Todo lo que necesitas para reconocer estructuras geológicas es visualizar características y prestar Atención *

Alvear Silvana[1 [msalvear@uce.edu.ec]], Caiza Verónica[1 [vccaiza@uce.edu.ec]], Changoluisa Viviana[1 [lvchangoluisa@uce.edu.ec]], and Acevedo David[2 [dfacevedoa@uce.edu.ec]]

Universidad Central del Ecuador, Av Universitaria Quito-Ecuador

Abstract. El tratamiento de problemas geológicos mediante Inteligencia Artificial se ha convertido en una herramienta que complementa al geólogo. La extracción y visualización de características, así como el uso de mecanismos de atención para reconocer estructuras geológicas implementando redes neuronales convolucionales y redes transformer han dado como resultado de precisión en la evaluación con valores de 0.8556 y 0.8944 respectivamente. Los resultados obtenidos radican en extraer características de forma jerárquica a medida que se profundiza en la red convolucional, identificando así patrones de textura, disposición de ejes, formas geometricas, lineaciones. Por su parte las redes transfomer mediante el mecanismo de autoatención van puntuado con altos valores de atención zonas o regiones propias e imprescindibles de cada estructura geológica.

Keywords: Inteligencia Artificial · Redes neuronales convolucionales · Redes Transformer · autoatención .

1 Introducción

Desde el comienzo de los estudios geológicos hace más de 200 años, los geólogos han demostrado que las capas de rocas en ciertas partes de los continentes están plegadas, fracturadas y deformadas en diferentes escalas [1]. La forma, la geometría y la deformación de cuerpos rocosos son parte de lo que llamamos su estructura.

Las estructuras geológicas son la evidencia de procesos ocurridos en la litósfera terrestre por ejemplo, eventos tectónicos estrechamente relacionados a procesos corticales que causan la formación de estructuras que dan testimonio al movimiento continuo de la litósfera y la deformación que produce. Se las puede catalogar como un código a decifrar para comprender los diferentes eventos: tectónicos, volcánicos, deformativos, y procesos que actúan sobre la superficie de la Tierra, todo lo mencionado constituye la historia geológica de una zona determinada [2]. Muchos procesos son tan graduales que se necesitan enormes

* Software aplicado a la Geología

lapsos de tiempo antes de que se produzcan resultados significativos. Es importante su análisis debido a que determinan fenómenos geológicos específicos; además permiten evidenciar cambios en un ambiente de formación e interpretar posibles zonas de interés económico. Esta labor de identificación se logra interpretando las **características geométricas** distintivas que sobresalen de cada una de ellas: tamaño, límites, orientación, tipo de material, distribución geográfica [3].

De acuerdo a su origen se pueden dividir en estructuras primarias y secundarias [4]. Las estructuras primarias son aquellas que se originan durante la formación de las rocas [5], reflejan las condiciones durante la sedimentación, el vulcanismo o la intrusión, algunas son de depósito, como la estratificación cruzada [6] y otras son deformacionales, como las columnas basálticas. Las estructuras primarias pueden ayudar a reconocer la dirección de la estratificación y flujo en rocas sedimentarias y volcánicas [7]. Mientras que las estructuras secundarias son aquellas que se forman durante un proceso de deformación por actividad tectónica, después de la formación de la roca. Las estructuras geológicas fundamentales (secundarias) en la naturaleza son diaclasas, fallas, pliegues, estructuras metamórficas y zonas de corte [7].

El reconocimiento de estructuras geológicas generalmente se realiza mediante fotografías áreas en base a ciertos parámetros como apariencia, cambios principalmente de tono, textura, forma, tamaño de los rasgos, sombras [8]. En las fotográfias áreas se pueden distinguir contactos entre distintas litologías, líneas de estratificación, pliegues, columnas basálticas y lineamientos [9]. Este método requiere de experiencia sobre conceptos de fotogeología. El conocimiento teórico es fundamental para reconocer cada una de las estructuras geológicas desarolladas por procesos geológicos específicos [10]. Mediante criterios "visu", es decir la visualización en campo como parte del trabajo del geólogo; el hecho de identificar cada estrutura geológica es una destreza, que se basa en descripción de las características regionales, físicas como la geometría, cambio de ángulos, planos de dirección, alineación, esta destreza se consigue a través de la experiencia [11]. La problemática que se aborda en este trabajo involucra la falta de experiencia por parte de los estudiantes que empiezan su formación en materia de reconocimiento de estructuras geológicas, pues dicha experiencia se apoya en el conocimiento de tectónica de placas y por ende el análisis dinámico, cinemático y geométrico a escala regional, local, de afloramiento; éste tipo de análisis puede a su vez involucrar elementos adicionales de sedimentología, paleontología, petrología, geofísica y otras subdisciplinas de la geociencia [12]. El reconocimiento de estructuras geológicas exige a su vez el desarrollo de habilidades espaciales, que incluyen dos categorías, 1) la capacidad de reconocer y comprender las relaciones entre las diversas partes de una configuración y la propia posición y 2) la habilidad de generar una imagen y operar varias manipulaciones mentales sobre esta imagen. Esta categoría corresponde a la "visualización espacial" [13].

En la actualidad, la Inteligencia Artificial es una herramienta muy útil para la **identificación automática** de cualquier característica observable, con lo cual esta tarea se efectuaría de manera más ágil y precisa. En particular, las técnicas

de Deep Learning se han convertido en el método de referencia para muchas tareas [14]. En el campo de la visión por computadora, las redes neuronales convolucionales profundas logran un rendimiento de vanguardia para tareas como clasificación, detección de objetos o segmentación de instancias [15]. Además de las Redes Neuronales Convolucionales (CNN), existe un nuevo modelo de aprendizaje profundo llamado *Transformer* que ha recibido cierta popularidad, y actualmente se implementa en el ámbito de visualización, procesamiento y clasificación de imágenes, pues logra excelentes resultados en comparación con las redes convolucionales, mientras que requiere sustancialmente menos recursos computacionales [16].

El propósito del presente trabajo es brindar una herramienta basada en una CNN que permita la visualización de características, detalles y patrones que van desde elementales a complejos de cada una de las estructuras geológicas. Es de gran ayuda tanto a profesionales como a personas que se inician en el estudio y reconocimiento de estructuras geológicas (fallas, pliegues, diques, columnas basálticas, augens, pillowlavas, concreciones, ripple marks, grietas de desecación y estratificación cruzada), pues sin lugar a duda existen detalles que en una salida de campo y al estar frente a un afloramiento se escapan de la vista.

Además, ponemos a disposición una herramienta alternativa para reconocer estructuras geológicas, basada en los mecanismos de atención, a los que debe su éxito el modelo de red Transformer, que permite visualizar mapas que resaltan aquellas ubicaciones críticas, regiones fundamentales en las que se debe enfocar el estudio de estructuras geológicas

A medida que las redes neuronales convolucionales se vuelven más complejas, su funcionamiento interno se vuelve cada vez más opaco, convirtiéndose en una "caja negra" pues el proceso de toma de decisiones ya no es comprensible [14]. La visualización de mapas de características brinda información sobre el funcionamiento interno de la red, es decir cómo la imagen de entrada es percibida por la primera capa de convolución y en su paso a las capas posteriores se van extrayendo de manera jerárquica rasgos, patrones, características, desde elementales hasta complejas, con la finalidad de categorizar a la imagen original. Por otro lado, las redes Transformer basan su funcionamiento en el mecanismo de atención, para resaltar regiones poderosas al enfocarse selectivamente en partes críticas de la imagen y luego procesarlas secuencialmente [16]. La visualización de mapas de atención permitirá esclarecer el funcionamiento del mecanismo de atención con el objetivo de determinar la relación entre regiones de las imágenes que son semánticamente relevantes para la clasificación [17].

De manera concreta, nuestras contribuciones principales son:

- Un dataset de 1800 imágenes de estructuras geológicas con 9 categorías, que servirán como base para la tarea de visualización de características y mapas de atención. El dataset es de libre acceso en la plataforma de Kaggle[1].

[1] https://www.kaggle.com/datasets/mariasilvanaalvear04/estructuras-geologicas/settings?select=EstructurasGeologicas

– Un modelo de red convolucional[2] y otro de red transformer[3], ambos basados en aprendizaje automático supervisado, para la visualización y ubicación de características fundamentales en la identificación de estructuras geológicas.

El resto del documento está organizado como sigue: en la sección 2 se citan algunos de los trabajos relacionados; la sección 3 describe en detalle la metodología empleada para la elaboración del dataset y la creación de los modelos; la sección 4 explica los experimentos realizados, cuyos resultados se presentan en la sección 5, con su respectiva discusión en la sección 6. Finalmente, en la sección 7 enunciamos las conclusiones y ciertas líneas de trabajo futuro.

2 Trabajos relacionados

El uso de redes neuronales convolucionales para la visualización de características es escaso, debido a que dichas redes se han utilizado comúnmente para la clasificación. De igual manera, las redes transfomer se las utiliza como clasificador más no para observar los mecanismos de atención con los cuales trabaja la red.

Baraboshkin en el 2019 realizó la descripción automática de núcleos de perforación, basada en el análisis de distribución de color y la extracción de características, mediante redes neuronales convolucionales; presenta resultados donde se visualizó cada mapa en donde resaltan diferentes áreas y peso para cada clase de roca que categorizaron. Este método permitió observar las regiones de una imagen de roca que tienen mayor influencia en los resultados de la predicción [18].

Chen en el 2021 realiza la clasificación automatizada de estructuras rocosas utilizando imágenes geológicas en la cara del túnel de la autopista Mengzi-Pingbian en la provincia de Yunnan, China. En este estudio se utiliza el modelo Inception-ResNet-V2, siendo el modelo que presenta mejor desempeño en términos de precisión, pérdida y tiempo de prueba por imagen [19]. En los resultados lograron evaluar el riesgo geológico y análisis de estabilidad estructural.

Yang en el 2021, utiliza redes neuronales convolucionales con modificación del modelo AlexNet, para llevar a cabo la tarea de clasificar imágenes en base al volumen de roca fragmentada durante la construcción de túneles con máquinas perforadoras, con el fin de identificar grandes fragmentos de roca que corresponderían a una tunelización en mal estado pudiendo causar daños físicos a las cintas transportadoras. Los resultados presentados con la visualización de mapas de características de la red, demostró que la metodología propuesta puede detectar con precisión grandes fragmentos de roca mediante imágenes en el sitio [20].

En el análisis de clasificación automatizada de estructuras geológicas basados en datos de imágenes y modelo de aprendizaje profundo de Ye, Gang y Mingchaor en el 2018, se desarolla el clasificador utilizando un modelo Inception V3 para estructuras geológicas mediante una CNN, vecino más cercano (KNN),

[2] https://www.kaggle.com/mariasilvanaalvear04/visualizaci-n-de-caracter-sticas-mediante-cnn

[3] https://www.kaggle.com/code/mariasilvanaalvear04/atenci-n-en-redes-transformers

red neuronal artificial (ANN) y aumento de gradiente extremo (XGBoost) [21]. Los resultados del estudio concluyeron con la clasificación de cada una de las estructuras geológicas.

Según Bazi en el 2021, en lugar de utilizar las capas neuronales convolucionales, utilizan mecanismos de atención de múltiples cabezales que se evidencian en mapas de atención de diferentes capas de codificadores [17].Este mecanismo de atención permite determinar ciertos patrones y áreas de una imagen para categorizarla como tal.

Black en 2022 plantea que debido a la falta de explicabilidad o interpretabilidad de las redes neuronales profundas, se presenta un método para representar visualizaciones interpretables, dado un par de imágenes codificadas con un transformer en las cuales se muestran las regiones en las que la red mostró mayor atención [22].

Si bien existen trabajos previos en los que emplean redes neuronales artificiales para tareas de clasificación de imágenes con enfoque geológico, el entendimiento de cómo trabajan las redes es pobre, pues los estudios no muestran el proceso de extracción de características, tampoco el fundamento en el que se basan los mecanismos de atención, esto significa una gran desventaja para aquellos estudiantes principiantes en el campo de la geología, pues se busca una comprensión intuitiva de cómo trabaja una red, ya sea convolucional o transformer, y de qué manera la red percibe a las imágenes y cómo esa percepción cambia a través de la arquitectura de los modelos.

El presente estudio pretende mostrar con detalle cómo trabajan las redes neuronales convolucionales, cómo extraen características de manera jerárquica, características que son propias, indispensables y únicas de cada estructura geológica y que sin su correcta identificación no podría concluirse de manera eficiente la clasificación. También se plantea el modelo alternativo utilizando redes Transformer, con el afán de llegar a entender el funcionamiento de los mecanismos de atención, cómo estos mecanismos logran generar características poderosas al enfocarse selectivamente en partes críticas de la imagen, generando mapas de atención con valores de pesos que indican dónde debemos centrar la atención al momento de analizar estructuras geológicas.

3 Metodología

Para el reconocimiento y visualización de las características de estructuras geológicas mediante Deep Learning, se ha seguido el flujograma de trabajo presentado en la figura 1.

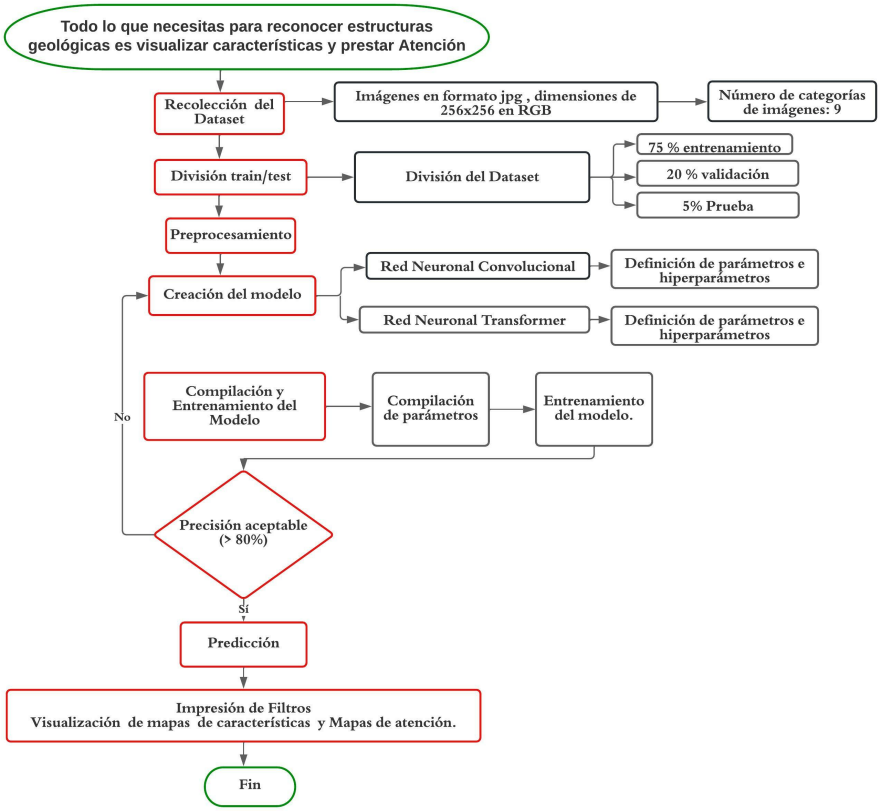


Fig. 1. Metodología del trabajo.

3.1 Recursos computacionales

Hadware Para la realización de este proyecto, se utilizaron los recursos de hadware local tipo laptop, cuyas características se describen a continuación:

Table 1. Hadware

DISPOSITIVO	
Tipo	Laptop
Modelo	HP Notebook
Procesador	Intel ® Core ™ i5-3230 M @2.60 Hz
Núcleos	4 núcleos (3ra Generación)
Procesador gráfico	4095 MB NVIDIA GeForce GTXX 1050
Tipo de sistema	Sistema Operativo de 64 bits, procesador en x64
UNIDADES DE ALMACENAMIENTO	
RAM	4.0 GB (3 GB usable)
Disco Sólido	SSD 256 GB
DISPOSITIVOS DE ENTRADA	
Mouse	Inalámbrico Anera
Teclado	P2/2 estándar
DISPOSITIVOS DE SALIDA	
Pantalla	14 pulgadas HD
COMUNICACIONES	
Conectividad	Modem Ancho de Banda 2,4 Hz

3.2 Software

Sistema Operativo
Windows 10 Home fue lanzado oficialmente al público por la empresa Microsoft en el 2015, este sistema fue implementado en una gama de dispositivos como laptos, tabletas, teléfonos inteligentes, Xbox entre otros. [23]. Entre las versiones que dispone Windows 10 Home, se tiene instalado Home Single Lenguaje, la cual posee las mismas características que la versión Home normal, pero usa solo el idioma predeterminado y no tiene la capacidad de cambiar a un idioma diferente.

Lenguaje de Programación El lenguaje implementado en la realización del proyecto es Python un lenguaje de alto nivel tipo interpretado y multiparadigma, caracterizado por tener una sintaxis fácil de aprendizaje. El Entorno de Desarrollo Integrado (IDE) que ayudó a ejecutar éste proyecto mediante el lenguaje editor de código en celdas, un compilador por celdas igualmente, además de guardar archivos generados es *Colab* el cual utiliza un entorno de Nube. Colaboratory o Colab es un producto de Google Research que permite a cualquier usuario ejecutar código arbritario de Python en el navegador, es un servicio de cuaderno alojado en Jupyter que no requiere configuración y ofrece acceso sin coste adicional. Permite el uso gratuito de GPUs y TPUs. La ventaja que Jupyter Nootebook se encuentre incorporado en Colab es que permite armar un cuaderno interactivo y al ser compartido de manera sencilla, los usuarios puedan ver los resultados sin necesidad de volver a ejecutar el código nuevamente. Aunque tiene algunas limitaciones que puede consultarse en la página de [4]. Es una herramienta

[4] https://research.google.com/colaboratory/intl/es/faq.html

ideal para el desarrollo de Machine Learningy Deep Learning sin tener que invertir en recursos de hadware.

Librerías de aprendizaje automático

- TensorFlow: es una librería OpenSource más importante de Deep Learning creada por Google. Es un conjunto de herramientas, librerías y recursos para Machine Learning, permite la compilación y entrenamiento de sus modelos de forma sencilla utilizando sus API en la cual se destaca a Keras [24] [25].
- Scikit-learn: formada por una cantidad de algoritmos de Machine Learning como clasificación, regresión, clustering.
- TorchVision va de la mano con PyTorch, util para transformaciones de imagen y vídeo. El torchvision es un paquete que consta de conjuntos de datos populares, arquitecturas modelo y transformaciones de imágenes comunes para la visión artificial [26].

3.3 Elaboración del Dataset

El principal recurso para el aprendizaje automático es el conjunto de datos, en este caso corresponde a imágenes de estructuras geológicas. Así, el primer paso fue la recopilación de dichas imágenes desde diversas fuentes como Atlas de Geología y buscadores de Internet (Image Finder [27], Pixabay [28], y Google Imágenes). Se obtuvieron 1800 imágenes organizadas en 9 categorías: augens, columnas basálticas, concreciones, dique, estratificación cruzada, grietas de desecación, pillowlavas, pliegue y ripple marks(Fig. 2).

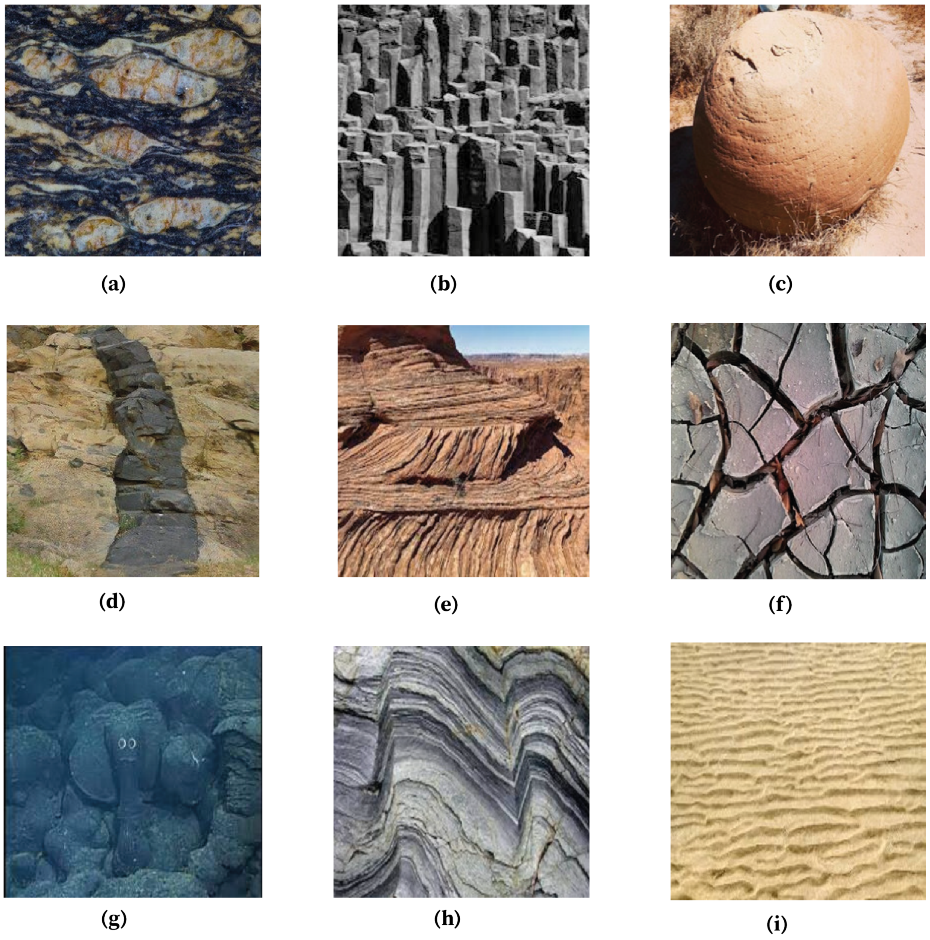


Fig. 2. Un ejemplo de cada categoría de estructuras geológicas: **a** Augen; **b** Columna Basáltica; **c** Concreción; **d** Dique; **e** Estratificación Cruzada; **f** Grietas de Desecación; **g** Pillow Lava; **h** Pliegue; **i** Ripple Marks.

Las categorías consideradas son las más frecuentes de encontrar en las salidas de campo, además su geometría ayuda al reconocimiento y extracción de características que posteriormente intervendrán para entender el contexto geológico, facilitando la comprensión de la teoría, los mecanismos y procesos geológicos que dieron origen a la formación de la estructura.

Debido a las diversas fuentes, las imágenes tienen diferente tamaño, resolución y formato (JPG y PNG). Fue necesario el uso de algunos programas de software para estandarizar tales características. Mediante *Let's Enhance* [29], se pudo mejorar la resolución y aumentar el tamaño de las imágenes a través de una tecnología basada en redes neuronales convolucionales profundas permitiendo ampliar las imágenes hasta una escala de 4x. Posteriormente, con el

fin de redimensionar las imágenes sin que se vea comprometida la calidad de las mismas, se utilizó un convertidor online denominado *Iloveimg*, llevando todas las imágenes a un mismo tamaño de 256x256 pixeles, profundidad de 24 bits, modo de color RGB, diferente resolución dada por defecto de la fuente, y un mismo formato JPG, el cual es muy ligero, permitiendo un mejor procesamiento de las imágenes en el entrenamiento. A continuación, se muestra las categorías del dataset con las siguientes descripciones: ambiente, se refiere al entorno geológico donde estas estructuras fueron formadas; origen, muestra el tiempo en que las estructuras fueron formadas [4] y la escala de estudio de las diferentes estructuras teniendo: cartográfico (m-km), afloramiento (m-cm), muestra de mano (cm) y óptico (mm-um) [30].

Table 2. Dataset de imágenes de estructuras geológicas.

Estructura Geológica	**No.**	**Ambiente**	**Origen**	**Escala**
Augen	200	Tectónico	Secundaria	Muestra de mano, óptico
Columna basáltica	200	Volcánica	Primaria	Cartográfico
Concrecion	200	Sedimentaria	Primaria	Afloramiento
Dique	200	Volcánica	Primaria	Cartográfico
Estratificación cruzada	200	Sedimentaria	Primaria	Afloramiento, muestra de mano
Grietas de desecación	200	Sedimentaria	Primaria	Afloramiento, muestra de mano
Pilow lava	200	Volcánica	Primaria	Afloramiento
Pliegue	200	Tectónico	Secundaria	Cartográfico
Ripple marks	200	Sedimentaria	Primaria	Afloramiento, muestra de mano

Por último, se aprovechó *Google Drive* para almacenar las imágenes, ya que es una plataforma gratuita que ofrece 15 GB de espacio de alojamiento, además para libre acceso al público también se encuentra disponible en la plataforma de Kaggle con el nombre "Estructuras Geológicas" que contiene las tres subcarpetas de train, validation y test, cada una con sus nueve categorías, ocupando un espacio de 69.78 MB. Las imágenes de cada categoría se ordenaron por fecha de modificación para nombrarlas con la primera palabra de la categoría y enumerarlas del 1 al 200 seguido del formato de imagen por ejemplo una imagen de la categoría de pillow lava se llamara pillow190.jpg

3.4 División en Train y Test

Una vez recopiladas las imágenes correspondientes al problema que se va a tratar, se realiza una de las tareas que caracteriza al aprendizaje automático y consiste en subdividir el dataset en tres conjuntos: entrenamiento, validación y prueba [31]. El **conjunto de entrenamiento** (Train) permite construir el modelo de reconocimiento; representa parte del "aprendizaje", donde el algoritmo analiza este conjunto para establecer un modelo matemático que represente al propio conjunto. Luego, el modelo toma el **conjunto de validación** (Validation) con el objetivo de proporcionar una evaluación imparcial del ajuste

de un modelo en el conjunto de datos de entrenamiento mientras se selecciona el mejor modelo [32]. Una vez que el rendimiento de la validación alcanza un nivel aceptable, el modelo se evalúa con el **conjunto de prueba** (Test) para verificar el rendimiento con datos totalmente nuevos. Dichos conjuntos a menudo se seleccionan en función de un muestreo aleatorio simple, el volumen y la variedad de los datos. Por lo general es preferible mantener tantos datos como sea posible para el conjunto de entrenamiento para tener un modelo más sólido [31]. La división del conjunto de datos se realizó de manera aleatoria, estableciendo un 75% para el entrenamiento, 20% datos de validación y 5% para el conjunto de prueba. Para determinar que imágenes del conjunto de datos formarán parte de cada subcarpeta, primero se genera la lista con todas las imágenes de cada categoría, después se utilizó la función *aleatorio.entre* de Excel y se generaron números aleatorios entre (1) Entrenamiento, (2) Validación y (3) Prueba; los valores resultantes se copiaron en una nueva columna manteniéndose fijos, se filtraron los valores iguales a 1, las imágenes asignadas al Entrenamiento se guardaron en la carpeta Train, se ejecutó aleatorio hasta completar el 75% del conjunto de datos, el procedimiento se repitió para la validación y las imágenes sobrantes al final se asignaron a la Prueba (Tabla 4).

Table 3. División del dataset

Categoría	Porcentaje	Imágenes por categoría	Total
Train	75 %	150	1350
Validation	20 %	40	360
Test	5 %	10	90
Total	100 %	200	1800

3.5 Preprocesamiento

Las imágenes preparadas hasta el momento serán el insumo que recibirá el modelo de aprendizaje automático. Para lograr un proceso de entrenamiento eficiente, en esta etapa es necesario generar lotes de imágenes que puedan almacenarse en la memoria de la computadora, redimensionar el tamaño y normalizar los valores de los píxeles. En la Red Neuronal Convolucional se incremento el conjunto de imágenes, generando así un dataset más extenso y capaz de una mayor generalización mediante la utilidad *ImageDataGenerator* de la librería *Tensorflow* se crean nuevas imágenes de forma artificial y aleatoria, transformadas a partir del conjunto original, utilizando los efectos: *rotation_range* para rotar las imágenes con ángulos aleatorios entre -20 y 20 grados, *zoom_range* que acerca o aleja la imagen en una determinada porción ,*width_shift_range* realiza desplazamientos horizontales, *height_shift_range* realiza cambios a la imagen de arriba hacia abajo,*horizontal_ flip* crea un reflejo de la imagen. Una vez construido, se puede crear un iterador para el conjunto de datos de imagen cargado en la memoria con

la instrucción, *flow_ from_directory* guarda el generador en una variable junto con la dirección del conjunto de datos del entrenamiento.

Las instruciones que se desriben a continuación son aplicadas a los tres subconjuntos de imágenes *batch_size = 32* indica que debe tomar grupos de 32 imágenes a partir del conjunto de datos, *label_mode = ćategorical´* especifica una etiqueta de tipo categórica ya que el dataset está dividido en 9 carpetas correspondientes a cada una de las categorías de estructuras geológicas. Con la opción *image_size = (224,224)*, se establece un tamaño uniforme para todas las imágenes en píxeles. Posteriormente, se procedió a la normalización de las imágenes con la instrucción *rescale = (1/255)* que convierten los valores de píxeles en un rango entre 0 y 1.

Para el caso de la red Transformer se definió una función con las siguientes instrucciones: *tf.random.uniform* genera valores aleatorios a partir de una distribución uniforme, *tf.image.random_flip_left_righ* a las imágenes las voltea izquierda y derecha , *tf.image.random_flip_up_down* voltea una imagen hacia arriba y abajo, *tf.image.transpose* intercambia las dimensiones de largo y ancho y *tf.image.rot* permite generar una imagen espejo, y para ajustar la saturación de las imágenes se utilizó la instrucción *tf.image.random saturation*, luego esto se guardó en un generador de imágenes mediante la instrucción de *ImageDataGenerador* aumentando la diversidad del conjunto de entrenamiento con las instrucciones dadas, luego un iterador con la instrucción, *flow_ from_directory* guarda el generador en una variable que tiene la dirección del conjunto de datos. Finalmente se procede a la normalización de las imágenes con la instrucción *rescale = (1/255)* donde se convierte los valores de píxeles en un rango entre 0 y 1.

3.6 Creación del Modelo

Para cumplir con la tarea de obtener mapas de características y mapas de atención, de las diferentes estructuras geológicas, se utilizó la técnica de *Transfer Learning*, que consiste en re-utilizar un modelo pre-entrenado con gran cantidad de datos, aprovechando al máximo el entrenamiento previo y reduciendo el tiempo de ejecución, además de utilizar parámetros establecidos de un modelo con un dominio similar y adaptarlos a nuestra necesidad con un pequeño cambio al momento de la clasificación [33].
El modelo que se utilizó para la red Convolucional fue *VGG16* [34], cuyo diagrama de arquitectura se indica en la Figura 3.

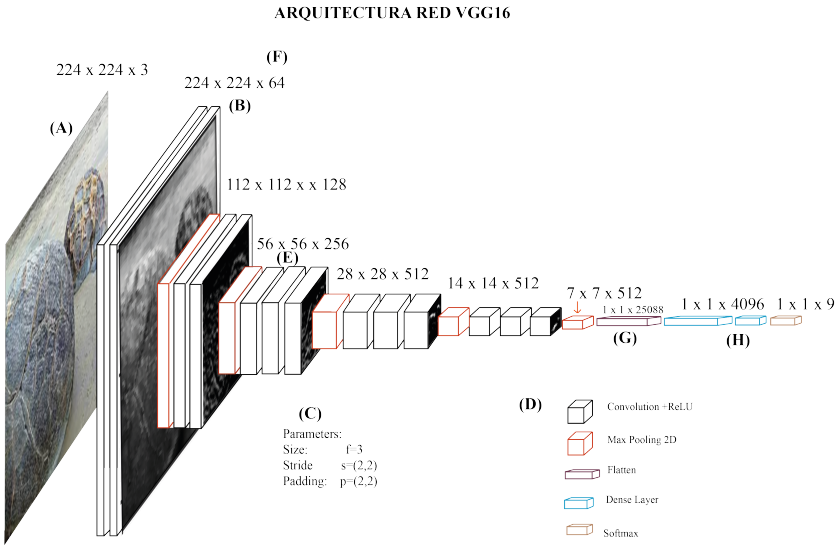


Fig. 3. Arquitectura del modelo VGG16.

La estructura y funcionamiento del modelo consiste en:

(A) El modelo recibe como entrada una imagen de (224,224,3) es decir un tamaño horizontal y vertical de 224 píxeles con tres canales RGB.

(B) Seguidamente, son 13 capas convolucionales las encargadas de extraer las características de forma jerárquica, las primeras aprenden bordes y texturas simples mientras que las posteriores aprenden patrones más complejos mediante el uso de kernels o filtros que se inicializan aleatoriamente. La convolución consiste en tomar un grupo de píxeles de la imagen de entrada (campo de recepción) e ir realizando un producto escalar con un kernel. El kernel recorrerá todas las neuronas de entrada y dará como resultado una nueva matriz, dependiendo del número de filtros se va a obtener el número de mapas de características para cada bloque de convolución. El tamaño de kernel que se utiliza para las 13 capas es de (3,3) siendo ideal para capturar las características con mayor detalle.

(C) Después de cada convolución se utiliza una operación de agrupación máxima denominada *MaxPooling*, la cual captura las características más importantes de una imagen por regiones, reduciendo así la muestra de la entrada a lo largo de sus dimensiones espaciales (alto y ancho), tomando el valor máximo sobre una ventana de entrada con un tamaño de *pool_size* de (2,2), esta operación está configurada con las siguientes funciones: zancadas o *padding* de (2,2), lo que permite agregar píxeles a los bordes de la imagen original para que al realizar la convolución, la imagen resultante sea del mismo tamaño que la imagen original para crear una red más profunda permitiendo extraer características más específicas de las imágenes durante el entrenamiento, un

stride de (2,2) obteniendo como resultado una matriz de menor tamaño en comparación con las obtenidas en la convolución original logrando reducir la cantidad de datos a procesar entre una y otra capa de la red.

(D) El modelo tiene varios bloques: el primero y segundo bloques con dos capas de convolución y una de maxpooling; el tercero, cuarto y quinto bloques tienen 3 capas de convolución y una de maxpooling.

(E) El número de filtros o kernels para las capas convoluciones son 64 para el primer bloque, 128 para el segundo, 256 para el tercero y 512 para el cuarto y quinto bloque. Como resultado se tendrá para cada filtro un mapa de características.

(F) La función de activación en cada bloque es *ReLu*, permitiendo transformar los valores negativos en cero y mantenerlos igual cuando son positivos.

(G) Una vez terminada la extracción de características con las capas convolucionales, se aplica la función *flatten()*, la cual permite aplanar los mapas de características en un vector de una dimensión.

(H) Dichas características en forma de vector son proporcionadas a un clasificador compuesto de: 3 capas densas de 4096 neuronas para la primera y segunda con activación *ReLu*, mientras que la tercera capa tiene 9 neuronas con activación *Softmax*, devolviendo la distribución de probabilidad de cada una de las clases en un rango de 0 a 1, y la probabilidad más alta será la que indique la categoría a la que pertenece la imagen.

Para el caso del modelo de red Transformer, aprovechamos el modelo de Visión Transformer *ViTBase 16* [16], cuya arquitectura se compone de una capa de embedding, un codificador y un clasificador, tal como se muestra en la Figura 4.

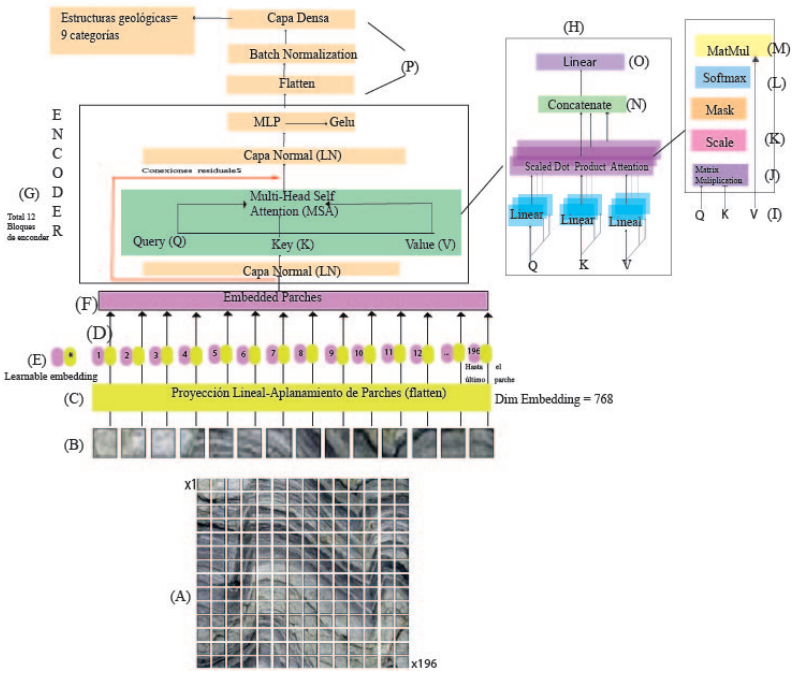


Fig. 4. Arquitectura de la Red Transformer

(A) La imagen ingresa con un tamaño de 224x224x3 (w=ancho, h=altura, c=canales), dando un total de 50176 píxeles. Seguidamente, es divida en parches, que son la entrada para este tipo de redes.

(B) Cada parche (p) tendrá un tamaño de 16x16 (256 pixeles), el cual se elige en función de $h \times w/p2$. En total se tienen 196 parches (14 parches de largo y 14 parches de ancho). Se debe tomar en cuenta que estos parches no se superponen unos con otros.

(C) Luego, para cada parche se hace una proyección lineal simple de aplanamiento (flatten) para obtener un vector denominado "Embedding". Este vector tiene una dimensión de 768, igual para cada parche, lo que es visto por el transformer como un token individual.

(D) Se agrega el codificador posicional a la representación de Embedding. Es importante este paso porque la secuencia de los parches será la misma aunque la imagen sea diferente, es decir conserva la información de posición.

(E) También se añade un *embedding learnable* adicional a la secuencia de los parches embedding. Es útil para ingresar la salida de esa dimensión en particular en la representación de la imagen del codificador.

(F) La secuencia resultante de vectores de embedding sirve como entrada para el codificador.

(G) El codificador se compone de un bloque de autoatención multicabezal, capas lineales, 12 bloques de encoder. El proceso que se explica a continuación se repetirá desde el bloque 0 hasta el bloque 11 de encoder.
(H) El Bloque de Auto-atención Multi-cabezal (MSA) es la parte central del tranformer, tiene la función de determinar la importancia relativa del embedding de un solo parche con respecto a los otros embedding en la secuencia, este bloque tiene cuatro capas: capa lineal, capa de auto-atención, capa de concatenación, la cual une las salidas de los múltiples cabezales de atención.
(I) La atención de múltiples cabezas consta de tres vectores que se pueden aprender: Q (query o consulta), K (clave o key) y V (valor o value). Esto se inspira en la recuperación de información en función de una consulta y un motor de búsqueda que compara su consulta con una clave y responde a un valor.
(J) Se hace una multiplicación de matrices Q y K, de producto punto escalares para producir una matriz de puntuación que representa cuánto tiene que atender un parche a otro, una puntuación más alta significara una atención más alta.
(K) La matriz de puntuaciones se reduce de acuerdo a las dimensiones de los vectores Q y K, para asegurar gradientes más estables.
(L) A la matriz resultante se le aplica una función Softmax y luego se la multiplica por el vector V, con la finalidad de que las puntuaciones de probabilidad más altas que el modelo haya aprendido sean las que tome como las más importantes.
(M) La salida concatenada de los vectores KQ y V se alimenta a la capa lineal para continuar su proceso.
(N) Una vez concatenados, los vectores se agregan a la conexión residual proveniente de la capa de entrada (esta conexión residual ayuda a los gradientes fluyan a través de la red).Luego se añade un MLP con activación Gelu (es útil esta activación para el desvanecimiento del gradiente).
(O) La representación resultante se pasa a una capa de normalización, lo que ayuda a reducir el tiempo de entrenamiento.
(P) Finalmente, se añade una capa de aplanamiento, Batch Normalization y un clasificador mediante una capa densa que tendrá las nueve categorías de las estructuras geológicas.

4 Experimentación

4.1 Compilación y Entrenamiento

Previo al entrenamiento de los modelos de red neuronal convolucional y transformer, mediante la instrucción *compile()* se establecieron los siguientes hiperparámetros: la función de pérdida o error denominada *categorical_crossentropy* debido a que se tiene más de dos clases en el conjunto de datos; el optimizador *Adam* para la red convolucional y *RectifiedAdam* para la red transformer, este último es una variante del optimizador Adam, cuya tasa de aprendizaje adaptativo es rectificada, la tasa de aprendizaje para la red convolucional fue de

$lr = 0.001$, mientras que para la red transfomer 0.0001; y la métrica de entrenamiento para ambas redes será almacenada en cada iteración denominada *accurancy*, es decir, la precisión del modelo.

El entrenamiento es un proceso iterativo que trata de asociar cada imagen del conjunto de entrenamiento con su respectiva categoría. Para guardar la arquitectura del modelo y los mejores pesos en un archivo con extensión *.h5*, tanto para la red convolucional como para la red transformer, se utilizó la instrucción *checkpoint*, esto permitirá cargar el modelo para la posterior evaluación.
Las opciones de configuración son los siguientes: *monitor=val_accuracy* monitoreará la métrica de precisión en la validación; *save_best_only=True* permite guardar exclusivamente el mejor modelo; *save_weights_only=False*, guardar los pesos y la arquitectura del modelo, *mode='max'*, permite guardar el mejor modelo de las cantidades monitoreadas. En la red transformer además se añadió la instrucción *ReduceLROnPlateau()* para reducir la tasa de aprendizaje cuando la métrica ha dejado de mejorar; *EarlyStopping()* para que se detenga el entrenamiento cuando una métrica monitoreada no mejore.
El entrenamiento se realiza mediante la instrucción *model.fit*, el número de épocas o iteraciones completas del conjunto de entrenamiento fueron de 20 para la red Convolucional con un tiempo de ejecución de 15 minutos, mientras que para la red Transformer con 25 épocas la ejecución duró 35 minutos.

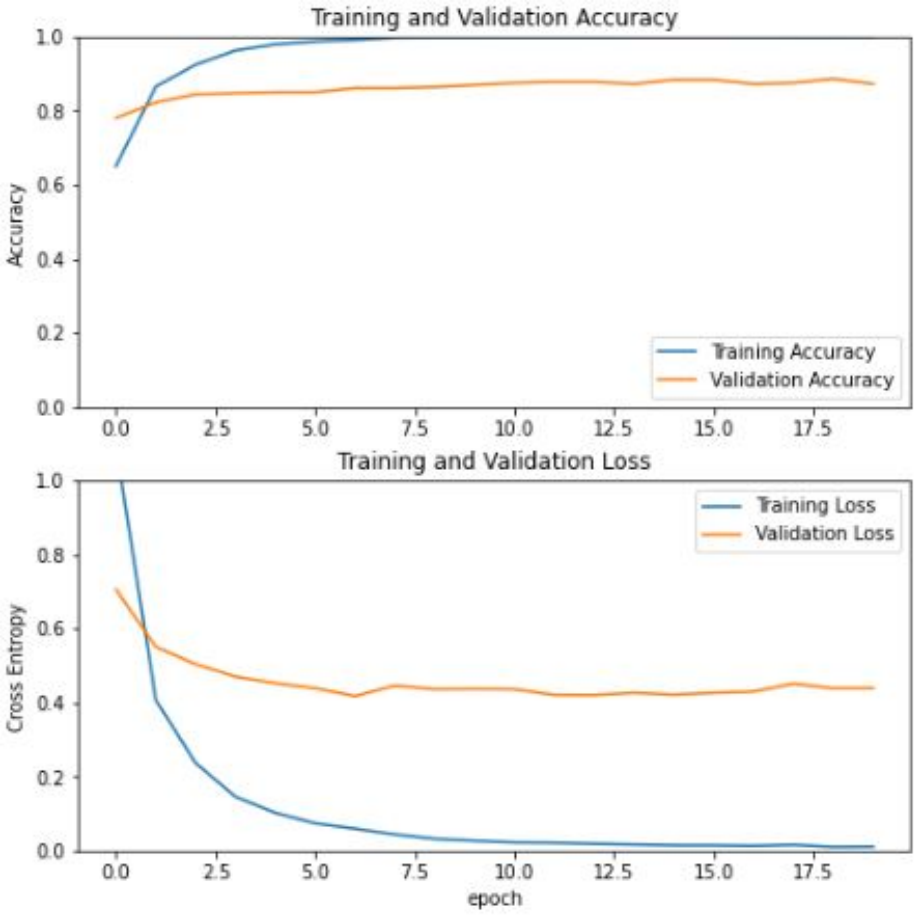


Fig. 5. Arriba, gráfica de precisión y pérdida del entrenamiento y Abajo, gráfica de precisión y pérdida de validación del modelo de Red Neuronal Convolucional.

En el caso de la red convolucional (Figura 5) se puede observar como la precisión del entrenamiento inicia con un valor de 0.6504 y alcanza el valor máximo en la época 11 donde la tendencia tiene forma de meseta lo que indica que el modelo deja de mejorar en esa iteración, la curva de pérdida muestra una tendencia a cero alcanzando finalmente el valor de 0.0104 en la época 20, estos valores se consideran como aceptables para el modelo. Para visualizar el comportamiento del entrenamiento y validación del modelo se realizó el ploteo de los valores de pérdida y precisión del entrenamiento y de la validación (Figura 5) y (Figura 6) a través del historial que se va almacenando durante el proceso y las instrucciones de *plt.figure()* y *plt.show()*.

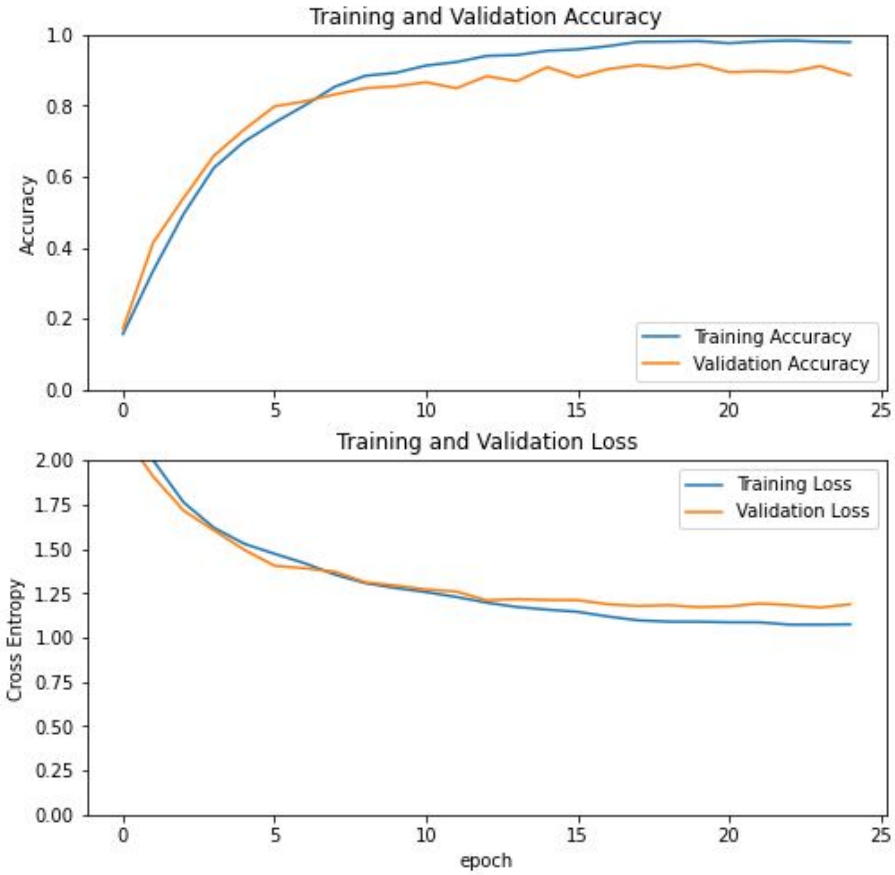


Fig. 6. Arriba, gráfica de precisión y pérdida del entrenamiento y Abajo, gráfica de precisión y pérdida de validación del modelo de la Red Transformer.

En las curvas de aprendizaje de la red transfomer (Figura 6) se observa como la precisión del entrenamiento tiende a 1, iniciando en 0.1562 y llegando a un máximo de 0.9841 en la época 23, mientras que la pérdida inicia en 2.4367 y termina en 1.0744, los valores aquí obtenidos se consideran aceptables.

Table 4. Comparación de pérdida y precisión del Entrenamiento de las redes

Red Convolucional			
Precisión época 1	Precisión época 20	Pérdida época 1	Pérdida época 20
0,6504	0,9993	1,0944	0,0104
Red Transformer			
Precisión época 1	Precisión época 25	Pérdida época 1	Pérdida época 25
0,1562	0,9795	2,4367	1,0744

4.2 Evaluación del Modelo

Para la realización de la evaluación del modelo mediante la instrucción *model.evaluate()* y utilizando imágenes del *conjunto de test*, se obtiene como resultado para la Red Neuronal Convolucional una precisión de 0.8556 y pérdida de 0.4707. En el modelo de Red Transformer se obtuvo una precisión de 0.8944 y pérdida de 1.1790. Los valores aquí obtenidos se consideran aceptables por tanto se puede continuar con la predicción de los modelos.

4.3 Predicción del Modelo

Para el modelo Convolucional se carga el modelo entrenado en formato .h5 y se utiliza la instrucción *model.predict*, se procede a realizar la predicción de una imagen fuera del dataset y se la carga desde el equipo, en la Figura 7 se observa el resultado de la predicción mediante un gráfico de barras en el que cada barra representa el valor de probabilidad que el modelo asigna a cada categoría.En este caso la predicción para una grieta de disecación es correcta.

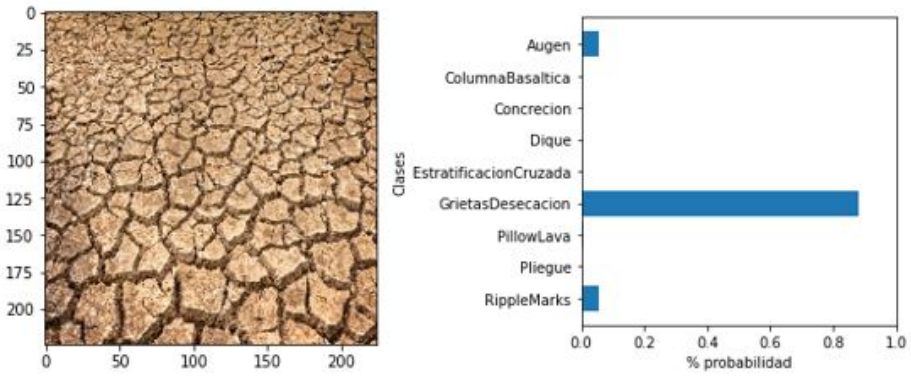


Fig. 7. Se muestra la imagen original junto al gráfico de probabilidad para cada categoría

Para el modelo Transformer se carga el modelo ya entrenado y guardado en formato .h5, mediante el comando *model.predict*, se lleva a cabo la predicción

utilizando una imagen que no se ha incluido en el dataset, cargada desde el equipo, en la figura 8 se observa el resultado mediante un gráfico de barras en el que cada barra representa el valor de probabilidad que el modelo asigna a cada categoría. En este caso la predicción para una concreción es correcta.

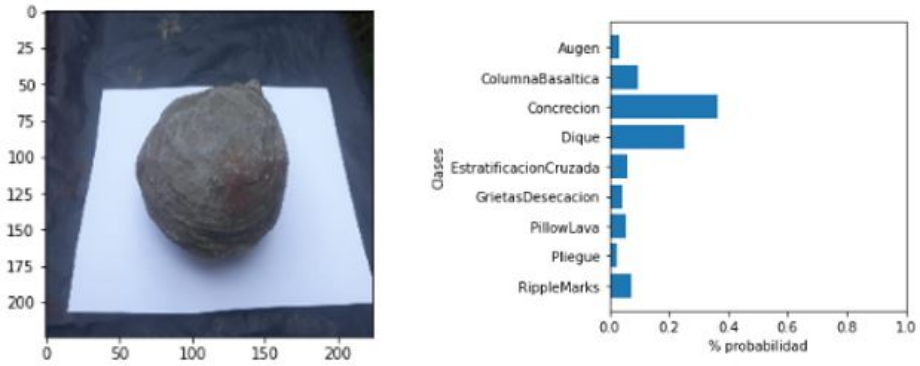
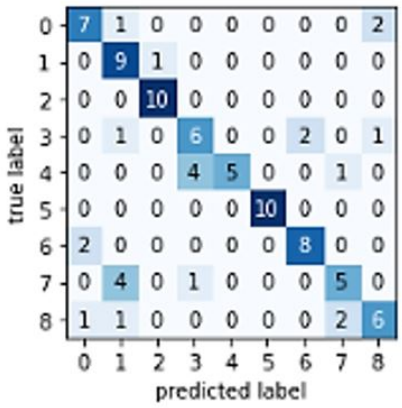


Fig. 8. Se muestra la imagen original junto al gráfico de probabilidad para cada categoría.

4.4 Matriz de Confusión

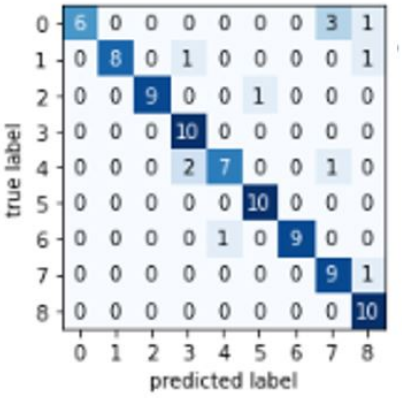
Para un mejor detalle de los resultados se implementa la matriz de confusión tanto para el modelo de Red Convolucional como para el modelo Transformer, está matriz es una tabla resumida, utilizada para poder evaluar el rendimiento del modelo de clasificación a partir del *Conjunto de Prueba*.



	precision	recall	f1-score	support
0	0.7000	0.7000	0.7000	10
1	0.5625	0.9000	0.6923	10
2	0.9091	1.0000	0.9524	10
3	0.5455	0.6000	0.5714	10
4	1.0000	0.5000	0.6667	10
5	1.0000	1.0000	1.0000	10
6	0.8000	0.8000	0.8000	10
7	0.6250	0.5000	0.5556	10
8	0.6667	0.6000	0.6316	10
accuracy			0.7333	90
macro avg	0.7565	0.7333	0.7300	90
weighted avg	0.7565	0.7333	0.7300	90

Fig. 9. A la derecha la Matriz de Confusión del modelo CNN, a la izquierda muestra las métricas de precisión, recall y acurracy determinadas a partir de la matriz de confusión

El número de predicciones correctas e incorrectas se resume en la figura 9 y 10 con un conteo y se muestra gráficamente para cada clase, en este caso se estableció números para cada categoría :0 Augen', '1 Columna Basáltica', '2 Concreción', '3 Dique', '4 Estratificación Cruzada', '5 Grietas de Desecación','6 Pillow Lava','7 Pliegue', '8 Ripple Marks'



	precision	recall	f1-score	support
0	1.0000	0.6000	0.7500	10
1	1.0000	0.8000	0.8889	10
2	1.0000	0.9000	0.9474	10
3	0.7692	1.0000	0.8696	10
4	0.8750	0.7000	0.7778	10
5	0.9091	1.0000	0.9524	10
6	1.0000	0.9000	0.9474	10
7	0.6923	0.9000	0.7826	10
8	0.7692	1.0000	0.8696	10
accuracy			0.8667	90
macro avg	0.8905	0.8667	0.8651	90
weighted avg	0.8905	0.8667	0.8651	90

Fig. 10. A la derecha la Matriz de Confusión del modelo Transformer y a la izquierda muestra las métricas de precisión, recall y acurracy determinadas a partir de la matriz de confusión

En la figura 9 y 10 , en la diagonal principal(tonalidades oscuras) se encuentran las predicciones correctas, al tomar una categoría de las 9 se la denom-

ina *verdadero positivo* y lo restante que no corresponde a esa categoría pero la predicción es correcta se denomina *verdadero negativo* . Mientras, el falso positivo y falso negativo son resultado de cuando el modelo predice incorrectamente, clasificando a la imagen de una estructura geológica dentro de una categoría a la que no corresponde. Se han obtenido las métricas de precisión, recall , f1-score, estos se obtienen mediante fórmulas utilizando valores: verdadero positivo, falso positivo y falso negativo, estos valores de desempeño indican que los modelos de clasificación de estructuras geológicas tiene un buen comportamiento para cada una de las categorías mostradas en las figuras 9 y 10.

5 Resultados

5.1 Visualización de filtros con el modelo convolucional

La extracción de características de las imágenes es posible a través de los filtros, estos son como pequeñas ventanas que recorren la imagen horizontal y verticalmente haciendo el proceso de convolución. Para imprimir los filtros de las convoluciones se utilizó la instrucción *model.layers* la cual fue configurada para que solo extraiga las capas convolucionales como muestra la figura 11.

```
1 block1_conv1 (None, 224, 224, 64)
2 block1_conv2 (None, 224, 224, 64)
4 block2_conv1 (None, 112, 112, 128)
5 block2_conv2 (None, 112, 112, 128)
7 block3_conv1 (None, 56, 56, 256)
8 block3_conv2 (None, 56, 56, 256)
9 block3_conv3 (None, 56, 56, 256)
11 block4_conv1 (None, 28, 28, 512)
12 block4_conv2 (None, 28, 28, 512)
13 block4_conv3 (None, 28, 28, 512)
15 block5_conv1 (None, 14, 14, 512)
16 block5_conv2 (None, 14, 14, 512)
17 block5_conv3 (None, 14, 14, 512)
```

Fig. 11. Enumaración de cada capa convolucional con su respectivo bloque.

Posteriormente se imprima los filtros y sus respectivos pesos y sesgos; en la primera capa convolucional se dispone de 64 filtros de tamaño 3x3, para cada canal de color RGB, es decir, un total de 64 x 3 filtros de tamaño 3x3. Se grafican estos filtros, cada canal de color en una nueva columna. En la Figura 12 se muestra los 3 primeros filtros de cada canal, además se ha recuperado los pesos de los filtros con la instrucción *model.layers[1].get_weights()* y fueron normalizados en un valor de 0 a 1 para su fácil visualización.

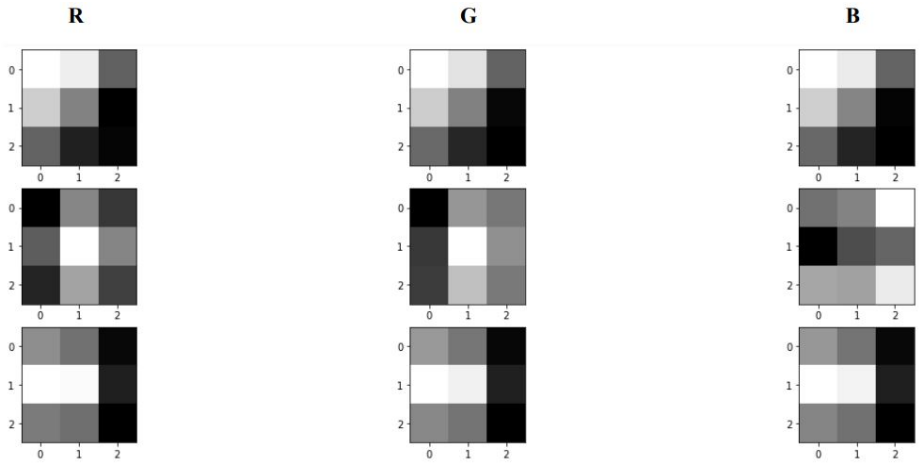


Fig. 12. Filtros pertenecientes a la primera capa convolucional del Bloque 1, cada columna corresponde a los tres canales que componen la imagen de entrada (RGB).

Se puede observar en la Figura 12 que, para la primera fila, el filtro es el mismo en todos los canales, y para el caso de las otras filas, los filtros difieren. Cada filtro está compuesto por cuadrados oscuros que indican pesos pequeños y cuadrados claros que representan pesos grandes.

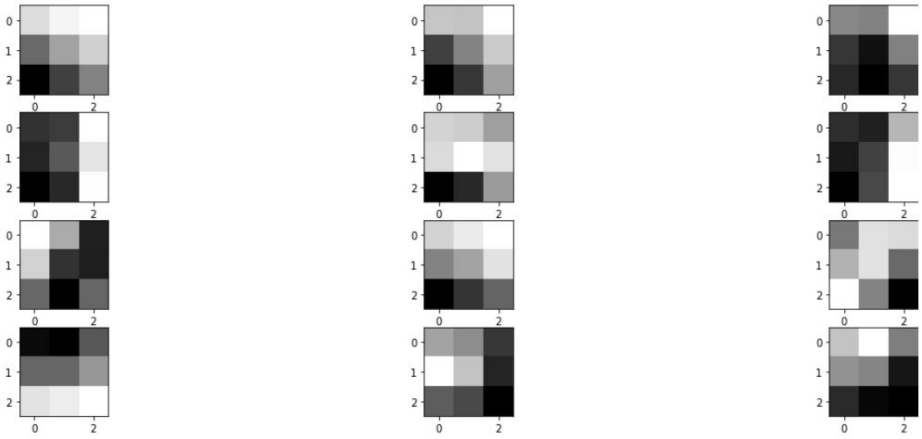


Fig. 13. Filtros pertenecientes a la primera capa convolucional del Bloque 3.

5.2 Visualización de mapas de características con el modelo convolucional

El proceso de generación de mapas de características se llevó a cabo en imágenes de las 9 categorías, la extracción de los mapas correspondientes a las últimas capas de convolución de los 5 bloques confirman lo que dice la teoría, pues se ha logrado evidenciar como el modelo aquí utilizado dada una imagen de entrada, extrae las características de manera jerárquica, como se detalla a continuación:

Bloque 1: Capa de convolución 2 En la figura 14 se puede observar los mapas características que resultan de la primera capa convolucional, la cual extrae características primitivas de la imagen, como bordes orientados, diferencia de tonalidades, marca las zonas de luz y oscuridad, se aplican efectos de saturación, contraste, con el objetivo de resaltar ejes. Para este caso, en el pliegue se puede diferenciar patrones de color, se intensifica la tonalidad y la diferencia de contraste entre capas de diferente composición, es muy evidente la orientación aproximadamente paralela en los ejes que conforman las capas del pliegue.

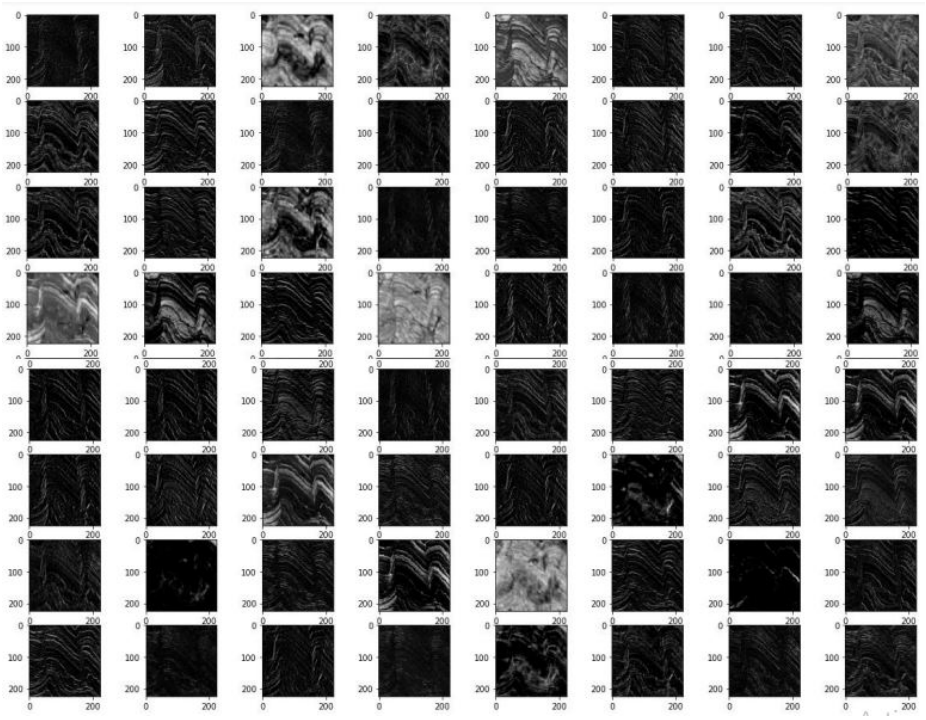


Fig. 14. Mapas de características extraídos de la última capa de convolución del Bloque 1.

Bloque 2: Capa de convolución 5

La figura 15 muestra los mapas de características resultantes de la segunda convolución, en ella se extraen patrones texturales, vértices, además de realizar la combinación de bordes formando líneas rectas, curvas, paralelas. Para este caso se observa una textura corrugada que sobresale evidenciando el plano del eje de pliegue, la textura granular se relaciona con la composición de materiales que lo conforman. Las líneas curvas y paralelas se relacionan con cada una de las capas, los vértices representan los puntos de máxima curvatura (charnela) así como los bordes exteriores determinan sus flancos.

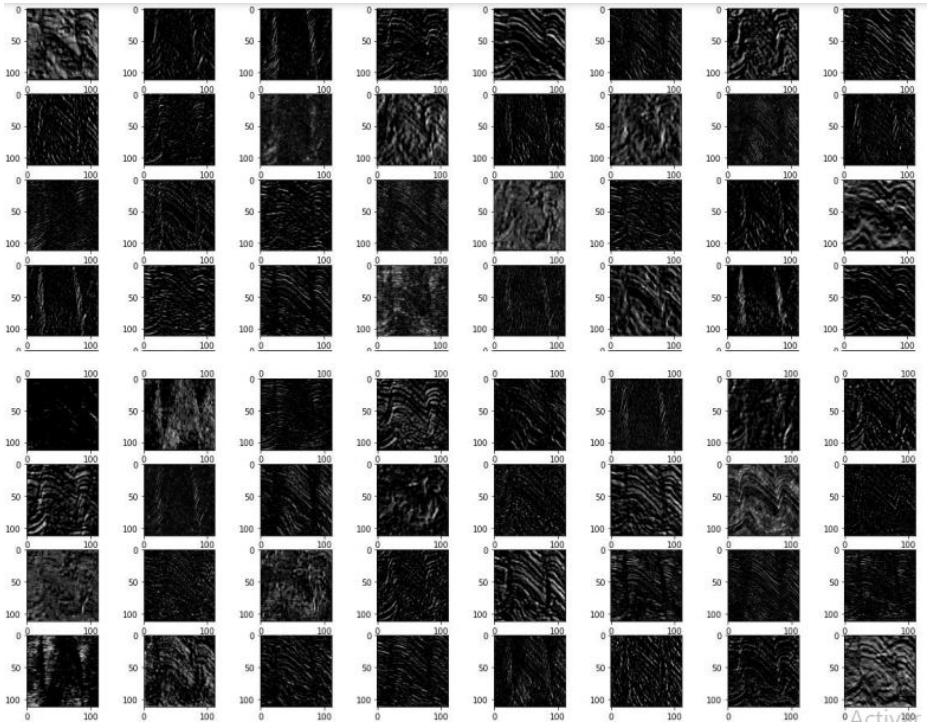


Fig. 15. Mapas de características extraídos de la última capa de convolución del Bloque 2.

Bloque 3: Capa de convolución 9

En la figura 14 se puede observar cómo a medida que la red alcanza mayor profunidad la dimensión de los mapas se reduce, pues luego de cada convolución, la capa de MaxPooling se aplica para lograr localizar las características más importantes para clasificar a la imagen sin importar la posición, rotación, tamaño de los mapas de entrada, se extraen texturas más complejas es así que, en la estructura del pliegue se analiza las características más complejas y a la vez más relevantes, por ejemplo, en éste punto se puede observar como una parte de las

capas se dividen de acuerdo a su espesor, siendo evidente su curvatura a causa de la deformación, el plano axial se muestra como una línea vertical claramente marcada que además indican inclinación mostrando su disposición geométrica. Por otro lado, las texturas junto con la variación de tonalidades muestran los diferentes estratos que conforman el pliegue.

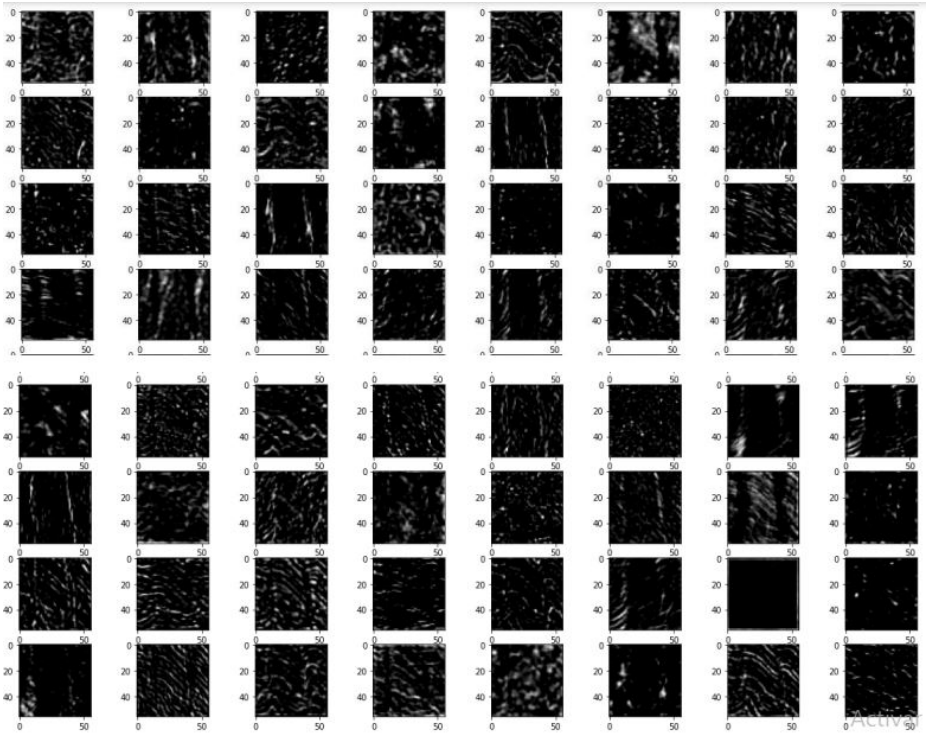


Fig. 16. Mapas de características extraídos de la última capa de convolución del Bloque 3.

Bloque 4: Capa de convolución 13

Para la visión humana las características obtenidas con esta convolución pueden parecer borrosas e impercetibles como se muestra en la figura 17, pero para la red la localización de esas características esenciales es bastante clara, se ubican regiones especificas propias y únicas de cada esctructura geológica. En el ejemplo se observa como la red localiza y resalta la forma del plano axial y marca aquellas capas de mayor espesor dentro del pliegue.

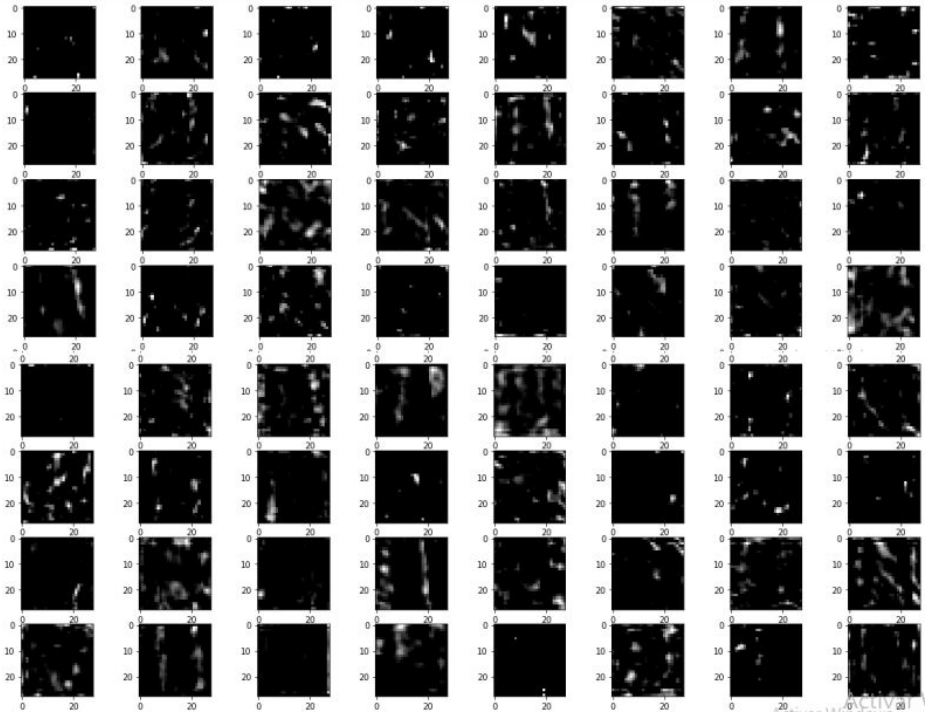


Fig. 17. Mapas de características extraídos de la última capa de convolución del Bloque 4.

Bloque 5: Capa de convolución 17

En esta última capa figura 18 la red localiza aquellas regiones que son imprescindibles para identificar una estructura geológica, se muestra el efecto acumulativo de las convoluciones anteriores, pero con menor resolución espacial. En este ejemplo, la red ubica y resalta parte del plano axial, representadose como puntos que corresponden a la máxima curvatura (charnela) del pligue.

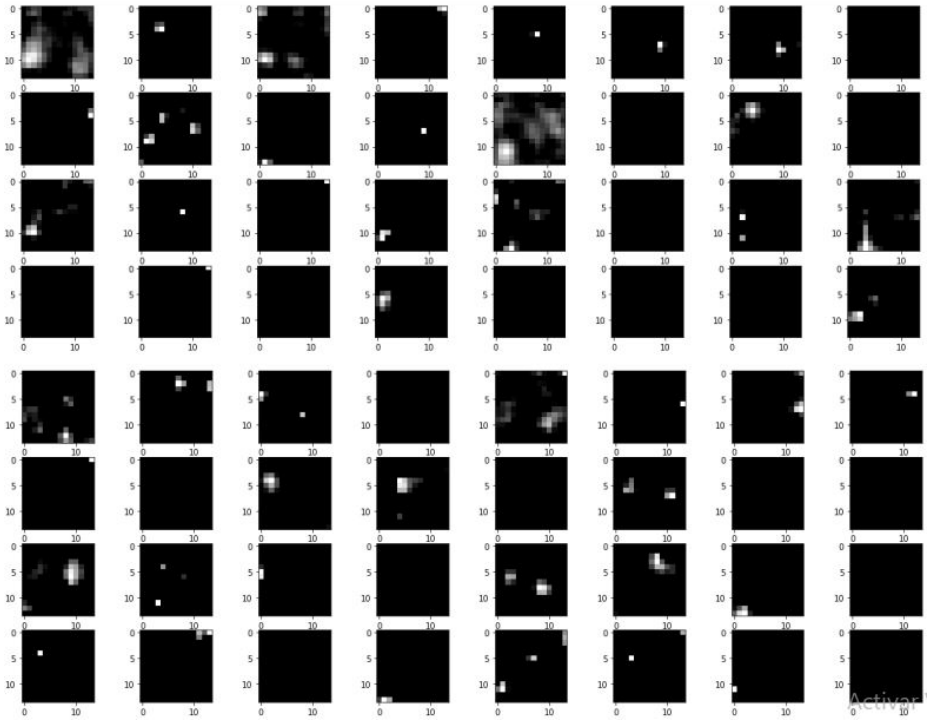


Fig. 18. Mapas de características extraídos de la última capa de convolución del Bloque 5.

5.3 Visualización de mapas de atención con la red Transformer

El modelo utilizado en este trabajo consta de 12 codificadores y cada uno se compone de *12* cabezales, la atención de múltiples cabezales permite que el modelo atienda conjuntamente la información de diferentes subespacios de representación en diferentes posiciones, es decir cada cabezal codifica la representación semántica consciente del contexto para obtener regiones discriminatorias que son necesarias para una clasificación precisa a nivel de parche.

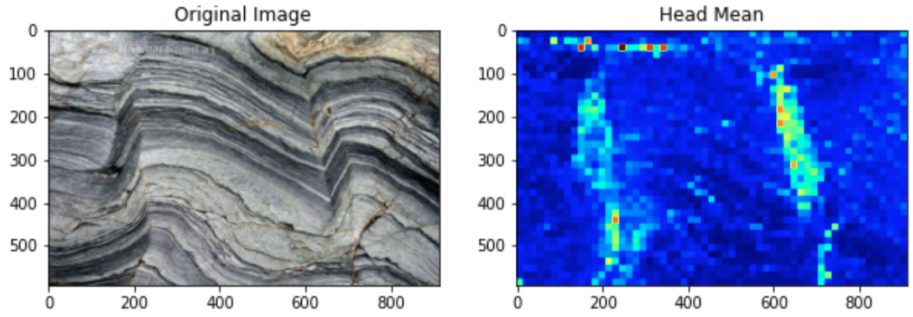


Fig. 19. Matrices de atención extraídos del Bloque Codificador 0, tonalidades verdosas y rojizas para ponderaciones de atención alta y baja para tonos en azul).

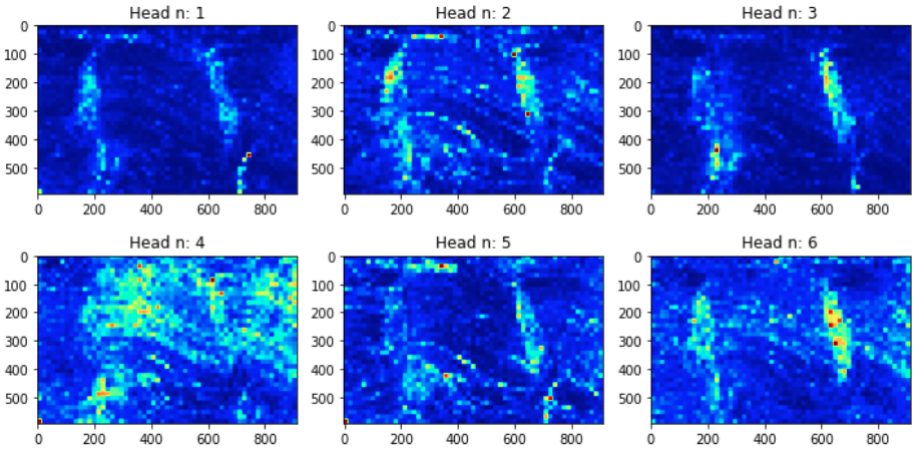


Fig. 20. Mapas de atención extraídos del Bloque Codificador 0 de una imagen aleatoria de test, tonalidades verdosas y rojizas para ponderaciones de atención alta y baja para tonos en azul).

Cada cabezal de auto-atención aprende las ponderaciones de atención en base al producto punto escalado, mismo que se calcula entre el vector Q (Consulta) de este elemento con los vectores K(Clave) de otros elementos, el resultado del producto punto se asigna al vector V (Valor) quién determina la importancia relativa de los parches en la secuencia.

En la figura 20 se puede observar el resultado del primer bloque Codificador (0), con sus respectivas *Head Attention*, la Head 1 centra la atención en los planos axiales de cada pliegue, la Head 2 y 3 centra la atención en los puntos de inflexión en los que el pliegue pasa de ser un anticlinal (cóncavo) a un sinclinal

(convexo), la Head 4 muestra claramente toda la zona central de los pliegues y los diferentes tamaños de capa, la Head 5 centra la atención en las fracturas presentes en la zona de debilidad de las capas, la Head 6 integra la atención de la Heads anteriores y realza las zonas de eje de pliegue que son las más importantes de la estructura entrenada.

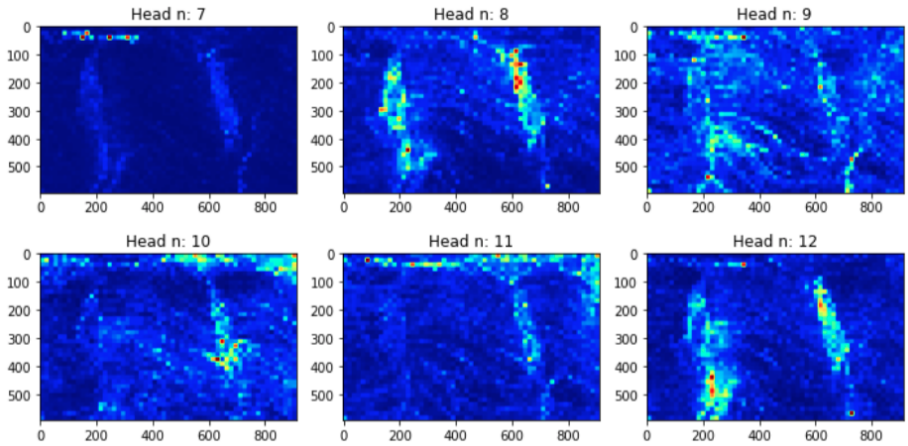


Fig. 21. Mapas de atención extraídos del Bloque Codificador 0, tonalidades verdosas y rojizas para ponderaciones de atención alta y baja para tonos en azul).

En la figura 21 se puede visualizar como la autoatención se utiliza para integrar información de la imagen, esto se relaciona con la distancia de atención que presta la red sin importar que tan lejos esté un parche del otro. En esta figura se observa el resultado de las seis heads resultantes del primer bloque codificador, la Head 7 y 8 centra la atención en ambos planos axiales y ya revela nueva información al realizar las sumas ponderadas del analisis de los heads previos, la Head 9 fija su atención en las partes no observadas en ninguno de los heads previos, la Head 10 y 11 enfoca la atención en el pliegue derecho y las zonas superiores que contienen información no reelevante de la imagen, y la head 12 finaliza con el realce de las zonas más importantes de la estructura, muy similar al cabezal 6, pero contrastando las partes de la imagen con menos reelevancia encontradas durante el entrenamiento como las partes exteriores del pliegue y las capas de la parte inferior.

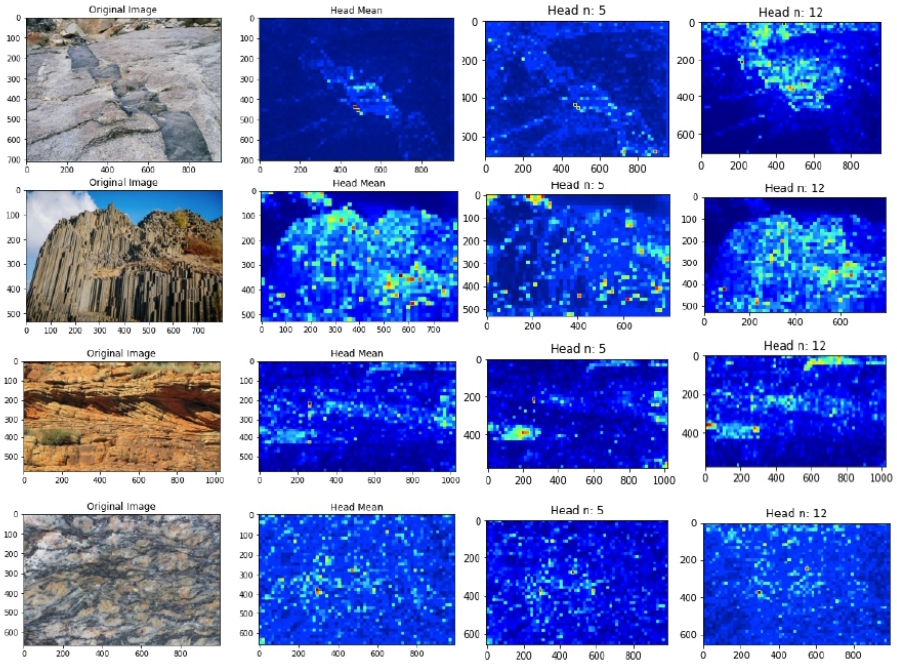


Fig. 22. Mapas de atención extraídos de imágenes del dataset "test", de arriba hacia abajo se observan Diques, Columnas Basálticas, Estratificación Cruzada y Augens

En la figura 22 y 23 se observan ejemplos de varias imágenes donde se han extraido los mapas de atención, los cuales pertenecen al primer bloque codificador con tres heads analizados, de arriba hacia abajo se expone a los diques, donde se observa la atención de la red exactamente a la estructura presente en el centro de la fotografía, en las columnas basálticas las zonas de atención se concentran en las columnas prismáticas originadas por las coladas de lava que destacan claramente, la estratificación cruzada evidencia atención en las capas depositadas en diferentes ángulos, los augens denotan claramente la atención en los cristales deformados, los ripple marks en las formas que denotan en el suelo, las grietas de desecación en sus respectivas zona de debilidad que agrietan los sedimentos, los pliegues marcan las principales zonas de atención en sus respectivos planos axiales como se analizó en la figura 21 y las concreciones en sus estructuras esféricas características.

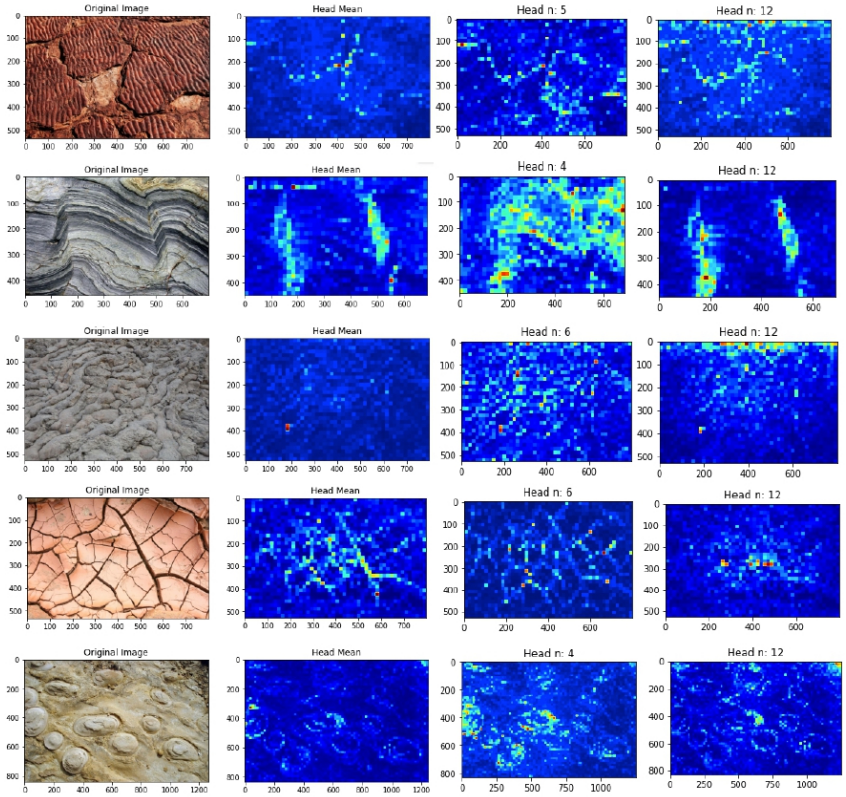


Fig. 23. Mapas de atención extraídos de imágenes del dataset "test", de arriba hacia abajo se observan Ripplemarks, Pliegues, Pillow Lavas, Grietas de desecación y Concreciones

6 Discusión

Utilizamos la técnica de Transfer Learning para ajustar nuestro conjunto de datos a modelos preentrenados que han dado resultados de aprendizaje aceptables. Primero, presentamos una forma de visualizar el funcionamiento interno de la red convolucional, extrayendo de manera jerárquica características propias de cada estructura geológica, se presentan mapas de características que ayudarán al estudiante que inicia su estudio en el fascinante mundo de la Geología, a reconocer ejes, patrones, formas, objetos, geometrías, tamaños, texturas, que a su vez estarían relacionados a procesos: téctonicos, volcánicos, intrusivos y sedimentarios.

Segundo, utilizando los mecanismos de atención de una red Transformer, se ha logrado obtener mapas de atención que resultan de operaciones matemáticas entre vectores generados por la red durante el entrenamiento, éstos mapas ayu-

dan a indentifcar regiones relevantes a nivel de parche, en las que se debe centrar la atención al momento de reconocer estructuras geológicas.

Finalmente, los resultados aquí presentados, difieren en que, la red neuronal convolucional muestra mapas características desde elementales a complejas, mismas que identifican a una estructura como tal, es decir formas, patrones, objetos que son propios y únicos de cada estructura geológica. Mientras que de la red Transformer se obtienen mapas de atención que exhiben zonas o regiones de imporntancia en las que se debe enfocar la atención. Haciendo una comparación entre los productos obtenidos se evidenció que las características más importantes que extrae la red convolucional se corresponden con las regiones donde la red transformer centra la atención. En geología, tanto como para el profesional como para el estudiante, la visualización de geometrías, texturas, alineaciones, orientación, tamaño, es parte fundamental para el reconocimiento de cada una de las estructuras geológicas, por lo tanto consideramos que la red convolucional proporciona una herramienta de mayor aplicabilidad al momento de aprender a reconocer estructuras geológicas.

7 Conclusiones

El proceso sobre el funcionamiento interno de cada una de las redes para el manejo de imágenes es diferente; mientras las redes neuronales convolucionales trabajan con filtros en cada una de sus capas para extraer de forma jerárquica, características simples a complejas, las redes transfomer mediante el mecanismo de autoatención van puntuado con altos pesos de atención en zonas o regiones propias e imprescindibles de cada estructura geológica.

En la evalación, yanto las redes convolucionales como las redes transformer alcanzan precisiones mayores al 80% con 20 y 25 épocas respectivamente y mediante al matriz de confusión se puede interpretar una mejor respuesta de las redes transformer sin embargo las diferencias con la red neuronal convolucional no son tan marcadas.

La implementación de estas redes representa significativa importancia en el campo geológico, pues además de categorizar las estructuras como tal, se puede ir observando rasgos, patrones, detalles, zonas donde se focaliza la red para distinguir a cada categoría de las estructuras geológicas.En el siguente enlace [5] se puede encontrar de manera detallada la descripción de los resultados obtenidos con los mapas de visualización para cada cateogría, para el caso de: **augen**, en las primeras capas de la red se obtiene trazas y formas en la que se presentan los lineamientos, así como la dirección principal de la estructura, también diferencias de tonalidad, conforme avanza la red se observa una forma de ojo que es la característica principal de esta estructura; **columna basáltica**, las características primitivas que se extrae de la red corresponden a formas y ejes de

[5] https://www.kaggle.com/datasets/vivianachangoluisa/mapas-de-caractersticas-de-estructuras-geolgicas

un prisma, conforme avanza la red se tiene un conjunto de lineas inclinadas , paralelas y puntos que corresponden a los vértices de los prismas que se forman; **concreción**, las primeras capas de convolución muestran bordes redondeados y líneas propias de una concreción con texturas irregulares, mientras se profundiza en la red se extrae solo los bordes redondeados dejando de lado las líneas rectas, las características más complejas de ésta estructura corresponden a zonas donde la concreción comprende una esfera completa con una textura lisa. **dique**, en la primeras capas de convolución se observa la forma tabular e inclinada, luego se extrae solamente líneas verticales que corresponden a los límites del dique, una textura que pasa de lisa a corrugada; en el caso de la **estratificación cruzada** se evidencia líneas inclinadas de cada estrato en contacto con líneas subhorizontales que caracterizan esta estructura, a media se profundiza en en la red las características evidencian cambios de textura corrugada a lisa. En las capas finales la red localiza el plano de truncamiento propio de una estratificación; **grieta de desecación** la red se enfoca en líneas que se intersecan entre sí dando la apariencia de textura enrejada, en las ultimas capas de la red se identifican límites entre cada grieta; **pilow lava** las características primitivas de esta estructura son las formas redondeadas y una textura rugosa que se asocia con su ambiente y condiciones de formación seguida de características más complejas como bordes redondeados, finalmente la particularidad que las distingue de otras estructuras son bordes que se redondean y textura corrugada; Finalmente para un **ripple marks** en las primeras capas se extraen bordes que forman ondículas, conforme se profundiza en la red se extraen líneas sinuosas y en las ultimas capas de la red se muestra las crestas de laminación siendo la particularidad que distingue a esta estructura de las demás.
Con la red Transformer se ha obtenido buen resultado en la tarea de clasificación de estructuras geológicas, y los mapas de atención son una herramienta muy valiosa para identificar aquellas regiones que son relevantes para identificar cada una de las estructuras geológicas, en este trabajo se reconoce que existe correspondencia entre aquellas características que las redes convolucionales extraen en sus capas más profundas y las regiones en las que la red transformer focaliza la atención.

8 Trabajos futuros

En trabajos futuros se puede ampliar la aplicación de visualización de características a otras áreas de la geología como: en geofísica con el reconocimiento de estructuras en perfiles sísmicos; paleontología para visualizar características de cada fósil; geomorfología para identificar las geoformas de los distintos relieves e incluso a otras ciencias como la botánica donde uno de sus aspectos principales es la visualización de bordes, texturas y patrones más relevantes que identifican a las plantas.
Para una interacción más amigable con el usuario se recomienda la creación de un sitio web que pueda implementar el código o la combinación de los dos códigos

presentados en este trabajo y así el usuario solo tenga que subir una imagen y con un solo clik obtener mapas de características y/o mapas de atención.

References

1. Hamblin, K., Christiansen, E.: Earth's Dynamic Systems. 10nd edn. Pearson, New York (2004)
2. Tarbuck, E., Lutgens, F.: Ciencias de la Tierra. Una Introducción a la Geología física 1st edn. Pearson, Madrid-España (2005)
3. Fossen, H.:Structural geology and structural analysis. 2da edn. Cambridge University Press,Reino Unido (2016)
4. Russell,W.:Structural Geology for Petroleum Geologists. 1ra edn. McGraw-Hill Book Co., New York (1955)
5. Waldron,J.,Snyder,M.,: Geological Structures: a Practical Introduction. Earth and Atmospheric Sciences, University of Alberta (2020)
6. Nichols,G.,: Sedimentology and Stratigraphy. Blackwell Publishing.Second Edition. (2009)
7. George,H.,Stephen,J.,: Structural Geology of Rocks and Regions.(2011)
8. Guerra, F.: Las doce principales reglas de la interpretación fotogeológica y bases fundamentales de que se derivan. Investigaciones Geográficas, Boletin del Instituto de Geografía UMAN, **42**(60) 42-66 (2003)
9. Echeveste, H.: Manual de levantamiento geológico. Catédra de Levantamiento Geológico Facultad de Ciencas Naturales, Museo Universidad de la Plata (2017)
10. Benninson, G.: An Introduction to Geological Structures and Maps. The Library University of Birmingham Edgbaston, Birmingham 5ed (1990)
11. Pozo, M.: Gelogía Práctica. Introducción de reconocimiento de materiales y análisis de mapas. Pearson Prentice Hall
12. Haakon, F.: Structural Geology. Cambridge University Press.(2010)
13. Kali, Y.,Orion, N. . Spatial abilities of high-school students in the perception of geologic structures. Journal of Research in Science Teaching, 33(4),369–391.(1996)
14. Zurowietw, M., Nattemper,T.: An Interactive Visualization for features localization in Deep Neural Networks **3**(49) 1-11 (2020)
15. Zeiler, M.: Visualizing and Understanding Convolutional Networks.ArXiv.1-11 (2013)
16. Dosovitskiy,D.: An Image is worth 16x16 words: Transformers for image recognition at scale.ArXiv.1-22 (2021)
17. Bazi, Y.,Bashmal, L.: Vision Transformers for Remote Sensing Image Classification.Remote sensing.1-19 (2021)
18. Baraboshkin, E., Ismailova, L .: Deep Convolutions for In-Depth Automated Rock Typiing ArXiv.(2019)
19. Chen, J.: Deep learning based classification of rock structure of tunnel face. Geoscience Frontiers **12** 395-404 (2021)
20. Yang, Z.: Classification of rock fragments produced by tunnel boring machine using convolutional neural networks. Automation in Construction **125** (2021)
21. Ye, Z., Gang,W.: Automated Classification Analysis of Geological Structures Based on Images Data and Deep Learning Model **8**(2493) 1-16 (2018)
22. Black, S.,Stylianou, A.: Visualizing Paired Image Similarity in Transformer Networks.Computer Vision Foundation.3164-3176

23. Caperton, R., Holdener, C.: Introducing Windows 10 for IT Professionals. Microsoft Press 1-5 (2016)
24. TensorFlow, https://www.tensorflow.org/resources/learn. Último acceso 01 de abril del 2022
25. Abadi,M., Barham, P., Chen,J.:TensorFlow: A system for large-scale machine learning. (2016)
26. Collobert,R.,Bengio,S., Mariéthoz, J.,: Torch: a modular machine learning software library. Idiap-RR-46 (2002)
27. ImageFinder, https://imagefinder.co/. Último acceso 15 Febrero 2021
28. Pixabay, https://pixabay.com. Último acceso 15 Febrero 2021
29. Let's Enhance, https://letsenhance.io/. Último acceso 19 Febrero 2021
30. Universidad de Granada, Escalas y métodos de estudio. https://https://www.ugr.es/. Último acceso 01 Marzo 2022
31. Genc,B.,Tunc,H.,: Optimal trining and test sets design for machine learning. Journal of Electrical Engineering and Computer Sciences 1-14 (2019)
32. Guy,T.,Zanhavy,T.,: Train on Validation: Squeezing the data lemon 1-17 (2018)
33. Tammina, s.,: Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images . International Journal of Scientific and Research Publications 149-150 Vol 9 (2019)
34. Simonyan,K.,Zisserman,A.,: Very deep convolutional networks for large-scale image recognition. p, Department of Engineering Science, University of Oxford (2015)