



ReCIBE. Revista electrónica de Computación, Informática,
Biomédica y Electrónica

ISSN: 2007-5448

recibe@cucei.udg.mx

Universidad de Guadalajara
México

Gutiérrez Velázquez, Miguel Ángel; Chacon-Murguía, Mario Ignacio; Ramírez-
Quintana, Juan Alberto; Arzate Quintana, Carlos; Corral-Saenz, Alma Delia
Clasificación de parásitos en imágenes de copro con transferencia de aprendizaje y aumento de datos

ReCIBE. Revista electrónica de Computación, Informática,
Biomédica y Electrónica, vol. 10, núm. 2, 2021, Noviembre-, pp. 1-15
Universidad de Guadalajara
Guadalajara, México

Disponible en: <https://www.redalyc.org/articulo.oa?id=512269058005>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica Redalyc

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso
abierto

Recibido 07 jul. 2021
Aceptado 16 mzo.2022

ReCIBE, Año 10 No. 2, Marzo 2022

Clasificación de parásitos en imágenes de copro con transferencia de aprendizaje y aumento de datos

Parasite classification in copro images with transfer learning and data augmentation

Miguel Ángel Gutiérrez Velázquez¹
m19061419@chihuahua.tecnm.mx

Mario Ignacio Chacon-Murguía¹
mario.cm@chihuahua.tecnm.mx

Juan Alberto Ramirez-Quintana¹
juan.rq@chihuahua.tecnm.mx

Carlos Arzate Quintana²
carzate@uach.mx

Alma Delia Corral-Saenz¹
alma.cs@chihuahua.tecnm.mx

¹ Tecnológico Nacional de México / I.T. Chihuahua, Chihuahua, Chihuahua, México.

² Universidad Autónoma de Chihuahua, Chihuahua, Chihuahua, México.

Resumen

Los seres humanos pueden albergar parásitos, por lo tanto, es fundamental una detección temprana para prevenir enfermedades. Los parásitos pueden observarse a través de imágenes microscópicas, con lo cual, la visión por computadora se muestra como un enfoque que pueda ayudar a la detección y clasificación de parásitos en imágenes digitales. Los modelos de aprendizaje profundo han mostrado un desempeño formidable en la clasificación de imágenes, debido a esto, en este artículo se presentan varios clasificadores profundos multiclase para reconocer 8 clases: 7 tipos de parásitos y la clase no parásitos. Los clasificadores diseñados utilizan transferencia de aprendizaje basada en la arquitectura AlexNet modificada. Al contar con una cantidad de muestras escasa de imágenes de parásitos, se realizó también un aumento de datos, tanto con el método tradicional como mediante la generación de imágenes con una red generativa adversaria (GAN) diseñada para ese propósito. El clasificador con mejor desempeño presentó una exactitud del 99.94%, 98.97% y 98.18% en el conjunto de entrenamiento, validación y de prueba, respectivamente.

Palabras clave: Clasificación de parásitos, aumento de datos, transferencia de aprendizaje, GAN, AlexNet.

Resume

Humans can harbor parasites; hence, it is fundamental an early detection to prevent diseases. Parasites can be observed in microscopic images, and computer vision may be a helpful approach to detect and classify those parasites in digital images. Deep learning models have shown to have a high performance in image classification. Therefore, this article presents various multi-class deep learning classifiers to recognize 8 classes: 7 types of parasites and non-parasite class. The designed classifiers are based on transfer learning from an AlexNet modified architecture. By having a reduce amount of parasite images samples, a data augmentation was done, employing traditional methods and images generation with an adversarial neural network (GAN) designed for this purpose. The classifier with best performance presented a 99.94%, 98.97% and 98.18% accuracy in the training, validation, and testing sets, respectively.

Keywords: Parasite classification, data augmentation, transfer learning, GAN, AlexNet.

1. Introducción

La visión por computadora es un campo científico cuyo objetivo es extraer información de imágenes digitales para su posterior análisis. Para efectuar tal extracción, se emplean algoritmos que realizan operaciones en las imágenes con el fin de mejorar la información pictórica para la interpretación humana, o para tareas como almacenamiento, transmisión y extracción de información en las imágenes. La visión por computadora aparece en innumerables campos, como: agricultura (Song C. *et al*, 2020), automatización (Chain E. *et al*, 2021), seguridad (Prakash R. *et al*, 2020), medicina (Meda K. *et al*, 2021), entre otros. La integración de la visión por computadora en el área de la medicina ayuda a los expertos a realizar los diagnósticos y tratamientos médicos de una manera más eficiente mediante el análisis de imágenes médicas. Las imágenes médicas predominantes son las obtenidas mediante rayos X, tomografía computarizada (CT), imagen de resonancia magnética (MRI), tomografía de emisión de positrones (PET) y ultrasonido. Además de los métodos anteriores, existen imágenes obtenidas a partir de un microscopio. Las imágenes conseguidas con un microscopio amplifican objetos diminutos, permitiendo observar lo que no es posible de otra manera. Gracias a esto, se puede analizar y examinar la composición, la estructura y las características de células y tejidos de seres vivos, así como la detección y clasificación de parásitos que habitan en el interior de humanos y animales.

El 2 de marzo del 2020, la Organización Mundial de la Salud (OMS) publicó una noticia acerca de la helmintiasis, enfermedad de parásitos (OMS, 2020). Se menciona la cifra escandalosa de 1500 millones de personas infectadas en el mundo. Esto aun cuando la OMS, en el 2001, solicitó a los países afrontar esta enfermedad. Dado que las cifras proporcionadas por la OMS son alarmantes y exponen un gran riesgo e impacto en la salud (para una sola enfermedad intestinal), por lo que es menester el desarrollo de tecnologías que detecten eficientemente parásitos intestinales. Una forma es mediante un sistema de análisis de imágenes digitales de muestras de copro. En esta investigación, se utilizan imágenes de parásitos y de no parásitos para clasificar el tipo de parásito presente en dichas imágenes o, por el contrario, clasificar el objeto que aparece en la imagen como no parásito. La clasificación de parásitos en imágenes de muestras de copro se realiza con una red neuronal profunda modificada de la red Alexnet, haciendo una transferencia de aprendizaje. Además, debido a que el número de muestras de imágenes de copro era reducido, se realizó un aumento de datos en el conjunto de entrenamiento mediante técnicas clásicas y con una red generativa adversaria, GAN, diseñada para ese propósito. Las contribuciones de este trabajo son: diseño de clasificadores mediante técnica de transferencia de aprendizaje, donde el clasificador con mejor desempeño presentó una exactitud del 98.18% en el conjunto de prueba, y el aumento de datos: una demostración de que las imágenes generadas con la GAN incrementan el desempeño del clasificar profundo.

El artículo está organizado como sigue: en la sección 2 se muestran trabajos relacionados, en la sección 3 se presenta el modelo de GAN diseñado para aumento de datos. la técnica de transferencia de aprendizaje utilizada y el clasificador, en la sección **Error! Reference source not found.** se describen los experimentos realizados y se analizan los resultados obtenidos. Finalmente, en la sección 1 se comentan las conclusiones del trabajo.

2. Trabajos relacionados

Existe en la literatura algunos trabajos relacionados al tema de análisis de imágenes médicas. En esta sección se presentan trabajos donde se realiza una clasificación de parásitos mediante clasificadores profundos. Por ejemplo, en (Shorten C. *et al*, 2021; Suriva *et al*, 2019; Shah *et al*, 2020) se reportan trabajos donde se realizan clasificadores binarios utilizando redes neuronales convolucionales (CNN). En (Shorten C. *et al*, 2021) se clasifica malaria en imágenes de células rojas. Las imágenes de entrada a la CNN son de 125x125x3, la CNN tiene 5 capas de convolución (con 32, 64, 128, 256 y 300 filtros respectivamente), 5 capas de max-pooling y 2 capas totalmente conectadas. El conjunto de imágenes utilizado es del conjunto de datos NIH Malaria, donde se tienen 13779 imágenes con parásitos y 13779 sin parásitos. Este conjunto de datos fue aumentado mediante técnicas clásicas hasta obtener 173700 imágenes. El clasificador tuvo una precisión de 96.33%. En (Suriva . *et al*, 2019) se realiza una clasificación de imágenes de células rojas infectadas debido a presencia de parásitos. Primeramente, la imagen es procesada para tener un tamaño de 128x128. Luego se entrena la red diseñada. La cual tiene 19 capas de convolución para la extracción de características, 6 capas de max

pooling para reducir el número de cálculos computacionales, 3 capas de normalización por lotes para normalizar las características, 4 capas totalmente conectadas y una capa plana. El conjunto de datos consiste en 26188 imágenes, 13105 pertenecen a la clase de imágenes parasitadas y 13083 a imágenes no parasitadas. La red tuvo un 98.9% de precisión. En (Shah . *et. al*, 2020) se identifica el parásito *Plasmodium* en imágenes de sangre. La CNN propuesta consta de 3 capas de convolución con 32, 64 y 128 filtros, respectivamente. Cada filtro tiene un tamaño de 3x3 y la función de activación es la ReLU. El algoritmo de aprendizaje es el optimizador Adam. Además, el modelo propuesto tiene 3 capas de pooling, 3 capas totalmente conectadas y la salida de la red está dada por la función sigmoidal. El conjunto de entrenamiento consiste en 8000 imágenes, mientras que el conjunto de prueba contiene 2000 imágenes. Para cuantificar la funcionalidad del modelo, se tienen 17460 imágenes de prueba, con 8760 imágenes con la presencia de parásito y 8700 sin presencia de parásito. La precisión del modelo en el conjunto de prueba mencionado fue de 95%.

Por otro lado, los enfoques basados en la utilización de redes profundas pre entrenadas se presentan en (Huq y Pervin, 2020; Nakasi *et. al*, 2021; Var y Boray, 2018; Bharadwaj y Sujitha, 2019; Chakradeo *et. al*, 2021). En (Huq y Pervin, 2020) se implementó la red VGG16 para identificar imágenes de células con parásitos y sin parásitos. Se utilizó un conjunto de datos disponible en National Institute of Health, que tiene 27558 imágenes disponibles. De esas imágenes, 13779 son imágenes sin parásitos y 13779 son imágenes infectadas. El 60% de las imágenes fueron utilizadas para el entrenamiento, el 10% para la validación y el 30% para la prueba. El desempeño fue de 95.96%. En (Nakasi *et. al*, 2021) se clasifican parásitos intestinales en muestras de heces. El conjunto de datos consta de 3327 imágenes con 11141 parásitos, este conjunto fue aumentado mediante técnicas clásicas para obtener 7 imágenes a partir de una sola. Se utilizaron las redes AlexNet y GoogleNet para la clasificación. Ambas redes se entrenaron durante 10 épocas con el optimizador Adam, con un factor de aprendizaje de 0.001 y un momentum de 0.9. La métrica de evaluación fue el área bajo la curva (AUC) de la región de convergencia (ROC). Utilizando la red GoogleNet se obtuvo un AUC de 0.99, mientras que con la AlexNet el AUC fue 1.00. En (Var y Boray, 2018) se clasifican parásitos en imágenes de muestras de sangre. Se aplicó transferencia de aprendizaje utilizando la red CNN pre entrenada VGG19. La red se entrenó durante 20 épocas para 5 clases: 1428 imágenes para la clase *Plasmodium Vivax*, 1425 para la clase *Plasmodium Ovale*, 1446 para la clase *Plasmodium Falciparum*, 1450 para la clase *Plasmodium Malariae* y 1440 para la clase de no parásitos. El modelo alcanza una precisión de 80%, 83%, 86% y 75% de precisión en 19 imágenes de prueba. En (Bharadwaj y Sujitha, 2019) se clasifican imágenes de células infectadas debido a presencia de parásitos implementando transferencia de aprendizaje. El conjunto de datos contiene 27558 imágenes de células infectadas y no infectadas, dicho conjunto de datos está disponible en el sitio web de National Library of Medicine. La red pre entrenada utilizada es la ResNet50. Se utilizó el optimizador gradiente descendiente estocástico (SGD) y la función de pérdida es la entropía cruzada. La red se entrenó durante 9 épocas. La precisión del entrenamiento fue de 95.91% y de la validación 95.4%. En (Chakradeo *et. al*, 2021) realiza una clasificación binaria de imágenes de células infectadas y no infectadas. Se tiene un conjunto de datos con 12058 imágenes infectadas y 14142 imágenes no infectadas. Luego, mediante aumento de datos se alcanzan las 22046 imágenes para cada clase para el entrenamiento. La red reportada en (Chakradeo *et. al*, 2021) utilizada está basada en la red VGG; la arquitectura de la red implementada en (Chakradeo *et. al*, 2021) tiene 6 capas de convolución, las capas 1 y 2 tienen 64 filtros, la 3 y 4 tienen 128 filtros y la 5 y 6 tienen 256 filtros. La función de activación es la Leaky ReLU, hay tres capas de max pooling con dropout y dos capas totalmente conectadas. La exactitud fue de 98.2%.

Al analizar los trabajos anteriores se observa que en la mayoría de los casos fue necesario contar con un gran número de muestras iniciales, con lo cual surge la necesidad de investigar si la clasificación de parásitos es posible cuando la cantidad de muestras disponibles es reducida. Por lo tanto, esta investigación tiene como fin diseñar un clasificador de parásitos intestinales con alto desempeño utilizando una red profunda partiendo de una muestra reducida de muestras, 358 muestras de imágenes de parásitos (estas 358 muestras se definirán como imágenes de parásitos originales, *IPO*) y 845 muestras negativas (no parásitos), este conjunto de muestras negativos se define como *NP*.

3. Aumento de datos, transferencia de aprendizaje y clasificadores

Esta sección presenta el aumento de datos mediante técnicas clásicas y mediante la generación de imágenes con una GAN. Posteriormente, se describe el concepto de transferencia de aprendizaje y el diseño de la red pre entrenada que se utiliza en esta investigación.

3.1 Aumento de datos

El aumento de datos tradicional está basado en transformar una imagen para obtener una nueva imagen que aporte en las muestras de entrenamiento y prueba para modelos de aprendizaje de máquina o profundo. Las transformaciones pueden ser: rotación, reflejo, adición de ruido, modificación del brillo, corrección gamma, borrado aleatorio, cambio en el espacio de color, etcétera (Shorten y Khoshgoftaar, 2019). En la Figura 1 se muestran algunas de estas transformaciones.

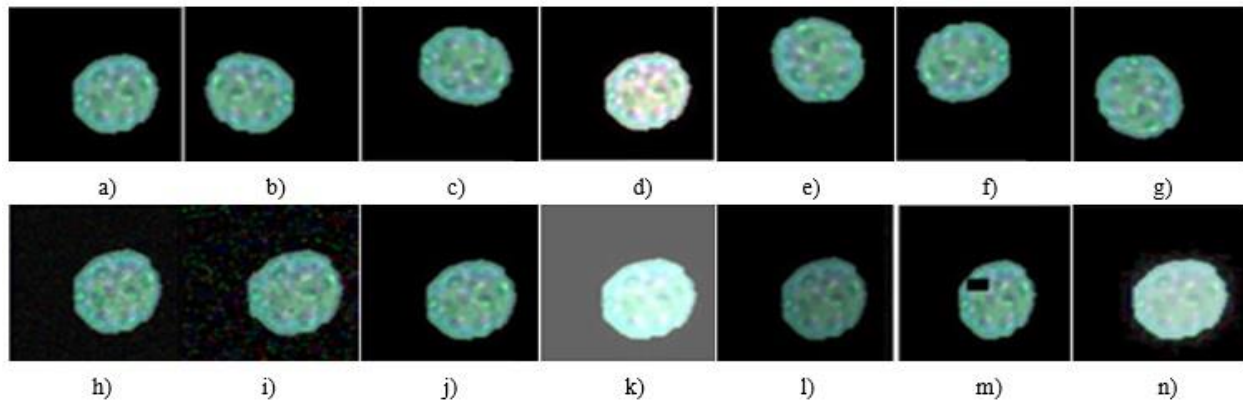
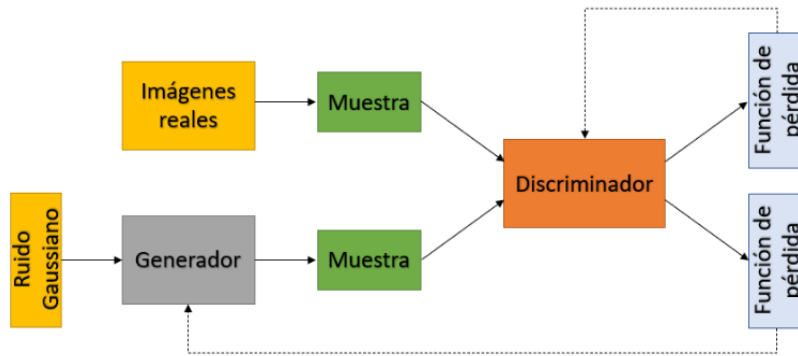


Figura 1. Aumento de datos mediante transformaciones, a) Imagen original, b) Imagen reflejada en el eje y, c) Imagen reflejada en el eje x, d) Aumento de contraste, e) Imagen rotada 90°, f) Imagen rotada 180°, g) Imagen rotada 270°, h) Imagen con ruido gaussiano, i) Imagen con ruido sal y pimienta, j) Imagen con ruido Poisson, k) Aumento de brillo, l) Disminución de brillo, m) Borrado aleatorio, n) Corrección gamma.

Adicional a estas transformaciones, existen paradigmas de aumento de datos que generan nuevas imágenes, por ejemplo, la GAN. La GAN es un paradigma basado en los métodos generativos propuesto por Goodfellow *et. al* (Goodfellow *et. al*, 2014). Mientras que los modelos discriminativos asocian una entrada a unas coordenadas en un hiperplano y de esa manera le asignan la etiqueta correspondiente, los modelos generativos intentan producir una salida que se parezca a la entrada.

En la Figura 2 se observa la arquitectura general de una GAN, donde existen dos clasificadores: el generador y el discriminador. El discriminador tiene dos entradas: las imágenes reales y las imágenes generadas por el generador. En cambio, el generador tiene como entrada ruido gaussiano. El discriminador aprende de las imágenes reales y de las imágenes creadas por el generador, de tal modo que clasifica una imagen como real o artificial. Por otra parte, el generador intenta engañar al discriminador creando imágenes parecidas al conjunto de muestras reales. Basado en este modelo de GAN, se presenta en este trabajo la red GAN profunda (DCGAN) (Radford y Metz, 2015), cuya arquitectura se muestra en la Tabla 1. La DCGAN se entrenó 7 veces para generar artificialmente imágenes de los parásitos estudiados en este trabajo. Estos 7 entrenamientos se deben a que se tienen 7 clases de parásitos, y cada entrenamiento corresponde a la generación de imágenes de un tipo específico de parásito. Para cada entrenamiento, las imágenes de cada parásito fueron aumentadas con métodos tradicionales, ya que esto mejora el resultado de las imágenes generadas, visualmente hablando.



Generador			Discriminador		
Capa	Tipo de capa	Dimensión de la activación	Capa	Tipo de capa	Dimensión de la activación
1	Ruido gaussiano de entrada	1x1x100	1	Conjunto de entrenamiento	64x64x3
2	Proyección y cambio de tamaño	4x4x512	2	Dropout	64x64x3
3	Convolución traspuesta	8x8x384	3	Convolución	32x32x64
4	Normalización por lotes	8x8x384	4	Leaky ReLU	32x32x64
5	ReLU	8x8x384	5	Convolución	16x16x128
6	Convolución traspuesta	16x16x192	6	Normalización por lotes	16x16x128
7	Normalización por lotes	16x16x192	7	Leaky ReLU	16x16x128
8	ReLU	16x16x192	8	Convolución	8x8x256
9	Convolución traspuesta	32x32x96	9	Normalización por lotes	8x8x256
10	Normalización por lotes	32x32x96	10	Leaky ReLU	8x8x256
11	ReLU	32x32x96	11	Convolución	4x4x512
12	Convolución traspuesta	64x64x3	12	Normalización por lotes	4x4x512

Figura 2. Modelo general de una red GAN.

13	Tanh	64x64x3	13	Leaky ReLU	4x4x512
			14	Convolución	1x1x1

Tabla 1. Arquitectura del generador y discriminador de la DCGAN.

A continuación, se presentan la cantidad de muestras de imágenes de parásitos de entrenamiento usadas para diseñar la DCGAN para cada tipo de parásito: 1548 *Blastocystis hominis* (BH), 984 *Entamoeba coli* (EC), 672 *Chilomastix Mesnili* (CM), 1440 *Entamoeba histolytica* (EHY), 384 *Giardia Lamblia* (GL), 251 *Endolimax nana* (EN) y 2881 de *Enteromonas hominis* (EHO). Algunas imágenes generadas con la DCGAN diseñada para cada tipo de parásito se muestran en la Figura 3. La DCGAN fue entrenada durante 1000 épocas con un factor de aprendizaje de 0.0002 y un valor de decaimiento de 0.7 (valores típicos), el tamaño del mini-batch es de 64. La cantidad de imágenes generadas con la DCGAN es: 218 BH, 475 EC, 157 CM, 204 EHY, 158 GL, 466 EN y 1014 EHO, para un total de 2692 imágenes generadas con la DCGAN.

3.2 Transferencia de aprendizaje

El escenario ideal para que una red profunda aprenda es aquel donde hay decenas de miles imágenes de entrenamiento etiquetadas, que tienen la misma distribución que las imágenes de prueba. Sin embargo, generar esta cantidad de imágenes en ocasiones no es viable por el costo, trabajo y tiempo necesarios para obtener este tipo de información y realizar el entrenamiento correspondiente.

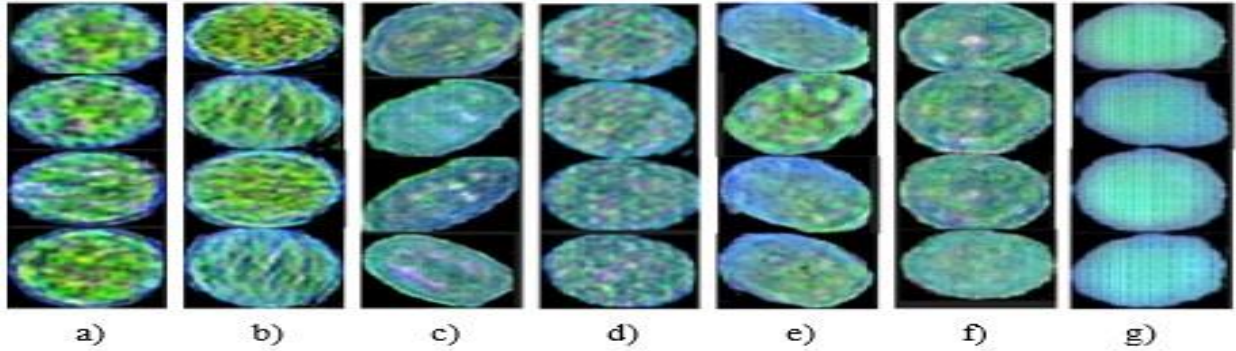


Figura 3. Algunas imágenes generadas con la GAN, a) *Blastocystis hominis*, b) *Entamoeba coli*, c) *Chilomastix mesnili*, d) *Entamoeba histolytica*, e) *Giardia lamblia*, f) *Endolimax nana*, g) *Enteromonas hominis*.

Para dar solución a este problema, existe la transferencia de aprendizaje. La transferencia de aprendizaje es utilizada para mejorar el aprendizaje de un dominio al transferir información de otro dominio relacionado (Weiss *et. al*, 2016). El aprender a transferir es el resultado de la generalización de la experiencia. Es decir, es posible realizar una transferencia de conocimiento de una situación a otra, siempre y cuando se generalice la experiencia.

El requisito de la transferencia es que exista una conexión entre las dos actividades de aprendizaje. La transferencia de aprendizaje se inspira en la habilidad de los seres humanos de transferir conocimientos a través de dominios, de un dominio fuente a un dominio objetivo. Donde el dominio fuente es el dominio del que ya se tiene conocimiento, y el dominio objetivo es el que se quiere aprender a partir de lo que se sabe sobre el dominio fuente.

De esta manera, esta forma de aprendizaje tiene como objetivo aprovechar el conocimiento de un dominio relacionado para mejorar el aprendizaje o minimizar el número de ejemplos etiquetados requeridos en otro dominio de prueba (Zhuang *et. al*, 2020) En este trabajo, el dominio fuente es lo que aprendió la red Alexnet durante el entrenamiento con millones de imágenes correspondientes a diversas clases, y el dominio objetivo son las imágenes de parásitos y no parásitos Para entender la transferencia de aprendizaje se requieren varias definiciones (Pan y Yang, 2009).

- Dominio: Un dominio D está formado de dos partes: el espacio de características χ y su distribución marginal $P(X)$.

$$D = \{\chi, P(X)\} \quad (1)$$

donde X representa el conjunto de instancias

$$X = \{x | x_i \in \chi, i = 1, \dots, n\} \quad (2)$$

- Tarea: Una tarea τ consiste en un espacio de etiqueta Y y una función de decisión f que se espera sea aprendida desde los datos de muestra

$$\tau = \{Y | f\} \quad (3)$$

Algunos modelos de aprendizaje dan como salida la distribución condicional predicha de los ejemplos

$$f(x_j) = \{P(y_k | x_j) | y_k \in Y, k = 1, \dots, |Y|\} \quad (4)$$

En la práctica, el dominio se observa por un número de instancias x con o sin la información etiquetada (en este trabajo, al ser aprendizaje supervisado, se tiene la información etiquetada). El dominio fuente D_S correspondiente a la tarea fuente τ_S se expresa como la tupla ejemplo-etiqueta

$$D_S = \{(x, y) | x_i \in \mathcal{X}^S, y_i \in Y^S, i = 1, \dots, n^S\} \quad (5)$$

- Transferencia de aprendizaje: Dado una o más observaciones correspondientes a $m^S \in \mathbb{N}^+$ dominios y tareas $\{(D_{Si} | \tau_{Si}) | i = 1, \dots, m^S\}$ y alguna o algunas observaciones $m^T \in \mathbb{N}^+$ sobre el dominio y tarea de prueba $\{(D_j | \tau_j) | j = 1, \dots, m^T\}$, donde m^S es la cantidad de dominios fuente y m^T la cantidad de tareas, la transferencia aprendizaje utiliza el conocimiento obtenido del dominio fuente al mejorar el desempeño de las funciones de decisión aprendidas f^T_j ($j = 1, \dots, m^T$) en el dominio de prueba. En este trabajo la cantidad de tareas son 8, ya que se clasificarán 7 clases de parásitos y la clase de no parásitos.

La transferencia de aprendizaje que se implementó fue con la red Alexnet, la cual es una CNN profunda utilizada en el concurso de ImageNet LSVRC en 2012 para clasificar imágenes entre 1000 clases. La red tiene 60 millones de parámetros y 650000 neuronas, 5 capas convolucionales, max pooling y tres capas totalmente conectadas (Krizhevsky *et. al*, 2012). AlexNet se entrenó con 1.2 millones de imágenes, con 50,000 imágenes de validación y 150,000 imágenes de prueba.

El conocimiento de la red son los valores de los filtros. Esto es, cada filtro activará ciertas regiones de la imagen de entrada, generando las características que analizará la red. Por ejemplo, la Figura 4a muestra una imagen de un parásito y en la Figura 4b se observan algunas activaciones en la primera capa de convolución de la red AlexNet. Estas activaciones en las capas de convolución son conocimiento sobre ciertas características que aprendió la red durante el entrenamiento, y ese conocimiento se utiliza para reconocer ciertas características de las imágenes de parásitos de entrada que no estaban en el conjunto de entrenamiento con el cual la red fue entrenada.

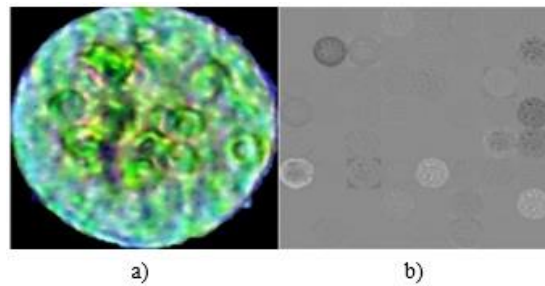


Figura 4. a) Imagen de entrada a la red AlexNet, b) Algunas características de la primera capa de convolución de la red AlexNet.

3.3 Clasificadores

Para el diseño de los clasificadores multiclase se hace una transferencia de aprendizaje de una red profunda ya entrenada (AlexNet). Para ello, se modificó la última capa de la red AlexNet, correspondiente a una capa totalmente conectada, para que sea de tamaño 8, pues ahora solo clasificarán 8 clases y no mil, para la cual fue diseñada originalmente. La Figura 5 muestra la arquitectura de la red que se entrenó con el gradiente estocástico descendente con momentum, con una tasa de aprendizaje inicial de 0.001. El total de épocas es de 20 con un mini lote de 128. Lo único que cambia en cada clasificador (antes de ser entrenados) es el conjunto de entrenamiento.

4. Aumento de datos, transferencia de aprendizaje y clasificadores

En esta sección se explican los experimentos realizados y sus resultados. Los diversos experimentos tienen como fin el ver cómo cambia el desempeño del clasificador profundo al utilizar distintas muestras de entrenamiento. Todos los clasificadores utilizaron la arquitectura mostrada en la Figura 5.

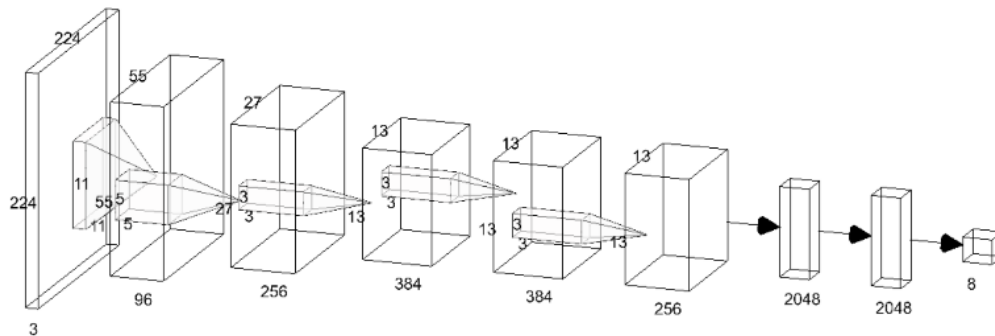


Figura 5. Arquitectura de la red para los clasificadores diseñados.

Las 358 muestras de parásitos reales están divididas de la siguiente manera: 72 *BH*, 41 *EC*, 28 *CM*, 60 *EHY*, 16 *GL*, 21 *EN* y 120 *EHO*.

4.1 Clasificadores multiclase

Clasificador 1, C1, se entrenó la red con las 358 muestras de parásitos reales y las 845 imágenes de *NP*. Las muestras fueron divididas en 70% para el entrenamiento, 15% para la validación y 15% para la prueba para los 4 clasificadores diseñados. El desempeño de C1 se muestra en la Tabla 2. Clasificador 2, C2, se entrenó la red con las 358 imágenes de parásitos reales más las 2692 imágenes generadas con la DCGAN y las 845 imágenes de *NP*. Los resultados de C2 se muestran en la Tabla 3. Clasificador 3, C3, se hizo un aumento de datos con técnicas clásicas a las 358 muestras de parásitos. El aumento de datos consistió en reflejar la imagen en el eje x y en el eje y, adición de ruido gaussiano, ruido sal y pimienta y ruido de poisson, y rotación de 90°, 180° y 270°. El conjunto de entrenamiento de C3 consiste en las muestras: 1548 *BH*, 984 *EC*, 672 *CM*, 1440 *EHY*, 384 *GL*, 251 *EN*, 2881 *EHO* y las 845 imágenes de *NP*. El desempeño de C3 se presenta en la Tabla 4. Clasificador 4, C4, este clasificador tiene como conjunto de entrenamiento el mismo que C3, más las 2692 imágenes generadas con la DCGAN. En la Tabla 5 se muestra el desempeño de C4

Conjunto	Exactitud
Entrenamiento	100%
Validación	92.22%
Prueba	91.67%

Tabla 2. Resultados del clasificador C1

Conjunto	Exactitud
Entrenamiento	99.89%
Validación	96.06%
Prueba	95.88%

Tabla 3. Resultados del clasificador C2.

Conjunto	Exactitud
Entrenamiento	100%
Validación	98.07%
Prueba	98.08%

Tabla 4. Resultados del clasificador C3.

Conjunto	Exactitud
Entrenamiento	99.94%
Validación	98.97%
Prueba	98.18%

Tabla 5. Resultados del clasificador C4.

El primer clasificador, C1, obtuvo un desempeño del 91.67% en el conjunto de prueba. El clasificador, C2, tuvo un desempeño superior a C1, del 91.67% al 95.88%. El clasificador C3 tuvo como conjunto de entrenamiento a las imágenes originales con aumento de datos tradicional. En la Tabla 4 se observa que su desempeño fue del 98.08% en el conjunto de prueba. Tal conjunto contiene imágenes reales, por lo que podría compararse con C1 directamente. Este experimento muestra que el desempeño con aumento de datos tradicional es mucho mejor que el obtenido utilizando solamente las muestras originales. Por último, se entrenó el clasificador C4.

El conjunto de entrenamiento de C4 fue una combinación de las muestras originales, muestras generadas con la DCGAN y muestras obtenidas con aumento de datos tradicional. En la Tabla 5 se observa un ligero incremento en el desempeño en relación con C3: de 98.08% a 98.18%. Con estos resultados se comprueba que el aumento de datos mejora el desempeño del clasificador.

En la Figura 6, Figura 7, Figura 8 y Figura 9 se muestra las matrices de confusión para C1, C2, C3 y C4 para el conjunto de prueba que le pertenece a cada clasificador. Las filas corresponden a las etiquetadas reales y las columnas a las salidas del clasificador. La penúltima fila muestra los porcentajes de clasificaciones correctas e incorrectas para cada clase real. La columna de la derecha muestra los porcentajes de las clasificaciones correctas e incorrectas para cada clase predicha por el clasificador.

En la diagonal de la matriz de confusión (las casillas verdes) se muestran las clasificaciones correctas para cada clase, y los elementos fuera de la diagonal en la matriz de confusión (casillas rojas) están las clasificaciones incorrectas.

BH	8	1	0	1	1	0	0	0	72.7%
CM	1	2	0	0	1	0	0	0	50%
EN	0	1	2	0	0	0	0	0	66.7%
EC	0	0	0	6	0	0	0	0	100%
EHY	2	0	0	0	7	0	0	0	77.8%
EHO	0	0	0	0	0	15	0	3	83.3%
GL	0	0	0	0	0	0	1	1	50%
NP	1	1	0	0	0	1	0	124	97.6%
	66.7%	40%	100%	85.7%	77.8%	93.8%	100%	96.9%	91.7%
BH	CM	EN	EC	EHY	EHO	GL	NP		

Figura 6. Matriz de confusión de C1 en el conjunto de prueba de C1.

BH	39	1	0	0	1	0	0	2	90.7%
CM	3	22	1	1	1	0	0	0	78.6%
EN	0	1	72	0	0	0	0	0	98.6%
EC	2	0	0	75	0	0	0	0	97.4%
EHY	0	0	0	1	38	0	1	0	95%
EHO	0	0	0	0	0	170	0	0	100%
GL	0	0	0	1	1	1	22	1	84.6%
NP	0	1	0	0	0	2	1	123	96.9%
	88.6	88%	98.6%	96.2%	92.7%	98.3%	91.7%	97.6%	99.2%
BH	CM	EN	EC	EHY	EHO	GL	NP		

Figura 7. Matriz de confusión de C2 en el conjunto de prueba de C2.

BH	225	0	0	0	6	0	0	1	97%
CM	0	97	0	0	4	0	0	0	96%
EN	0	1	36	0	1	0	0	0	94.7%
EC	0	0	0	148	0	0	0	0	100%
EHY	7	0	0	3	206	0	0	0	95.4%
EHO	0	0	0	0	0	431	0	1	99.8%
GL	0	1	0	0	0	0	57	0	98.3%
NP	0	0	0	0	0	0	1	126	99.2%
	97%	98%	100%	98%	94.9%	100%	98.3%	98.4%	99.2%
	BH	CM	EN	EC	EHY	EHO	GL	NP	

Figura 8. Matriz de confusión de C3 en el conjunto de prueba de C3.

BH	264	0	0	1	0	0	0	0	99.6%
CM	3	108	10	0	1	2	0	0	87.1%
EN	0	0	103	0	4	0	1	0	95.4%
EC	0	0	0	219	0	0	0	0	100%
EHY	2	0	0	2	243	0	0	0	98.4%
EHO	0	0	0	0	0	584	0	0	100%
GL	0	0	0	1	0	0	79	1	97.5%
NP	0	0	0	0	0	4	0	123	96.9%
	98.1%	100%	91.2%	98.2%	98%	99%	98.8%	99.2%	99.2%
	BH	CM	EN	EC	EHY	EHO	GL	NP	

Figura 9. Matriz de confusión de C4 en el conjunto de prueba de C4.

Finalmente, se realiza un experimento, E1, que tiene como objetivo evaluar de mejor manera el desempeño de C2. Pues C2 fue entrenado con un alto porcentaje de imágenes generadas con la DCGAN, y en el conjunto de prueba también hay un alto porcentaje de imágenes generadas con la DCGAN en relación con las imágenes reales. E1 consiste, entonces, en clasificar el conjunto de imágenes utilizadas en C3 (puras imágenes reales con aumento de datos) sin las IPO (imágenes de parásitos originales). Es decir, clasificar solamente las imágenes obtenidas con el aumento de datos tradicional. El resultado de C2 en la clasificación del conjunto de prueba descrito anteriormente se muestra en la 0. En la 0 se observa que el desempeño de C2 al clasificar imágenes reales es muy bajo. Pero hay clases como BH, EHY y EHO que tuvieron buen desempeño (88.6%, 90.4% y 99.2% respectivamente). Con lo cual se puede concluir que las imágenes generadas de ciertos parásitos son mejores que para otro tipo de parásito. Como se describió en la sección 3.1, en el entrenamiento de la DCGAN se utilizaron más imágenes de BH, EHY y EHO que de los otros 4 tipos de parásitos. Es decir, quizá si se generan más imágenes con la DCGAN utilizando un conjunto pequeño de imágenes, el desempeño de un clasificador entrenado con muestras generadas por esa DCGAN resultaría mejor al clasificar imágenes reales.

BH	838	0	321	294	7	0	4	12	56.8%
CM	2	294	280	1	28	3	17	19	45.7%
EN	0	10	209	0	11	0	0	0	90.9%
EC	41	0	1	898	3	0	0	0	95.2%
EHY	65	21	442	214	64	0	11	13	44.5%
EHO	0	23	505	0	0	2055	4	173	74.5%
GL	0	50	122	11	15	9	101	60	27.4%
NP	0	1	0	0	1	4	1	836	99.2%
	88.6%	73.7%	11.1%	63.3%	90.4%	99.2%	73.2%	75.2%	67.6%
	BH	CM	EN	EC	EHY	EHO	GL	NP	

Figura 10. Matriz de confusión del experimento E1.

1. Conclusiones

En este trabajo se presentaron varios clasificadores profundos que fueron entrenados utilizando diversos conjuntos de entrenamiento. Con lo cual, no solamente se presenta un clasificador profundo con un alto desempeño, sino que diversos experimentos para saber cómo varía el desempeño si se modifica el conjunto de entrenamiento. Tales conjuntos de entrenamiento consisten en alguna combinación de *IPO* e *NP*, imágenes con aumento de datos tradicional e imágenes generadas con la DCGAN. El clasificador C1, cuyo conjunto de muestras únicamente presenta imágenes de parásitos reales y las imágenes de *NP* obtuvo un desempeño del 91.67% en el conjunto de prueba. El clasificador C3 tiene como muestras las mismas imágenes de parásitos reales que C1, pero fueron incrementadas mediante aumento de datos, y las imágenes de *NP*. El desempeño de C3 fue del 98.08% en el conjunto de prueba. Por último, el clasificador C4 contiene como muestras las mismas que C3, pero con la adición de las imágenes de parásitos generadas con la DCGAN. Este clasificador presentó una exactitud del 98.18% en el conjunto de prueba.

Estos resultados son importantes para demostrar que el aumento de datos mejora notablemente la clasificación multiclase de imágenes de parásitos, incluso sin tener miles de imágenes de muestra, como en otros trabajos. Además, se observó que un clasificador entrenado con imágenes generadas con la DCGAN, en combinación con imágenes reales, obtiene un mejor desempeño que sin ellas. En cambio, diversas investigaciones se ven necesarias: primeramente, entrenar un clasificador con un alto porcentaje de imágenes generadas con la DCGAN y que tenga un buen desempeño al clasificar imágenes reales, ya que en el experimento E1, se observa que C2 no tiene un buen desempeño en la clasificación de imágenes de parásitos reales para todos los tipos de parásitos; también se presenta la necesidad de investigar qué características aprende el modelo profundo a partir del aumento de datos tradicional y DCGAN. Como se comprobó en este trabajo, el aumento de datos tradicional y DCGAN incrementa el desempeño de un clasificador profundo. Se intuye, por lo tanto, que una red profunda obtiene más información y encuentra más características cuando se tiene aumento de datos, sería muy interesante y útil saber cuáles son, es decir: qué características se obtienen de las imágenes de parásitos con ciertas modificaciones, que no se tienen en las imágenes de parásitos originales.

Agradecimientos

Este proyecto se realizó bajo el apoyo del Tecnológico Nacional de México con número 10071.21-P.

REFERENCIAS

- Bharadwaj A. y Sujitha D. (2019). Transfer Learning with ResNet-50 for Malaria Cell-Image Classification. *2019 International Conference on Communication and Signal Processing (ICCSP)*, 946-949. DOI: 10.1109/ICCSP.2019.8697909
- Chain E., Fang W., Miang Y y Tian J., (2021). Computer vision approaches for detecting missing barricades. *Automation in Construction*, 131, 103852. <https://doi.org/10.1016/j.autcon.2021.103862>
- Chakradeo K., Delves M. y Titarenko S., (2019). Malaria Parasite Detection Using Deep Learning Methods. *2019 International Conference on Communication and Signal Processing (ICCSP)*, 15(2), 175-182. <https://doi.org/10.1016/j.jns.2017.08.946>
- Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S, Courville A. y Bengio Y., (2014). Generative adversarial networks. *Advances in neural information processing systems*, 27, 2672–2680.
- Huq A. y Pervin T., (2020). Robust Deep Neural Network Model for Identification of Malaria Parasites in Cell Images. *2020 IEEE Region 10 Symposium (TENSYP)*, 1456-1459, <https://doi.org/10.1109/TENSYP50017.2020.9230832>.
- Krizhevsky A. Sutskever I. y Hinton G., (2012). ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-1105.
- Meda K., Milla S. y Rostad B., (2021). Artificial intelligence research within reach: an object detection model to identify rickets on pediatric wrist radiographs, *Pediatric Radiology*, 51, 782-791. <https://doi.org/10.1007/s00247-020-04895-8>
- Nakasi R., Alija E. y Nakatumba J., (2021). A Poster on Intestinal Parasite Detection in Stool Sample Using AlexNet and GoogleNet Architectures, *ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS)*, 389-395. <https://doi.org/10.1145/3460112.3472309>
- OMS (2 de marzo 2020). Helminthiasis transmitidas por el suelo. <https://www.who.int/es/news-room/fact-sheets/detail/soil-transmitted-helminth-infections>
- Pan S. y Yang Q., (2009). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*. 22(10), 1345-1359. <https://doi.org/10.1109/TKDE.2009.191>
- Prakash R. y Chithaluru P., (2020). Active Security by Implementing Intrusion Detection and Facial Recognition. *Nanoelectronics, Circuits and Communication Systems*, 692, 1-7. https://doi.org/10.1007/978-981-15-7486-3_1
- Radford A. y Metz L., (2015). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arVix Preprint*.
- Shah D., Kawale K., Shah M., Randive S. y Mapari R., (2020). Malaria Parasite Detection Using Deep Learning. *2020 4th Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS)*, 984-988. <https://doi.org/10.1109/ICICCS48265.2020.9121073>
- Shorten C. y Khoshgoftaar T., (2019). A survey on Data Augmentation for Deep Learning. *Journal of Big Data*, 6(60), 1-48. <https://doi.org/10.1186/s40537-019-0197-0>
- Shorten C., Maqsood A., Shahid M., Hassan M. y Grzegorzec M., (2021). Deep Malaria Parasite Detection in Thin Blood Smear Microscopic Images. *Applied Sciences*, 11(5), 2284. <https://doi.org/10.3390/app11052284>
- Song C., Wang C. y Yang Y., (2020). Automatic Detection and Image Recognition of Precision Agriculture for Citrus Diseases. *2020 IEEE Eurasia Conference on OIT, Communication and Engineering*

(ECICE), 187-190. <https://doi.org/10.1109/ECICE50847.2020.9301932>

Suriva M., Chandran V. y Sumithra M., (2019). Enhanced deep convolutional neural network for malaria parasite classification. *International Journal of Computers and Applications*, 1-10.
<https://doi.org/10.1080/1206212X.2019.1672277>

Var E. y Boray F., (2018). Malaria Parasite Detection with Deep Transfer Learning. *2018 3rd International conference on computer science and engineering (UBMK)*, 298-302.
<https://doi.org/10.1109/UBMK.2018.8566549>

Weiss K., Khoshgoftaar T. y Wang D., (2016). A survey of transfer learning. *Journal of Big Data*. 3(9), 1-40. <https://doi.org/10.1186/s40537-016-0043-6>

Zhuang F., Qi Z., Duan K., Xi D., Zhu Y., Zhu H., Xiong H y He Q., (2020), A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, 109(1), 43-76.
<https://doi.org/10.1109/JPROC.2020.3004555>



Esta obra está bajo una licencia de Creative Commons
Reconocimiento-NoComercial-CompartirIgual 2.5 México.