

<https://doi.org/10.23913/ride.v11i22.856>

*Artículos científicos*

## **Clasificación automática de anastomosis mediante redes neuronales convolucionales en video fetoscópico**

*An Automatic Classification of Anastomosis by Convolutional Neural Networks in Fetoscopic Video*

*Classificação automática de anastomoses usando redes neurais convolucionais em vídeo fetoscópico*

**Ángel Mario Lerma Sánchez**

Universidad Da Vinci, México

[alerna14@udavinci.edu.mx](mailto:alerna14@udavinci.edu.mx)

<https://orcid.org/0000-0002-5759-1302>

**Adriana Mexicano Santoyo**

Tecnológico Nacional de México, Instituto Tecnológico de Cd. Victoria, México

[mexicanao@gmail.com](mailto:mexicanao@gmail.com)

<https://orcid.org/0000-0002-8045-7558>

**Fabiola Mirosłaba Villalobos Castaldi**

Universidad Da Vinci, México

[mwillaloboscald@gmail.com](mailto:mwillaloboscald@gmail.com)

<https://orcid.org/0000-0002-6967-5706>

**Pedro Damián Reyes**

Universidad de Colima, México

[damian@uclm.mx](mailto:damian@uclm.mx)

<https://orcid.org/0000-0003-1114-9612>

## Resumen

El síndrome de transfusión gemelo a gemelo (*the twin-twin transfusion syndrome* o *TTTS* por sus siglas en inglés) es el resultado de un flujo sanguíneo desigual a través de anastomosis (conexión de los vasos sanguíneos) vasculares placentarias que enlazan las dos circulaciones fetales. Las anastomosis vasculares en la placenta compartida están presentes en prácticamente todos los embarazos gemelares monocoriónicos (*monochorionic twins* o *MC* por sus siglas en inglés), pero solo en aproximadamente 10 % conducen al síndrome de transfusión gemelo a gemelo. Sin intervención, la condición es a menudo fatal para ambos gemelos. Una alternativa para el tratamiento del TTTS es el procedimiento de láser placentario, conocido como *cirugía fetal*, que consiste, de manera general, en dividir la placenta en dos mediante la cauterización láser de los vasos sanguíneos entre los fetos, equilibrando así los flujos sanguíneos.

En la actualidad la cirugía fetoscópica es un procedimiento que se realiza con frecuencia en el país, por lo que la clasificación de anastomosis es vital para esta cirugía, la cual constituye el tratamiento más recomendado. Por su grado de complejidad, esta intervención quirúrgica presenta múltiples dificultades: los fetos se mueven durante el procedimiento, la orientación del video empleado no es adecuada para un análisis más preciso y el campo de visión generado por el fetoscópio es muy pequeño; por lo tanto, es necesario contar con una herramienta que ayude al médico novel a poder diferenciar y clasificar las anastomosis de una manera más adecuada. El objetivo de este trabajo, por tanto, es presentar el desarrollo de una herramienta computacional que contribuya a la clasificación automática de anastomosis dentro de un video fetoscópico mediante redes neuronales convolucionales de manera que sirva de soporte para el médico novel en su etapa de entrenamiento.

Para ello, se construyó un DataSet (conjunto de imágenes) a partir de videos fetoscópicos sin clasificar que luego se catalogaron en tres categorías, seleccionadas conjuntamente con los expertos, mediante una herramienta computacional *ex profeso* para ello (VideoLabel). La técnica de aumento de datos sirvió para construir imágenes artificiales a partir de las reales ya clasificadas, dado que el número de imágenes etiquetadas no eran suficientes; en ese contexto, se seleccionó la arquitectura AlexNet para realizar una transferencia de aprendizaje y para entrenarse, lo que arrojó resultados con una efectividad superior a 90 %. Estos datos permiten concluir que de no contar con una herramienta que permitiera entrenar a los médicos noveles en la identificación de anastomosis, su capacitación

tomaría más tiempo en llevarse a cabo, ya que se realizaría de manera *in situ* en cada cirugía fetal. Sin embargo, en esta investigación se generó el diseño de un *software* con el que resulta factible la clasificación automática de anastomosis a partir de un video fetoscópico mediante una red neuronal convolucional, con resultados prometedores como herramienta de apoyo para los médicos en su etapa de entrenamiento.

**Palabras clave:** aprendizaje de máquina, aprendizaje supervisado, inteligencia artificial, síndrome de transfusión gemelo a gemelo.

### Abstract

Twin-twin transfusion syndrome (*TTTS*) is the result of uneven blood flow through placental vascular anastomosis (blood vessel connection) that link the two fetal circulations. Vascular anastomosis in the shared placenta are present in virtually all monochorionic twin pregnancies (MCs), but in only about 10% lead to twin-twin transfusion syndrome. Without intervention, the condition is often fatal for both twins. An alternative to *TTTS* treatment is the placental laser procedure known as fetal surgery, which consists, in a very general way, of splitting the placenta in two, by laser cauterization of blood vessels between fetuses, thus balancing blood flows.

Currently fetoscopic surgery is a procedure that is performed frequently in Mexico and the appropriate classification of anastomosis is vital for this surgery, since it represents the most recommended treatment. However, due to its degree of complexity, this surgical intervention presents multiple difficulties, such as fetuses moving during the procedure, the orientation of the video used is not suitable for a more accurate analysis, the field of view generated by the fetus is very small. Therefore, it is necessary to have a tool that helps the doctor to be able to differentiate and classify anastomosis in a more appropriate way. The objective of this work is to present the development of a computational tool that contributes to the automatic classification of anastomosis within a fetoscopic video through convolutional neural networks in a way that serves as a support for the unexperienced doctor in his training stage.

A DataSet (image set) was built from fetoscopic videos first unclassified, then cataloged into three categories, selected in conjunction with experts, using a computational tool specifically created for this purpose (VideoLabel). The data augmentation technique served to build artificial images from the actual ones already classified, since the number of

tagged images were not sufficient; in the same context, the AlexNet architecture was selected to perform a learning transfer and be trained to obtain results above 90% effectiveness in the classifications made by the computational tool created. This data allows us to conclude that if we did not have a tool that would allow first-time physicians to train in the identification of anastomosis as a practice prior to fetal surgery, their training will take longer since it is done in situ in each fetal surgery. As a result of this research, the design of a software was generated with which it is feasible to automatically classify anastomosis from a fetoscopic video through a convolutional neural network with promising results as a tool to support doctors in their training stage, allowing them to carry out their training in a more agile and less timely way.

**Keywords:** Machine Learning, Supervised Learning, Artificial Intelligence, Twin to Twin Transfusion Syndrome.

## Resumo

A síndrome da transfusão de gêmeos (TTTS) é o resultado do fluxo sanguíneo desigual através das anastomoses vasculares da placenta (conexão dos vasos sanguíneos) ligando as duas circulações fetais. Anastomoses vasculares na placenta compartilhada estão presentes em praticamente todos os gêmeos monócóricos (MCs), mas apenas em cerca de 10% levam à síndrome de transfusão de gêmeos. Sem intervenção, a condição costuma ser fatal para ambos os gêmeos. Uma alternativa para o tratamento do TTTS é o procedimento a laser placentário, conhecido como cirurgia fetal, que geralmente consiste em dividir a placenta em duas por meio da cauterização a laser dos vasos sanguíneos entre os fetos, equilibrando assim os fluxos sanguíneos.

Atualmente, a cirurgia fetoscópica é um procedimento muito realizado no país, portanto a classificação da anastomose é fundamental para esta cirurgia, que é o tratamento mais recomendado. Devido ao seu grau de complexidade, esta intervenção cirúrgica apresenta múltiplas dificuldades: os fetos movimentam-se durante o procedimento, a orientação do vídeo utilizado não é adequada para uma análise mais precisa e o campo de visão gerado pelo fetoscópio é muito pequeno; portanto, é necessário um instrumento que ajude o médico novato a diferenciar e classificar as anastomoses de forma mais adequada. O objetivo deste trabalho, portanto, é apresentar o desenvolvimento de uma ferramenta computacional que contribua para a classificação automática de anastomose em um vídeo fetoscópico por meio

de redes neurais convolucionais de forma a servir de suporte para o médico iniciante em seu estágio de treinamento.

Para isso, um DataSet (conjunto de imagens) foi construído a partir de vídeos fetoscópicos não classificados que foram então catalogados em três categorias, selecionadas em conjunto com os especialistas, usando uma ferramenta computacional especialmente projetada (VideoLabel). A técnica de aumento de dados serviu para construir imagens artificiais a partir das reais já classificadas, uma vez que o número de imagens rotuladas não era suficiente; Neste contexto, a arquitetura AlexNet foi selecionada para realizar uma transferência de aprendizagem e para treinamento, que rendeu resultados com uma eficácia superior a 90%. Esses dados permitem concluir que, na ausência de uma ferramenta que possibilitasse o treinamento de novos médicos na identificação de anastomoses, o seu treinamento demoraria mais, pois seria realizado in loco em cada cirurgia fetal. No entanto, esta pesquisa gerou o desenho de um software com o qual é viável a classificação automática de anastomoses de um vídeo fetoscópico utilizando uma rede neural convolucional, com resultados promissores como ferramenta de apoio a médicos em estágio de formação.

**Palavras-chave:** aprendizado de máquina, aprendizado supervisionado, inteligência artificial, síndrome de transfusão de gêmeos.

**Fecha Recepción:** Julio 2020

**Fecha Aceptación:** Febrero 2021

---

## Introducción

El síndrome de transfusión gemelo a gemelo (*the twin-twin transfusion syndrome* o *TTTS* por sus siglas en inglés) es el resultado de un flujo sanguíneo desigual a través de anastomosis (que es la conexión de los vasos sanguíneos) vasculares placentarias que enlazan las dos circulaciones fetales. Estas anastomosis vasculares en la placenta compartida están presentes en prácticamente todos los embarazos gemelares monocoriónicos (*monochorionic twins* o *MC* por sus siglas en inglés), pero solo en aproximadamente 10 % conducen al síndrome de transfusión gemelo a gemelo.

La transfusión neta de sangre sucede a expensas del llamado *gemelo donante*, el cual se vuelve hipovolémico y anémico, mientras que el *gemelo receptor* se vuelve hipervolémico y policitémico. Esto conduce al rápido desarrollo de una discordancia significativa en el volumen de líquido amniótico entre los gemelos que sin la debida intervención puede ser fatal para ambos gemelos (Spruijt, Lopriore, Steggerda, Slaghekke y van Klink, 2019).



Actualmente, el síndrome de transfusión gemelo a gemelo es una de las circunstancias más mortales en medicina fetal, y sigue siendo un gran reto para obstetras y neonatólogos de todo el mundo. En tal sentido, el mejor tratamiento lo representa la cirugía con láser fetoscópico. La técnica, actualmente vigente, utiliza dicho láser para interrumpir el flujo sanguíneo a través de las comunicaciones vasculares con el fin de coagular solo los vasos anastomóticos en el ecuador vascular que está a cierta distancia de la membrana entrelazada debido a los volúmenes discordantes de líquido amniótico entre los gemelos (Glennon, Shemer, Palma-Dias y Umstad, 2016).

La cirugía láser para síndrome de transfusión gemelo a gemelo se puede dividir en dos pasos fundamentales: 1) identificación endoscópica de las anastomosis vasculares placentarias, y 2) ablación con láser de las anastomosis. Por lo tanto, el objetivo de la cirugía con láser es identificar y extirpar correctamente todas las anastomosis vasculares. Esto plantea dos cuestiones: 1) ¿se pueden identificar todas las anastomosis vasculares placentarias?, y 2) ¿se pueden extirpar todas las anastomosis vasculares placentarias? (Quintero, Kontopoulos y Chmait, 2016).

Ahora bien, centrándose en la detección endoscópica de las anastomosis vasculares placentarias, surge la siguiente pregunta: ¿cómo se pueden identificar las anastomosis? Algunos autores sugirieron que las anastomosis podrían distinguirse según su apariencia, es decir, en función de patrones específicos, ángulos, con dibujos destinados a ayudar en este proceso, esto es, la identificación basada en el reconocimiento de patrones. Desafortunadamente, esto no fue posible para los médicos debido a la gran cantidad de patrones que las anastomosis pueden tener (De Lia, Kuhlmann, Cruikshank y O'Bea, 1993).

De lo anterior surgen las siguientes preguntas de investigación: ¿puede una herramienta computacional ayudar en la identificación de las anastomosis dentro de un video fetoscópico?, y ¿qué tan exacta puede ser dicha identificación? El presente trabajo plantea como hipótesis que una técnica computacional puede identificar las anastomosis con cierto porcentaje de exactitud.

Distinguir las anastomosis para realizar una adecuada intervención en la cirugía láser para síndrome de transfusión gemelo a gemelo constituye una problemática para el médico con menos experiencia, por lo que, en aras de atender esta necesidad, se plantean los siguientes objetivos: demostrar que una herramienta computacional puede ayudar en la identificación de anastomosis en un video fetoscópico con cierto porcentaje de exactitud, y

crear una herramienta computacional que ayude en la clasificación automática de anastomosis, de manera que sirva al médico novel como herramienta de apoyo en su etapa de entrenamiento y habilitación.

La ciencia de solucionar problemas clínicos, a través del análisis de imágenes generadas en la práctica clínica, se conoce como *análisis de imágenes médicas*, y su objetivo es extraer información de manera eficiente para mejorar el diagnóstico clínico. En el presente artículo se pone de manifiesto este avance, el cual consiste en la aplicación de técnicas de aprendizaje automático para el análisis de imágenes médicas. Presenta, en la primera sección, una descripción del desarrollo del trabajo de investigación realizado con varios apartados que dan claridad sobre el proceso seguido en aras de alcanzar los objetivos previstos, destacando la manera en que el aprendizaje profundo se utiliza con éxito como una herramienta para el aprendizaje automático, donde una red neuronal es capaz de aprender características automáticamente. De las técnicas de aprendizaje profundo, las redes convolucionales profundas se usan activamente para el análisis de imágenes médicas. Esto incluye áreas de aplicación como segmentación, detección de anomalías, clasificación de enfermedades, diagnóstico asistido por computadora y recuperación (Anwar *et al.*, 2018). Pero ¿a qué se hace referencia cuando se habla de redes neuronales convolucionales?

Una red neuronal convolucional es un tipo de modelo de aprendizaje profundo para procesar datos que tiene un modelo de cuadrícula —como las imágenes— inspirado en la organización de la corteza visual animal y diseñado para aprender de forma automática y adaptativa jerarquías espaciales de características, desde niveles de patrones bajos hasta de alto nivel. Una neuronal convolucional es una construcción matemática que generalmente se compone de tres tipos de capas (o bloques de construcción): convolución, agrupación y capas completamente conectadas. Las dos primeras realizan extracción de características, mientras que la tercera —una capa totalmente conectada— asigna las características extraídas a la salida final, como la clasificación.

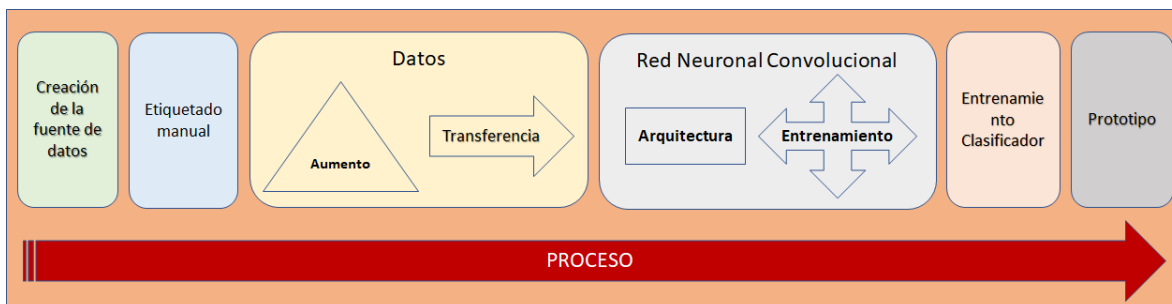
Una capa de convolución juega un papel clave en la neuronal convolucional, que se compone de una pila de operaciones matemáticas, como la convolución, un tipo especializado de operación lineal. En las imágenes digitales, los valores de píxeles se almacenan en una cuadrícula bidimensional (dos dimensiones), es decir, una matriz de números, y se aplica una pequeña cuadrícula de parámetros llamada *kernel*, un extractor de características optimizable, en cada posición de imagen, lo que hace que las neuronales

convolucionales sean altamente eficientes para el procesamiento de imágenes, ya que puede aparecer una característica en cualquier lugar de la imagen. A medida que una capa alimenta su salida a la siguiente capa, las entidades extraídas pueden volverse jerárquicas y progresivamente más complejas. El proceso de optimización de parámetros —como los núcleos— se denomina *capacitación*, que se realiza para minimizar la diferencia entre las salidas y las etiquetas de verdad del terreno a través de un algoritmo de optimización llamado *propagación hacia atrás y descenso de gradiente*, entre otros (Yamashita, Nishio, Do y Togashi, 2018).

## Método

En la presente sección se explica la metodología que se siguió para proponer una forma de resolver la problemática presentada, desde la recolección de los datos de entrada, su tratamiento y transformación hasta la propuesta de una herramienta que ayude a su resolución; en la figura 2 se puede visualizar en forma gráfica la metodología seguida.

**Figura 1.** Metodología seguida



Fuente: Elaboración propia

### Creación de la fuente de datos

En el contexto de este trabajo, el proceso general del origen de datos proporcionados fueron videos en formato MPG, grabados durante la intervención quirúrgica a pacientes reales cuando se les aplicó la cirugía fetal; los videos fueron extraídos directamente del fetoscópio por el médico a cargo de la intervención e incluyen los siguientes datos: fecha, tratamiento, cirujano, ID del paciente, fecha de nacimiento y sexo, en un archivo XML o HTML. Son generalmente  $N$  videos con una estructura de nombre *Video\_001.mpg*,



...*Video\_00N.mpg* para cada paciente; para esta investigación se tuvo acceso a  $N$  videos de 12 pacientes.

Posteriormente, se creó un programa para unir todos los videos separados de cada paciente (de dicho proceso se obtuvieron 12 videos completos), de ese modo se pudo tenerlos en formato MP4 para su posterior uso de clasificación de anastomosis con la herramienta Videolabel por el médico experto.

Aunado a esto, de los 12 videos completos se extrajeron los *frames* de cada uno y se renombraron con una nomenclatura adecuada para tener un *dataset* de imágenes sin clasificar. La nomenclatura fue la siguiente:

$F = \text{Frame}$ .

$P1, P2, P3 \dots PN = \text{Paciente}$ .

0000000001 = Número consecutivo.

Ejemplo:

$F-P1-0000000001.png$

Para concluir, en esta fase se elaboró un programa en MatLab, donde se revisó todo el directorio que contiene imágenes y se las recortó de manera automática con la finalidad de dejar solo la imagen en sí y eliminar lo sobrante. Esta actividad se realizó porque se trabaja con miles de imágenes, las cuales tendrán que ser procesadas por la red neuronal convolucional. Esta tarea de eliminación de segmentos permite ahorrar tiempo, y arroja como resultado un DataSet de imágenes sin clasificar.

## Etiquetado manual

Para el etiquetado del DataSet de imágenes se desarrolló un *software* que ayuda a clasificar las imágenes (llamado VideoLabel). Este *software* clasificador fue usado por los expertos médicos con conocimientos del tema; su funcionamiento, en forma general, consiste en leer un video ya procesado. El *software* permite a un médico experto —en caso de identificar alguna anastomosis— pausar el video, marcar el elemento y guardar el tipo de anastomosis identificada mientras internamente el programa guarda dicha imagen clasificada. Las clases de anastomosis que ya tiene cargado el programa VideoLabel son las siguientes: normal, arterio-venosa, veno-arterial, arterio-arterial, veno-venosa, borde placentario, membrana interfetal, inserción del cordón del gemelo donante, inserción del cordón del gemelo receptor, capilar.

La salida del programa VideoLabel es el almacenamiento de las imágenes clasificadas en una carpeta que conforman el DataSet de imágenes etiquetadas por el experto. En este trabajo de investigación se trabajó con las categorías *arterio-venosa*, *capilar* y *veno-arterial*; en la figura 2 se pueden visualizar los tres tipos de anastomosis con los que se trabajó:

**Figura 2.** Categorías de imágenes del trabajo de investigación



Fuente: Elaboración propia

Cabe resaltar que no se presentan todos los tipos de anastomosis en las intervenciones; de igual manera, es complicado reunir un DataSet de imágenes numeroso, ya que el tiempo dedicado al etiquetado manual por el médico experto es escaso. En la siguiente sección se especifican las cantidades de imágenes reunidas y etiquetadas por el experto.

### **Aumento de datos**

En general, las redes neuronales convolucionales son demasiado profundas como para entrenarlas desde cero con conjuntos de datos pequeños (se necesitan al menos mil imágenes clasificadas por categoría) para entrenamiento desde cero. Para eso, se recomiendan las técnicas de transferencia de aprendizaje y aumento de datos; en ese sentido, la técnica de aumento de datos reside en incrementar artificialmente el grupo de imágenes de entrenamiento empleando diversos reajustes a las imágenes originales, como modificar el brillo, escalar, acercar, girar, reflejar vertical y horizontalmente, etc. Las imágenes obtenidas deben ser lo más aceptablemente reales como para que un espectador externo no pueda distinguir entre una imagen creada por aumento de datos y una original (Gómez-Ros *et al.*, 2019).

Aunque existen diversas estrategias para aumento de datos, las anteriores se han explorado solo con imágenes naturales, y no con imágenes médicas. Entre las estrategias para aumento de datos en imágenes médicas se recomiendan las siguientes, ya que fueron las que dieron mejores resultados, pues incrementaron el porcentaje de efectividad en el entrenamiento de redes neuronales convolucionales: *flips* (dar la vuelta), *rotate* (rotación), *gaussian blur* (difuminado gaussiano) (Hussain, Gimenez, Yi y Rubin, 2018).

En esta investigación se crearon 2320 imágenes artificiales para aumentar el número de imágenes totales por categoría; en la tabla 1 se muestra el conjunto de las imágenes reales y artificiales que forman el DataSet.

**Tabla 1.** Resumen de imágenes clasificadas y generadas

<b>CLASIFICACIÓN</b>	<b>REALES</b>	<b>FLIP</b>	<b>ROT</b>	<b>BLUR</b>	<b>TOTAL</b>
<b>Arterio-venosa</b>	72	216	792	144	1224
<b>Capilar</b>	29	87	319	58	493
<b>Veno-arterial</b>	44	132	484	88	748
<b>Total</b>	145	435	1595	290	2465

Fuente: Elaboración propia

### **Transferencia de aprendizaje**

La transferencia de aprendizaje es una estrategia común y efectiva para entrenar una red con un conjunto pequeño de datos, donde hay una red preentrenada en un conjunto de datos extremadamente grande, como AlexNet, que contiene 1.2 millones de imágenes con 1000 clases, que luego se reutiliza y aplica a la tarea de interés dada. La suposición subyacente en la transferencia de aprendizaje es que las características genéricas (es decir, lo aprendido en un conjunto de datos lo suficientemente grande) se pueden compartir entre conjuntos de datos aparentemente dispares. Esta portabilidad de características genéricas aprendidas es una ventaja única del aprendizaje profundo que es útil en varias tareas de dominio con pequeños conjuntos de datos (Yamashita *et al.*, 2018). De igual manera, se realizó transferencia de aprendizaje de la red AlexNet preentrenada, donde se ajustaron las últimas capas (esto se explica en siguientes secciones de este artículo).

## Arquitectura de la red neuronal convolucional

Desde hace algunos años, se han expuesto varios modelos de redes neuronales convolucionales profundas, como AlexNet, ZFNet, VGG, GoogleNet, Residual Net, DenseNet, FCN y U-Net. Dichos modelos no solo han proporcionado un rendimiento de vanguardia para diversas tareas como la clasificación de imágenes, la segmentación, la detección de objetos y las tareas de seguimiento, sino que también han ofrecido un nuevo punto de vista para la fusión de imágenes (Zhou, Ruan y Canu, 2019).

Para este trabajo de investigación se ha usado la red preentrenada AlexNet —como se sugiere en Tajbakhsh *et al.* (2016)—, ya que es considerada como adecuada para realizar transferencia de aprendizaje en imágenes médicas. Además, como los conjuntos de datos siguen siendo demasiado pequeños, en lugar de ajustar todos los pesos de las capas, se han añadido dos capas completamente conectadas al final de las capas anteriores, para lo cual se agregó una capa a la arquitectura de la red neuronal convolucional. Esta capa se llama *fully connected layer* (23 en el listado de capas para AlexNet), la cual se encarga de clasificar los datos en varias clases donde las capas anteriores a ella (de la 1 a la 22) realizan la tarea de extraer las características de las imágenes.

En lo que respecta a la *classification output layer* (25 en el listado de capas de AlexNet), son las encargadas de calcular la pérdida de entropía cruzada para problemas de clasificación de múltiples clases con clases mutuamente excluyentes. La capa infiere el número de clases del tamaño de salida de la capa anterior.

Cabe señalar que la transferencia de aprendizaje ayuda a entrenar a las dos capas que se hayan modificado según las necesidades, y no a toda la red completa; en la figura 3 se puede visualizar la arquitectura de capas final de la red neuronal convolucional.

**Figura 3. Capas de la red neuronal convolucional**

1	data 227x227x3 images with 'zerocenter' normalization	Image Input	227x227x3	-	0
2	conv1 96 11x11x3 convolutions with stride [4 4] and padding [0 0 0]	Convolution	55x55x96	Weights 11x11x3x96 Bias 1x1x96	34944
3	relu1 ReLU	ReLU	55x55x96	-	0
4	norm1 cross channel normalization with 5 channels per element	Cross Channel Nor...	55x55x96	-	0
5	pool1 3x3 max pooling with stride [2 2] and padding [0 0 0]	Max Pooling	27x27x96	-	0
6	conv2 256 5x5x48 convolutions with stride [1 1] and padding [2 2 2]	Convolution	27x27x256	Weights 5x5x48x256 Bias 1x1x256	307456
7	relu2 ReLU	ReLU	27x27x256	-	0
8	norm2 cross channel normalization with 5 channels per element	Cross Channel Nor...	27x27x256	-	0
9	pool2 3x3 max pooling with stride [2 2] and padding [0 0 0]	Max Pooling	13x13x256	-	0
10	conv3 384 3x3x256 convolutions with stride [1 1] and padding [1 1 1]	Convolution	13x13x384	Weigh... 3x3x256x3... Bias 1x1x384	885120
11	relu3 ReLU	ReLU	13x13x384	-	0
12	conv4 384 3x3x192 convolutions with stride [1 1] and padding [1 1 1]	Convolution	13x13x384	Weigh... 3x3x192x3... Bias 1x1x384	663936
13	relu4 ReLU	ReLU	13x13x384	-	0
14	conv5 256 3x3x192 convolutions with stride [1 1] and padding [1 1 1]	Convolution	13x13x256	Weigh... 3x3x192x2... Bias 1x1x256	442624
15	relu5 ReLU	ReLU	13x13x256	-	0
16	pool5 3x3 max pooling with stride [2 2] and padding [0 0 0]	Max Pooling	6x6x256	-	0
17	fc6 4096 fully connected layer	Fully Connected	1x1x4096	Weights 4096x9216 Bias 4096x1	37752832
18	relu6 ReLU	ReLU	1x1x4096	-	0
19	drop6 50% dropout	Dropout	1x1x4096	-	0
20	fc7 4096 fully connected layer	Fully Connected	1x1x4096	Weights 4096x4096 Bias 4096x1	16781312
21	relu7 ReLU	ReLU	1x1x4096	-	0
22	drop7 50% dropout	Dropout	1x1x4096	-	0
23	fc 3 fully connected layer	Fully Connected	1x1x3	Weights 3x4096 Bias 3x1	12291
24	prob softmax	Softmax	1x1x3	-	0
25	classoutput crossentropyx with 'Arterio_venosa' and 2 other classes	Classification Output	-	-	0

Fuente: Elaboración propia

## Entrenamiento de la red neuronal convolucional

Para la realización tanto del entrenamiento de la red neuronal convolucional como de las pruebas en general, se presentan las características técnicas del equipo de cómputo usado: procesador Intel Core i5 7200U a 2.5Ghz, memoria RAM de 8 Gb, sistema operativo Windows 10 de 64 bits, y tarjeta gráfica Nvidia GeForce 940mx 2 Gb de RAM.

El total de imágenes (entre reales y artificiales) para entrenar a la red neuronal convolucional fue de 2465. Al respecto, es importante resaltar que para probar la red una vez entrenada se emplearon 23 imágenes inéditas (es decir, la red no las conocía).

A continuación, se enumeran las actividades seguidas para proceder al entrenamiento de una red neuronal convolucional:

1. Inicializar todos los parámetros o pesos con valores aleatorios de las primeras capas de la red.
2. Utilizar un grupo de imágenes de entrenamiento y usarlas en el modelo.
3. Computar el total de error de las probabilidades consecuencia del modelo.
4. Extender hacia atrás para computar el error de gradiente de todos los pesos en la red y emplear el gradiente descendiente para modificar dichos valores y disminuir el error de salida (Quintero, Merchán, Cornejo y Sánchez Galán, 2018).

Un parámetro importante se refiere al número de épocas que tendrá la red neuronal convolucional; una época se refiere a una sola pasada del conjunto de entrenamiento completo (Schilling, 2016). Conocer el número de épocas para entrenar a la red neuronal convolucional en imágenes médicas es un factor importante para tener un parámetro de decisión; algunos autores recomiendan 50 épocas (Kumar, Kim, Lyndon, Fulham y Feng, 2017; Petscharnig y Schöffmann, 2018).

La figura 4 muestra la red neuronal convolucional realizando el entrenamiento para un máximo de 50 épocas. Como resultado se obtuvo un porcentaje de exactitud de 100 % y el tiempo empleado por el programa fue de 2 horas y 21 minutos aproximadamente.

**Figura 4.** Red neuronal convolucional entrenándose

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:06	12.50%	2.6806	0.0010
1	50	00:01:24	29.69%	1.8023	0.0010
1	100	00:02:45	29.69%	1.7946	0.0010
2	150	00:04:10	51.56%	1.5030	0.0010
2	200	00:05:29	48.44%	1.2927	0.0010
3	250	00:06:50	59.38%	1.1698	0.0010
3	300	00:08:06	64.06%	1.0036	0.0010
4	350	00:09:23	67.19%	0.8340	0.0010
4	400	00:10:46	85.94%	0.5112	0.0010
5	450	00:12:04	87.50%	0.4826	0.0010
5	500	00:13:27	73.44%	0.6337	0.0010
•					
•					
•					
46	4550	02:12:12	100.00%	0.0045	0.0010
46	4600	02:13:16	98.44%	0.0130	0.0010
47	4650	02:14:19	100.00%	0.0042	0.0010
47	4700	02:15:22	100.00%	0.0023	0.0010
48	4750	02:16:26	100.00%	0.0018	0.0010
48	4800	02:17:29	100.00%	0.0137	0.0010
49	4850	02:18:32	100.00%	0.0123	0.0010
49	4900	02:19:35	100.00%	0.0022	0.0010
50	4950	02:20:40	100.00%	0.0195	0.0010
50	5000	02:21:42	100.00%	0.0036	0.0010

Fuente: Elaboración propia

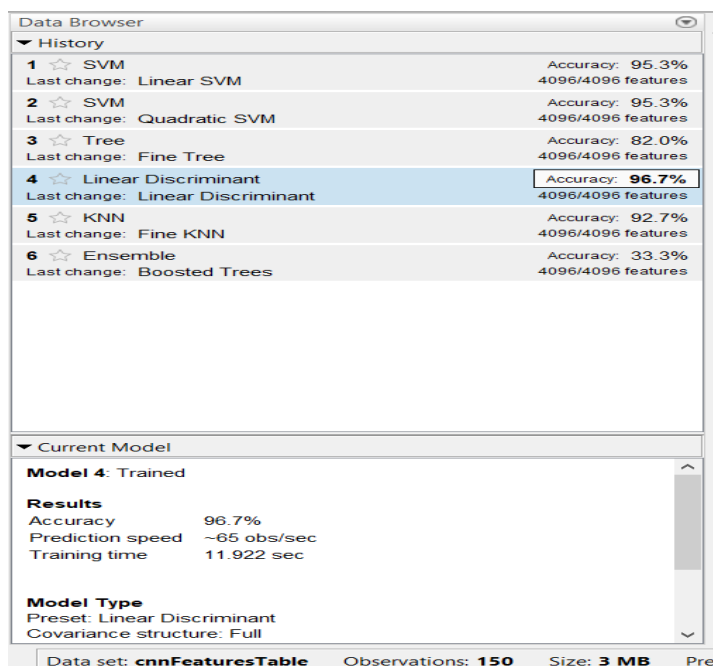
### Entrenamiento de un clasificador

Catalogar un objeto consiste en designarlo a una de las clases disponibles. Los objetos se pueden determinar por una lista de particularidades, como la textura, el tamaño o el color de sus píxeles; para lograr catalogar objetos es necesario precisar los límites entre las diferentes clases. Normalmente estos límites se computan mediante un tratamiento de entrenamiento en el que se usan las particularidades de una serie de modelos de ejemplo de las clases. Se mencionan los límites por claridad; en general, un clasificador deduce unos patrones de decisión durante el entrenamiento.

De eso se desprende que catalogar un objeto desconocido consiste en designarlo a la clase en la cual las particularidades usadas durante el entrenamiento tienen más equivalencia con las particularidades del objeto (Garrido Satué, 2013).

Una vez que se ha terminado el entrenamiento de la red neuronal convolucional, al resultado de dicho entrenamiento se le aplica el entrenamiento de un clasificador, ya que se pueden explorar los datos, seleccionar particularidades, especificar esquemas de validación cruzada, entrenar modelos y evaluar resultados. Es probable realizar entrenamiento automático para averiguar el tipo preferible de modelo clasificador (como árboles de decisión, análisis discriminante, máquinas de vector soporte, regresión logística, vecinos próximos, clasificación ensamblada). La figura 5 muestra el clasificador seleccionado con los datos de diversos clasificadores entrenados y cuál obtiene el mejor modelo de clasificador para los datos presentados.

**Figura 5.** Clasificador seleccionado



History		Accuracy
1	☆ SVM Last change: Linear SVM	95.3% 4096/4096 features
2	☆ SVM Last change: Quadratic SVM	95.3% 4096/4096 features
3	☆ Tree Last change: Fine Tree	82.0% 4096/4096 features
4	☆ Linear Discriminant Last change: Linear Discriminant	<b>96.7%</b> 4096/4096 features
5	☆ KNN Last change: Fine KNN	92.7% 4096/4096 features
6	☆ Ensemble Last change: Boosted Trees	33.3% 4096/4096 features

Current Model	
<b>Model 4: Trained</b>	
<b>Results</b>	
Accuracy	96.7%
Prediction speed	~65 obs/sec
Training time	11.922 sec
<b>Model Type</b>	
Preset:	Linear Discriminant
Covariance structure:	Full

Data set: **cnnFeaturesTable** Observations: **150** Size: **3 MB** Pre

Fuente: Elaboración propia

### Funcionamiento del prototipo

Se desarrolló un prototipo para probar el funcionamiento de la red neuronal convolucional, lo cual se hizo pensando en un usuario final que no tuviera muchos conocimientos informáticos. La pantalla cuenta con tres botones de acción: examinar, pausar y salir. En la figura 6 se puede ver una pantalla del prototipo funcionando en la clasificación de anastomosis.



La fuente de datos debe ser una carpeta que contenga el video de alguna intervención quirúrgica en formato MP4; posteriormente, el prototipo reproducirá el video de forma automática mientras la red neuronal convolucional selecciona cada *frame* o fotograma, lo analiza, y al encontrar alguna coincidencia con las categorías entrenadas las marcará con un color (verde para arterio-venosa, amarillo para capilar y azul para veno-arterial); al final guardará dicho *frame* en el dispositivo. La nomenclatura con la que se guarda la imagen es la siguiente, según las categorías seleccionadas:

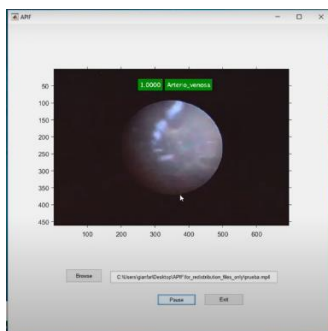
*AV-fecha-hora.png* para la categoría *arterio-venosa*.

*C-fecha-hora.png* para la categoría *capilar*.

*VA-fecha-hora.png* para la categoría *veno-arterial*.

La salida del prototipo es un conjunto de imágenes clasificadas y etiquetadas por la red neuronal convolucional, de acuerdo con las tres categorías previamente mencionadas. Con el uso del prototipo, el médico especialista pueda tener resultados para validarlos.

**Figura 6.** Muestra de funcionamiento del prototipo de *software*



Fuente: Elaboración propia

## Resultados

La primicia principal de este trabajo de investigación era determinar si una herramienta computacional podría ayudar a identificar las anastomosis dentro de un video fetoscópico y qué tan exacta sería dicha identificación. En tal sentido, durante el desarrollo se obtuvieron algunos resultados del trabajo realizado. Por ejemplo:

- Se obtuvieron 12 videos unidos completos de las intervenciones quirúrgicas con más de ocho horas de video fetoscópico.
- De los 12 videos se obtuvo un DataSet de 111 273 imágenes sin clasificar.

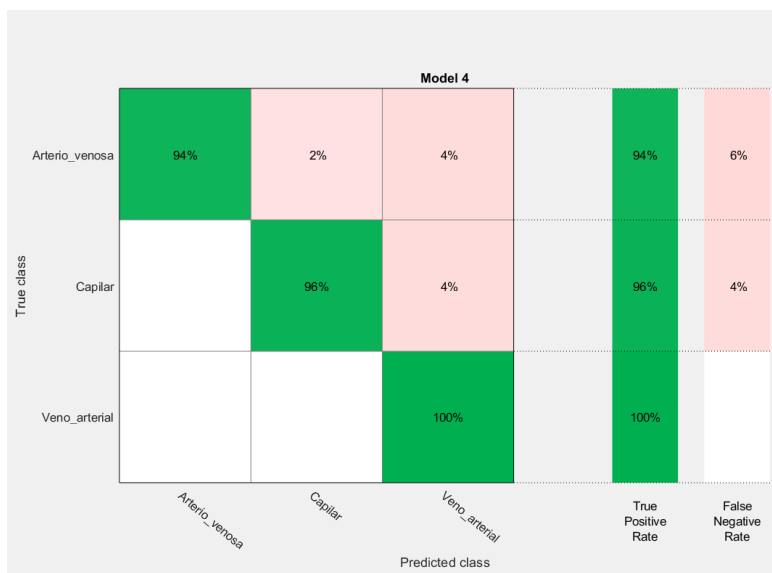
- Producto del etiquetado manual por parte del experto médico, se obtuvieron las imágenes clasificadas por categorías mostradas en la tabla 1.

Una parte importante de los resultados permite sostener que la solución propuesta funciona adecuadamente para lo que fue diseñada, por lo que se consideró esencial aplicar técnicas como la matriz de confusión, la curva ROC y el AUC para indicar si dicha metodología es aceptable.

En tal sentido, y de manera general, se puede mencionar que una matriz de confusión es una herramienta que se usa comúnmente para delinear el rendimiento de un sistema de clasificación en un cúmulo de datos en el que se conocen como los valores verdícos. La matriz abarca información de las clasificaciones reales y las predichas elaboradas por el sistema, y se emplea para computar el desempeño del algoritmo. Otra forma de examinar los rendimientos de los clasificadores son los gráficos ROC (*receiver operating characteristics*), método para visualizar, organizar y seleccionar los clasificadores fundamentado en su rendimiento, donde la curva ROC contiene toda la información de la matriz de confusión (Fos Guarinos, 2016). Asimismo, el área bajo la curva AUC (*area under curve*) es el estadístico por excelencia para medir la capacidad discriminante de la prueba y para comparar pruebas entre sí y determinar cuál es la más eficaz (Del Valle Benavides, 2017).

En la figura 7 se puede ver la matriz de confusión resultante de la experimentación llevada a cabo con tres clasificaciones distintas en la red neuronal convolucional después de entrenar la red. Véase que para la clasificación *arterio-venosa* se tiene 94 % de clasificación de verdaderos positivos y 6 % de falsos negativos en el conjunto de valores de prueba proporcionados a la red neuronal convolucional. Para la clasificación *capilar* se tiene 96 % de clasificación de verdaderos positivos y 4 % de falsos negativos en el conjunto de valores de prueba suministrados a la red neuronal convolucional. Finalmente, para la clasificación *veno-arterial* se tiene 100 % de clasificación de verdaderos positivos y 0 % de falsos negativos en el conjunto de valores de prueba proporcionados a la red neuronal convolucional.

**Figura 7.** Matriz de confusión resultante



Fuente: Elaboración propia

En cuanto a la curva ROC, se alcanzaron los siguientes resultados para cada una de las tres clasificaciones (la tabla 2 muestra la AUC que se obtuvo por cada categoría).

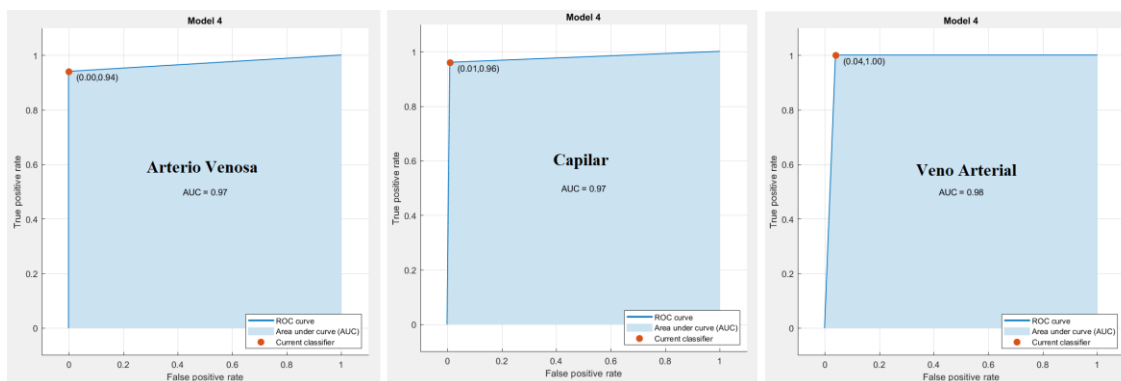
**Tabla 2.** Resultados de AUC por clasificación

CLASIFICACION	AUC
Arterio-venosa	0.97
Capilar	0.97
Veno-arterial	0.98

Fuente: Elaboración propia

Un resultado de la ROC y su AUC cuyo valor esté entre 0.5 y 1 indica que la predicción del clasificador es correcta (si tiende a 1 es casi perfecta); por el contrario, un resultado por debajo de 0.5 indica que existen inconsistencias en la clasificación de los datos. En la figura 8 se muestra cada una de las categorías con su curva ROC y su AUC.

**Figura 8.** Categorías con su curva ROC y AUC



Fuente: Elaboración propia

## Discusión

Como se mencionó en los resultados, se lograron porcentajes de efectividad superiores a 90 % en general para las tres categorías, con una tasa de imágenes reales de cerca de 150. Una limitación y área de debilidad de la investigación fue el relativo bajo número de imágenes reales que se solventó con la técnica del aumento de datos; pero ¿qué pasaría si el número de imágenes reales aumentara? Aventurando una hipótesis, se puede suponer un incremento en la tasa de efectividad.

No obstante lo anterior, la gran fortaleza de la investigación fueron los porcentajes de efectividad obtenidos teóricamente por las técnicas mencionadas, las cuales se pudieron demostrar con el prototipo de *software* creado. En consecuencia, se puede afirmar que se logró conseguir el objetivo planteado, es decir, desarrollar una herramienta computacional para contribuir a la clasificación automática de anastomosis dentro de un video fetoscópico mediante redes neuronales convolucionales, las cuales sirven de soporte en la etapa de entrenamiento del médico novel.

## Conclusiones

En el presente trabajo se explicó la creación de una herramienta computacional para ayudar al médico novel en la identificación de anastomosis dentro de un video fetoscópico, lo cual forma parte de su preparación para realizar una adecuada intervención con cirugía láser para el síndrome de transfusión gemelo a gemelo.

En este sentido, cabe indicar que la creación de un prototipo que lee y reproduce videos fetoscópicos contiene dentro de sí una red neuronal convolucional debidamente entrenada con tres categorías diferentes de imágenes de anastomosis, lo cual permite clasificarlas con un porcentaje de efectividad superior a 90 % mientras se reproduce el video. Sin embargo, vale apuntar que la capacidad de fluidez del prototipo está ligada directamente al equipo de cómputo usado, por lo que se sugiere uno con altas prestaciones (p. ej., con disco duro de estado sólido de 500 Gb como mínimo, al menos 8 Gb de memoria RAM y un sistema operativo de 64 bits).

Como producto adicional, se lograron obtener dos DataSets de imágenes: uno sin clasificar y otro debidamente clasificado por el médico o experto, que fue la base con la que se entrenó a la red neuronal convolucional empleada en la herramienta generada. Por eso, se puede asegurar que el prototipo creado será de gran ayuda para que el personal médico novel en proceso de entrenamiento realice cirugías fetoscópicas, ya que le permitirá contar con una herramienta de apoyo.

### **Futuras líneas de Investigación**

Finalmente, como líneas de investigación futuras se puede realizar un estudio comparativo y pruebas con otras arquitecturas de red neuronal convolucional (aparte de AlexNet), como VGGNet, GoogleNet, ResNet y ZFNet con la finalidad de comparar sus resultados, incrementar el número de categorías para que sea más completo el prototipo, aumentar el número de imágenes de las categorías para optimizar la exactitud de los resultados, así como aplicar aprendizaje no supervisado para clasificar videos con el objetivo de realizar ajustes en la red y fortalecerla.

### **Agradecimientos**

Este trabajo forma parte de la tesis doctoral de Ángel Mario Lerma Sánchez para su título de doctor en Sistemas Computacionales. Fue dirigida por la Dra. Adriana Mexicano Santoyo (asesora); Dra. Fabiola Mirosłaba Villalobos Castaldi, codirectora; y Dr. Pedro Damián Reyes, codirector. De igual manera, se agradece al Dr. Rogelio Cruz-Martínez por todas las facilidades para el proyecto y a la Dra. María Belén Nieto Castro por su arduo trabajo en la clasificación de imágenes y retroalimentación. Al Dr. José Ruiz Pinales por su

valioso apoyo con la herramienta VideoLabel. Asimismo, se expresa el agradecimiento a la Universidad Da Vinci por todo el apoyo proporcionado.

## Referencias

- Anwar, S. M., Majid, M., Qayyum, A., Awais, M., Alnowami, M. and Khan, M. K. (2018). Medical Image Analysis using Convolutional Neural Networks: A Review. *J Med Syst*, 42(11). Doi: 10.1007/s10916-018-1088-1
- De Lia, J. E., Kuhlmann, R. S., Cruikshank, D. P. and O'Bee, L. R. (1993). Current topic: Placental surgery: a new frontier. *Placenta*, 14(5), 477–485. Doi: 10.1016/s0143-4004(05)80201-2
- Del Valle Benavides, A. (2017). *Curvas ROC (Receiver-Operating-Characteristic) y sus aplicaciones* (tesis de pregrado). Sevilla, España: Universidad de Sevilla. Recuperado de <https://idus.us.es/bitstream/handle/11441/63201/Valle%20Benavides%20Ana%20Roc%20C3%ADo%20del%20TFG.pdf?sequence=1>
- Fos Guarinos, B. (2016). *Diseño de técnicas de inteligencia artificial aplicadas a imágenes médicas de rayos X para la detección de estructuras anatómicas de los pulmones y sus alteraciones* (tesis de pregrado). Valencia, España: Universidad Politécnica de Valencia.
- Garrido Satué, M. (2013). *Reconocimiento de señales de tráfico para un sistema de ayuda a la conducción* (tesis de maestría). Sevilla, España: Universidad de Sevilla, Escuela Técnica Superior de Ingeniería. Recuperado de [https://riunet.upv.es/bitstream/handle/10251/70103/73657675Z\\_TFG\\_14733814699245714722695301721227.pdf?sequence=2&isAllowed=y](https://riunet.upv.es/bitstream/handle/10251/70103/73657675Z_TFG_14733814699245714722695301721227.pdf?sequence=2&isAllowed=y)
- Glennon, C. L., Shemer, S. A., Palma-Dias, R. and Umstad, M. P. (2016). The History of Treatment of Twin-to-Twin Transfusion Syndrome. *Twin Research and Human Genetics*, 19(3), 168-174. Doi: 10.1017/thg.2016.27
- Gómez-Ros, A., Tabik, S., Luengo, J., Herrera, F., Shihavuddin, A. and Krawczyk, B. (2019). Redes neuronales convolucionales para una clasificación precisa de imágenes de corales. *XVIII Conferencia de la Asociación Española para la Inteligencia Artificial*, 1171-1176. Recuperado de [https://backend.orbit.dtu.dk/ws/files/166595939/CAEPIA2018\\_paper\\_212.pdf](https://backend.orbit.dtu.dk/ws/files/166595939/CAEPIA2018_paper_212.pdf)

- Hussain, Z., Gimenez, F., Yi, D. and Rubin, D. (2018). Differential Data Augmentation Techniques for Medical Imaging Classification Tasks. *AMIA Annu Symp Proc.*, 2017, 979–984. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977656/pdf/2730723.pdf>
- Kumar, A., Kim, J., Lyndon, D., Fulham, M. and Feng, D. (2017). An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification. *IEEE Journal of Biomedical and Health Informatics*, 21(1), 31-40. Doi: 10.1109/JBHI.2016.2635663
- Petschnig, S. and Schöffmann, K. (2018). Learning laparoscopic video shot classification for gynecological surgery. *Multimedia Tools and Applications*, 77, 8061–8079. Doi: <https://doi.org/10.1007/s11042-017-4699-5>
- Quintero, C., Merchán, F., Cornejo, A. y Sánchez Galán, J. (2018). Uso de redes neuronales convolucionales para el reconocimiento automático de imágenes de macroinvertebrados para el biomonitorio participativo. *KnE Engineering*, 3, 585-596. Recuperado de <https://knepublishing.com/index.php/KnE-Engineering/article/view/1462/3527>
- Quintero, R. A., Kontopoulos, E. and Chmait, R. H. (2016). Laser Treatment of Twin-to-Twin Transfusion Syndrome. *Twin Research and Human Genetics*, 19(3), 197-206. Doi: <https://doi.org/10.1017/thg.2016.28>
- Schilling, F. (2016). *The Effect of Batch Normalization on Deep Convolutional Neural Networks* (master thesis). Stockholm, Sweden: KTH Royal Institute of Technology School of Computer Science and Communication. Retrieved from <https://kth.diva-portal.org/smash/get/diva2:955562/FULLTEXT01.pdf>
- Spruijt, M. S., Lopriore, E., Steggerda, S. J., Slaghekke, F. and van Klink, J. M. (2019). Twin-twin transfusion syndrome in the era of fetoscopic laser surgery: antenatal management, neonatal outcome and beyond. *Expert Review of Hematology*, 13(3), 259-267. Doi: <https://doi.org/10.1080/17474086.2020.1720643>
- Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B. and Liang, J. (2016). Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Transactions on Medical Imaging*, 35(5), 1299-1312. Doi: <https://doi.org/10.1109/tmi.2016.2535302>

- Yamashita, R., Nishio, M., Do, R. and Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imaging*, 9, 611–629. Doi: <https://doi.org/10.1007/s13244-018-0639-9>
- Zhou, T., Ruan, S. and Canu, S. (2019). A review: Deep learning for medical image segmentation using multi-modality fusion. *Array*, 3-4, 100004. Doi: <https://doi.org/10.1016/j.array.2019.100004>



<b>Rol de Contribución</b>	<b>Autor (es)</b>
<b>Conceptualización</b>	Angel Mario Lerma Sánchez, Fabiola Miroslaba Villalobos Castaldi, Adriana Mexicano Santoyo, Pedro Damián Reyes Grado de contribución igual
<b>Metodología</b>	Angel Mario Lerma Sánchez, Adriana Mexicano Santoyo, Fabiola Miroslaba Villalobos Castaldi, Pedro Damián Reyes Grado de contribución igual
<b>Software</b>	Angel Mario Lerma Sánchez (Principal), José Ruiz Pinales (Apoyo)
<b>Validación</b>	María Belén Nieto Castro (Principal), Angel Mario Lerma Sánchez (Apoyo)
<b>Análisis Formal</b>	Angel Mario Lerma Sánchez, Adriana Mexicano Santoyo, Fabiola Miroslaba Villalobos Castaldi, Pedro Damián Reyes Grado de contribución igual
<b>Investigación</b>	Angel Mario Lerma Sánchez, Fabiola Miroslaba Villalobos Castaldi Grado de contribución igual
<b>Recursos</b>	Rogelio Cruz-Martínez (Principal),
<b>Curación de datos</b>	María Belén Nieto Castro (Principal), Angel Mario Lerma Sánchez (Apoyo)
<b>Escritura - Preparación del borrador original</b>	Angel Mario Lerma Sánchez (Principal)
<b>Escritura - Revisión y edición</b>	Adriana Mexicano Santoyo (Principal), Angel Mario Lerma Sánchez (Apoyo)
<b>Visualización</b>	Fabiola Miroslaba Villalobos Castaldi (Principal), Angel Mario Lerma Sánchez (Apoyo)
<b>Supervisión</b>	Adriana Mexicano Santoyo (Principal), Angel Mario Lerma Sánchez (Apoyo)
<b>Administración de Proyectos</b>	Angel Mario Lerma Sánchez, Adriana Mexicano Santoyo, Fabiola Miroslaba Villalobos Castaldi, Pedro Damián Reyes Grado de contribución igual
<b>Adquisición de fondos</b>	Angel Mario Lerma Sánchez