos propiedades petrofísicas a partir de imágenes de TC en escala de grises, registros RHOB y PEF. Se crearon varios registros de porosidades y permeabilidades mediante la predicción de los datos existentes con los modelos entrenados.

Un algoritmo de aprendizaje profundo usando la suite de Tensorflow de Google fue implementado, lo que facilita la generación, entrenamiento, validación y uso de estas redes neuronales incluyendo las convolucionales las cuales se aplican para imágenes.

Ambos modelos, porosidad y permeabilidad, mostraron excelentes resultados, prediciendo correctamente. Aunque el modelo de permeabilidad a permeabilidades muy altas presenta un poco más de ambigüedad, se considera un modelo robusto.

6. Recomendaciones

Se recomienda entrenar el modelo de ser posible con mayor número de secciones de imágenes de tomografía, radial y sección YZ, para tener resultados más robustos.

Se recomienda complementar el algoritmo con un módulo de determinación de tipos de roca o unidades de flujo con base en relaciones permeabilidad y porosidad (ej. Winland 35 o FZI).

Se recomienda integrar en este trabajo otros registros e información de pozos, por ejemplo, Gamma Ray, Core Gamma, Registros Resistivos, etc., con el fin de obtener modelos petrofísicos más robustos para todo el pozo

Referencias Bibliográficas

- A. Kantzas D.F. Marentette K.N. Jha. (1992). *Visualization To Quantitative Core Analysis*.
- Akin, S., Demiral, M., & Okandan, E. (1996). *A novel method of porosity measurement utilizing computerized tomography*. <https://avesis.metu.edu.tr/yayin/b00ba0b0-be8a-4be7-bbea-20eddb17c7b0/a-novel-method-of-porosity-measurement-utilizing-computerized-tomography>
- Alqahtani, N., Armstrong, R. T., & Mostaghimi, P. (2018). Deep learning convolutional neural networks to predict porous media properties. *Society of Petroleum Engineers - SPE Asia Pacific Oil and Gas Conference and Exhibition 2018, APOGCE 2018*. <https://doi.org/10.2118/191906-ms>
- Amini, A. (n.d.). *MIT Deep Learning 6.S191*. Retrieved January 21, 2021, from <http://introtodeeplearning.com/>
- Andrä, H., Combaret, N., Dvorkin, J., Glatt, E., Han, J., Kabel, M., Keehm, Y., Krzikalla, F., Lee, M., Madonna, C., Marsh, M., Mukerji, T., Saenger, E. H., Sain, R., Saxena, N., Ricker, S., Wiegmann, A., & Zhan, X. (2013). Digital rock physics benchmarks-Part I: Imaging and segmentation. *Computers and Geosciences*, 50, 25–32. <https://doi.org/10.1016/j.cageo.2012.09.005>
- Bhattacharya, S., & Nikolaou, M. (2011). Using data from existing wells to plan new wells in unconventional gas field development. *Society of Petroleum Engineers - Canadian Unconventional Resources Conference 2011, CURC 2011*, 1, 766–778. <https://doi.org/10.2118/147658-ms>
- Brunette, E. S., Flemmer, R. C., & Flemmer, C. L. (2009). A review of artificial intelligence. *ICARA 2009 - Proceedings of the 4th International Conference on Autonomous Robots and*

- Agents, February 2009*, 385–392. <https://doi.org/10.1109/ICARA.2000.4804025>
- Cuddy, S. J., & Putnam, T. W. (1998). Litho-facies and permeability prediction from electrical logs using fuzzy logic. *Society of Petroleum Engineers - Abu Dhabi International Petroleum Exhibition and Conference 1998, ADIPEC 1998, January*, 137–147. <https://doi.org/10.2523/49470-ms>
- Elmousalami, H. H., & Elaskary, M. (2020). Drilling stuck pipe classification and mitigation in the Gulf of Suez oil fields using artificial intelligence. *Journal of Petroleum Exploration and Production Technology*, 10(5), 2055–2068. <https://doi.org/10.1007/s13202-020-00857-w>
- Estudio detallado del AA: Regresión lineal*. (n.d.). Retrieved January 18, 2021, from <https://developers.google.com/machine-learning/crash-course/descending-into-ml/linear-regression>
- Goodfellow, I., Bengio, Y., & Courville, A. (2019). Deep Learning. In *MIT* (Vol. 29, Issue 7553).
- Hall, B. (2016). Facies classification using machine learning. *Leading Edge*, 35(10), 906–909. <https://doi.org/10.1190/tle35100906.1>
- Honarpour, M. M., Cromwell, V., Hatton, D., & Satchwell, R. (1985). Reservoir rock descriptions using computed tomography (CT). *Proceedings - SPE Annual Technical Conference and Exhibition, 1985-Septe*. <https://doi.org/10.2523/14272-ms>
- Kohli, A., & Arora, P. (2014). Application of artificial neural networks for well logs. *Society of Petroleum Engineers - International Petroleum Technology Conference 2014, IPTC 2014: Unlocking Energy Through Innovation, Technology and Capability*, 3(January), 2276–2283. <https://doi.org/10.2523/iptc-17475-ms>
- Leu, L., Berg, S., Enzmann, F., Armstrong, R. T., & Kersten, M. (2014). Fast X-ray Micro-Tomography of Multiphase Flow in Berea Sandstone: A Sensitivity Study on Image

- Processing. *Transport in Porous Media*, 105(2), 451–469. <https://doi.org/10.1007/s11242-014-0378-4>
- Li, S., Yang, C., Sun, H., & Zhang, H. (2019). Seismic fault detection using an encoder-decoder convolutional neural network with a small training set. *Journal of Geophysics and Engineering*, 16(1), 175–189. <https://doi.org/10.1093/jge/gxy015>
- Mitchel, T. M. (1997). *Machine Learning*. McGraw-Hill Science/Engineering/Math. http://www.goodreads.com/book/show/213030.Machine_Learning
- Mohaghegh, S. D., Grujic, O., Zargari, S., & Kalantari, M. (2011). Modeling, history matching, forecasting and analysis of shale reservoirs performance using artificial intelligence. *Society of Petroleum Engineers - SPE Digital Energy Conference and Exhibition 2011, January*, 274–287. <https://doi.org/10.2118/143875-ms>
- Mohanty, A. (2021). *Multi layer Perceptron (MLP) Models on Real World Banking Data*. Becoming Human: Artificial Intelligence Magazine. <https://becominghuman.ai/multi-layer-perceptron-mlp-models-on-real-world-banking-data-f6dd3d7e998f>
- Ortiz, A. F., Herrera, E. H., & Santos, N. (2020). Porosity prediction from X-ray computed tomography logs (RHOB and PEF) using Artificial Neural Networks (ANN). *Boletín de Geología*, 42(3), 141–149. <https://doi.org/10.18273/revbol.v42n3-2020>
- Ortiz Meneses, A. F., Plata Chaves, J. M., Herrera Otero, E., & Santos Santos, N. (2015). Caracterización Estática De Rocas Por Medio De Tomografía Computarizada De Rayos-X TAC. *Revista Fuentes El Reventón Energético*, 12(2), 57–63. <https://doi.org/10.18273/revfue.v13n1-2015005>
- Pandey, Y. N., Rastogi, A., Kainkaryam, S., Bhattacharya, S., & Saputelli, L. (2020). Machine Learning in the Oil and Gas Industry. In *Machine Learning in the Oil and Gas Industry*.

<https://doi.org/10.1007/978-1-4842-6094-4>

Pittman, E. D. (1992). Relationship of porosity and permeability to various parameters derived from mercury injection-capillary pressure curves for sandstone. In *American Association of Petroleum Geologists Bulletin* (Vol. 76, Issue 2, pp. 191–198).
<https://doi.org/10.1306/bdff87a4-1718-11d7-8645000102c1865d>

Saputelli, L. A., Nikolaou, M., & Economides, M. J. (2005). Self-learning reservoir management. *SPE Reservoir Evaluation and Engineering*, 8(6), 534–547. <https://doi.org/10.2118/84064-pa>

Shahkarami, A., Mohaghegh, S. D., Gholami, V., & Haghigat, S. A. (2014). Artificial intelligence (AI) assisted history matching. *Society of Petroleum Engineers - SPE Western North American and Rocky Mountain Joint Meeting, January 2019*.
<https://doi.org/10.2118/169507-ms>

Siddiqui, S., & Khamees, A. A. (2004). Dual-energy CT-scanning applications in rock characterization. *Proceedings - SPE Annual Technical Conference and Exhibition*, 2961–2969. <https://doi.org/10.2118/90520-ms>

Sudakov, O., Burnaev, E., & Koroteev, D. (2019). Driving digital rock towards machine learning: Predicting permeability with gradient boosting and deep neural networks. *Computers and Geosciences*, 127(February), 91–98. <https://doi.org/10.1016/j.cageo.2019.02.002>

Sun, J., Ma, X., & Kazi, M. (2018). Comparison of decline curve analysis DCA with recursive neural networks RNN for production forecast of multiple wells. *SPE Western Regional Meeting Proceedings, 2018-April*. <https://doi.org/10.2118/190104-ms>

Unrau, S. (2017). *Machine Learning Algorithms Applied to Detection of Well Control Events*. 24–27. <https://www.onepetro.org/download/conference-paper/SPE-188104-MS?id=conference->

paper%2FSPE-188104-MS

Withjack, E. M., Devier, C., & Michael, G. (2003). The Role of X-Ray Computed Tomography in Core Analysis. *SPE Western Regional/AAPG Pacific Section Joint Meeting*, 41–52.

<https://doi.org/10.2523/83467-ms>

Wu, X., Shi, Y., Fomel, S., & Liang, L. (2019). Convolutional neural networks for fault interpretation in seismic images. *2018 SEG International Exposition and Annual Meeting, SEG 2018, August*, 1946–1950. <https://doi.org/10.1190/segam2018-2995341.1>

Zangl, G., Neuhofer, L., Zabel, D., Toppel, P., & Pantazescu, C. (2016). *SPE-181072-MS Smart and Automated Workover Candidate Selection Concept of the Expert System*.

Apéndice A

La canalización del aprendizaje profundo

La infraestructura que rodea un algoritmo de aprendizaje automático. Una canalización incluye la recopilación de los datos, la colocación de los datos en archivos de datos para entrenamiento, el entrenamiento de uno o más modelos y la exportación de los de los modelos para la producción.

La canalización del aprendizaje profundo y en general del aprendizaje automático es prácticamente lineal, pero aun así complejo, solo se necesita información, limpiar la información, diseñar el modelo, estudiar el modelo, rediseñarlo y cuando se esté conforme con el modelo desarrollado, predecir para valores nuevos.

No es necesario (aunque ayuda) conocer la matemática y estadística compleja detrás de estos modelos, solo hay que conocer su empleabilidad, restricciones y beneficios para desarrollar un modelo de aprendizaje automático, cualquiera puede seguir la canalización del aprendizaje profundo.

Se recomienda enormemente el libro de (Goodfellow et al., 2019) y los cursos ofrecidos por el MIT gratis para interesarse y embarcarse en la inteligencia artificial.

Apéndice B

Script para seccionar las imágenes para entrenamiento

```
1. ##Importa la librería a emplear
2. from PIL import Image #Libreria para manejar imágenes
3.
4. img = Image.open("direccion/directorio/imagen.png").convert('L') #Abre la imagen y la
   pasa a escala de grises
5. img = img.resize((301,img.height), Image.ANTIALIAS) #Hace un redimensionamiento de la
   imagen de alto a ancho de 301 px, manteniendo proporciones
6. h=1297 #Alto donde empiezan a tomarse valores, valor tomando de la tabla de datos donde
   tiene valores de permeabilidad y porosidad
7. img = img.crop((0,h,301,h+41))#Corta la imagen en una imagen de alto 41 px y ancho 301
   px
8. img.show()
9. img.save("direccion/directorio/guardar/imagen.png") #Salva la imagen con un nombre
```

Apéndice C

Script para el preprocesamiento de imágenes

```
1. #Importa librerías a emplear en el preparamiento de la información para entrenamiento,
   validación y prueba
2. import os #Librería para manejar funciones del sistema operativo, explorador de archivo
   s, etc...
3.
4. import pandas as pd #Librería para trabajar con conjunto de datos
5. import numpy as np #Librería para operaciones matemáticas con tensores
6. import matplotlib.pyplot as plt #Librería para graficar variables
7. import seaborn as sns #Librería que trae por defecto gráficas como la regresión lineal
8.
9. import tensorflow as tf #Suite ofrecida por Google para aprendizaje profundo
10. gpus = tf.config.experimental.list_physical_devices('GPU') #Verifica que exista GPU y l
   a usa
11. tf.config.experimental.set_memory_growth(gpus[0], True)
12.
13. from tensorflow import keras #dependencias de Tensorflow
14. from tensorflow.keras import layers
15. import tensorflow_docs as tfdocs
16.
17. from sklearn.preprocessing import StandardScaler #Librería que agiliza la estandarizaci
   ón de datos.
18.
19.
20.
21.
22. #Toma las imágenes y las compone como tensores para entrenarlas
23. img_train_dir = 'D:/Tomografia/Training_Images/' #Lee el directorio con las imágenes
24. train_imgs = [] #Crea una lista vacía donde se pondrán las imágenes ya para entrenamien
   to
25.
26. for img in os.listdir(img_train_dir):
27.
28.     img = tf.io.read_file(filename=os.path.join(img_train_dir,img)) #Lee la imagen
29.     img = tf.io.decode_image(img, dtype=tf.dtypes.uint8) #La transforma en tensor con f
   ormato a nivel de 255
30.     img = tf.keras.preprocessing.image.img_to_array(img) #Pasa la imagen a un arreglo
31.     img /= 255.0 #Divide el arreglo en 255.0 (Estandarización de la imagen)
32.     for i in range(41):
33.         train_imgs.append(img[i]) #Agrega cada línea de píxel de imagen a la lista vací
   a de imagen
34. train_imgs_np = np.array(train_imgs) #Transforma la lista en arreglos de numpy, estos a
   rreglos llevan menos tiempo de cómputo y son más rápidos de entrenar
35. train_imgs_np = np.reshape(train_imgs_np, [1, 1, 301, 1])#Transforma el arreglo en un t
   ensor de (x cantidad de píxeles de imagen, 1 píxel de altura, 301 píxeles de ancho, 1 ca
   nal de color escala de grises)
```

Apéndice D

Script del preprocesamiento de datos numéricos

```
1. df = pd.read_csv('D:\Tomografia\Registros_Con_Porosidad_Permeabilidad.csv') #Lee el conjunto de datos, en este conjunto de datos hay 8 columnas, Profundidad, RHOB, PEF, Zeff, Core, Seccion, Porosidad, Permeabilidad
2.
3. df =df.dropna() #Elimina los valores nulos, es decir todos los puntos donde no hay porosidad ni permeabilidad
4.
5.
6. df_numerico = df.drop(['Core', 'Seccion', 'Profundidad'], axis=1) #Se eliminan las columnas que no se necesitan para el entrenamiento
7.
8. targets = df_numerico[['Permeabilidad']] #Se elige el objetivo del modelo en este caso permeabilidad
9. numerical_features = df_numerico[['Densidad', 'PEF', 'Zeff']] #Se eligen los inputs numericos
10.
11. ### Se importa un escalador estandar, sencillamente estandariza el conjunto de datos ##
12. standardScaler = StandardScaler()
13.
14. features_procesados = pd.DataFrame(standardScaler.fit_transform(numerical_features),
15.                                   columns=numerical_features.columns,
16.                                   index= numerical_features.index)
17.
18. #Features procesados son los datos numéricos de entrada para el entrenamiento preprocesados
```

Apéndice E

Script del diseño del modelo de aprendizaje profundo

```

1. #Importar libreria para separar los datos en entrenamiento y prueba
2. from sklearn.model_selection import train_test_split
3.
4. train_numerical, test_numerical, train_images, test_images, train_target, test_target =
   train_test_split(numerical_features, training_images, targets, test_size=0.2, random_s
   tate =1)
5.
6. #Se separan los datos en entrenamiento y prueba, tomando una porción de datos para prue
   ba del 20%, además, se barajan estos datos para evitar un sobreajuste en una sola zona
   de datos.#
7. ##Modelo de RNCs para las imágenes
8. conv_model = tf.keras.models.Sequential([
9.
10.     layers.Conv2D(16, (3,3), padding='same', activation='relu', #Capa de entrada Conv2D
11.         input_shape=train_images[0].shape),
12.     layers.MaxPooling2D(pool_size=(1,2)), #MaxPooling (1,2)
13.     layers.BatchNormalization(axis=-1),
14.     layers.Conv2D(32, (3,3),padding='same', activation='relu'),
15.     layers.MaxPooling2D(pool_size=(1,2)),
16.     layers.Conv2D(64, (3,3),padding='same', activation='relu'),
17.     layers.Dropout(0.2), #Dropout para evitar sobreajuste
18.
19.     layers.Flatten(), #se transforma el tensor de imagenes en un vector
20.
21.     layers.Dense(512, activation='relu'), #Capas ocultas
22.     layers.Dense(256, activation='relu'),
23.     layers.Dense(128, activation='relu'),
24.     layers.Dense(4, activation='relu')
25. ])
26. ])
27. ## Modelo de PMCs para manejo numérico de datos
28. reg_model = tf.keras.models.Sequential([
29.     layers.Dense(32, activation='relu', input_shape=[len(numerical_features.keys())]),
   #Datos de entrada
30.     layers.Dense(64, activation='relu'), #Capas ocultas
31.     layers.Dense(128, activation='relu'),
32.     layers.Dense(256, activation='relu'),
33.
34.     layers.Dropout(0.2), #Capa dropout para evitar sobreajuste
35.     layers.Dense(128, activation='relu'),
36.     layers.Dense(64, activation='relu'),
37.     layers.Dense(32, activation='relu'),
38.     layers.Dense(16, activation='relu'),
39.
40.     layers.Dense(4, activation='relu')
41.
42.
43. ])

```

```
44. # Se concatena el modelo
45. combined_input = layers.concatenate([reg_model.output, conv_model.output])
46. x = layers.Dense(4, activation='relu')(combined_input)
47. x = layers.Dense(2, activation='relu')(x)
48. x = layers.Dense(1)(x)
49.
50. #Se declara un modelo combinado que toma los datos de entrada del modelo de regresión y
    RNCs
51. combined_model = tf.keras.models.Model(inputs=[reg_model.input, conv_model.input],
52.                                         outputs = x)
53.
54. #Se compila el modelo con perdida (principal metrica) de ECM y tras metricas adicionales
    como EAM
55. combined_model.compile(loss="mse",
56.                         optimizer=keras.optimizers.Adam(0.001),
57.                         metrics=['mae', 'mse'])
58.
```

Apéndice F

Script para el entrenamiento y estudio de la métrica del modelo de AP

```
1. #Entrenamiento del modelo
2. trainig_hist = combined_model.fit(x=[train_numerical, train_images], y=train_público,
3.                                 validation_split= 0.2,
4.                                 epochs=300, batch_size=8,
5.                                 verbose = False,
6.                                 )
7. #300 epocas, tamaño de lote de 8 y porción para validación de 0.2
8.
9. #Evaluar el modelo#
10. y_pred = combined_model.predict([test_numerical, test_images]) #Se predice con la porción para prueba
11.
12. from sklearn.metrics import r2_score
13. r2_score(test_target, y_pred) #Se calcular el puntaje R2 para determinar la eficacia de l modelo
14.
15. #Se grafica el predecido contra el actual, en este caso permeabilidad#
16. perm_predicted= y_pred
17.
18. actual_pereability= np.array(test_target['Permeabilidad']).reshape(-
19. 1,1) #Transforma la permeabilidad en un tensor de numpy para graficar
20. #Se escoge un grafico de regresión y se ajusta para una visualización más amigable.
21. ax=sns.regplot(x=perm_predicted,y=actual_permeability)
22. ax.set_xlabel('Predicted permeability')
23. ax.set_ylabel('Actual permeability')
24. ax.set_axisbelow(True)
25. ax.yaxis.grid(color='lightgray', linestyle='-')
26. ax.xaxis.grid(color='lightgray', linestyle='-')
27.
28. #Importa libreria para estimar las demás perdidas, y se calculan#
29. from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error, max_error
30. print('R2_score: {:.3}\nMSE: {:.3}\nMAE: {:.3}\nMax Error: {:.3}'.format(r2_score(actual_permeability,perm_predicted),
31. mean_squared_error(actual_permeability,perm_predicted),
32. mean_absolute_error(actual_permeability,perm_predicted),
33. max_error(actual_permeability, perm_predicted)))
```