

Automatic Blood-Cell Classification via Convolutional Neural Networks and Transfer Learning

Luis Claudio Soto-Ayala, Jose Antonio Cantoral-Ceballos *Member, IEEE*

Abstract—The evaluation and diagnosis of cancer related diseases can be complex and lengthy. This is exacerbated due to manual analyses based on techniques that may take copious amount of time. In the last decade, different tools have been created to detect, analyze and classify different types of cancer in humans. However, there is still a lack of tools or models to automate the analysis of human cells to determine the presence of cancer. Such a model has the potential to improve early detection and prevention of said diseases, leading to more timely medical diagnoses. In this research, we present our current effort on the development of a Deep Learning Model capable of identifying blood cell anomalies. Our results show an accuracy that meets or exceeds the current state of the art, particularly achieving lower false negative rate in comparison to previous efforts reported.

Index Terms—Cancer Cell Detection, Convolutional Neural Networks, Residual Neural Network, Transfer Learning.

I. INTRODUCCIÓN

Realizar diagnósticos médicos es una tarea complicada que se debe llevar a cabo en un tiempo limitado y en ocasiones crítico, dependiendo de la gravedad del paciente. Por otro lado, la intensa carga de trabajo que experimentan los médicos durante su jornada laboral, en combinación con el limitado número de especialistas, pueden retrasar dichos diagnósticos aún más. Estos retrasos pueden llevar a consecuencias irreversibles en donde un tratamiento crítico se posponga debido a un diagnóstico tardío. Un caso particularmente difícil de diagnosticar es la leucemia, dado que los síntomas son similares a los de otras enfermedades, además de no presentar tumores. Por esto, su diagnóstico es mediante un proceso de laboratorio con base en juicios de expertos, mediante la visualización y clasificación manual de los diferentes tipos de células presentes en la sangre periférica.

Mediante el uso de algoritmos de *Machine Learning* (ML) se han desarrollado herramientas para analizar, detectar, e incluso clasificar algunos tipos de cáncer y otras enfermedades, sin embargo, ninguno de los trabajos estudiados durante esta investigación realiza un análisis y clasificación de imágenes a nivel celular [1]–[5]. No obstante, es posible encontrar aplicaciones de ML relacionadas con el estudio de órganos como el corazón [6], los pulmones [7], la piel [8], y otros tejidos del cuerpo humano [9], [10].

En este trabajo se presenta un algoritmo basado en Redes Neuronales Convolucionales (CNNs) capaz de discriminar de manera binaria la existencia de cáncer en los glóbulos blancos

pertenecientes a la sangre periférica en los seres humanos, mediante el uso de imágenes de las mismas, con la finalidad de, eventualmente, poder utilizarse como herramienta de apoyo durante diagnósticos médicos.

De esta forma, la principal contribución de este trabajo es definida como:

- Modelo basado en redes neuronales convolucionales y *transfer learning* para la detección de anomalías en células sanguíneas por leucemia, con bajo índice de falsos negativos.

El resto de este artículo se divide de la siguiente forma. La Sección II describe los trabajos más relevantes relacionados a esta investigación, basados en aplicaciones de *Machine Learning* en el campo médico. Además, se comparan los trabajos señalados durante este capítulo y las áreas de oportunidad en los mismos. En la Sección III se detalla la problemática existente dentro del campo médico y se justifica el desarrollo de una herramienta auxiliar para realizar diagnósticos médicos referentes al cáncer. La Sección IV y V presentan la propuesta de solución para el problema mencionado, así como la metodología para evaluar la solución planteada, respectivamente. El proceso de experimentación y resultados son presentados en la sección VI. Finalmente, en la sección VII se presentan las conclusiones finales a partir de los resultados obtenidos.

II. ANTECEDENTES

A. El Cáncer y sus Antecedentes

El cáncer es el nombre asignado a la colección de enfermedades en las cuales algunas células del cuerpo comienzan a dividirse sin control, y así se extienden a tejidos circundantes. El cáncer puede originarse en prácticamente cualquier parte del cuerpo, mismo que está formado por miles de millones de células. En situaciones normales, el cuerpo genera células sanas, las cuales maduran, envejecen y mueren, para que posteriormente otra célula tome su lugar. Este proceso de regeneración celular se ve afectado en presencia de células cancerígenas en un ser humano, ya que las células malignas tienen la capacidad de sobrevivir y multiplicarse, al punto de crear estructuras llamadas tumores [11]. Según la Sociedad Americana de Cáncer (ACS), ésta es la diferencia principal entre las células cancerígenas y las normales, [12], su capacidad de dividirse sin control ignorando las señales de muerte y sustitución celular.

La leucemia es un tipo de cáncer relacionado con las células de la sangre llamadas glóbulos blancos (leucocitos), que no suele desarrollar tumores. Además, se desarrolla principalmente en los tejidos blandos que se encuentran en el interior de los huesos donde se forman dichas células. Los leucocitos tienen como objetivo combatir cualquier tipo de afección en los seres humanos, sin embargo, en el caso de leucemia, la médula ósea produce una cantidad anormal de glóbulos blancos que al desplazarse al torrente sanguíneo impiden que se produzcan glóbulos rojos, plaquetas, y leucocitos maduros sanos [13].

B. Trabajos Relevantes

En años recientes, se han desarrollado métodos efectivos para el análisis y clasificación de imágenes [14]–[18]. Con el desarrollo e implementación de redes neuronales profundas (DNN) y, en particular, Redes Neuronales Convolucionales (CNN), el análisis de imágenes ha alcanzado un desempeño que en ocasiones excede el de un humano, con importantes casos de éxito en el área médica [1]. Por ejemplo, permitiendo la identificación automática de patrones en condiciones como el cáncer, reduciendo errores humanos y aumentando la velocidad de detección [1].

Otra aplicación de DNN es ADSC (Automatic Detector of Skin Cancer) [1], que presenta un modelo capaz de realizar detecciones tempranas de cáncer de piel en sus diferentes categorías, utilizando un nuevo tipo de optimización de los algoritmos ya existentes, basándose enteramente en el uso de imágenes para abstraer patrones más propensos a ser clasificados como un tipo de cáncer de piel. Esta solución se basa en la optimización de un método estocástico denominado algoritmo de optimización de ballenas (WOA por sus siglas en inglés) [19]. Este método comienza con un set de candidatos tipo solución para encontrar el menor y mayor óptimos, mientras que continúa actualizando la solución durante todo el proceso, hasta satisfacer la condición de finalización, logrando así 91% de efectividad, aunque con un alto costo computacional [1].

T. Zhang, et al. [2] expresan que la combinación de las cargas de trabajos excesivas de los médicos y el uso bases de datos incompletas en el entrenamiento de algunas redes neuronales pueden derivar en errores de diagnóstico y baja eficiencia de los algoritmos. Así, ellos se enfocan en el desarrollo de una herramienta con capacidades de clasificación de lesiones cervicales mediante un pre-procesamiento de los datos e incremento de la información utilizada. Para conseguirlo, se utilizaron 1,709 casos de información experimental de pacientes con uso de imágenes de colonoscopia cervical. Posteriormente, se realiza un pre-procesamiento de las imágenes mediante Extracción de la Región de Interés (por sus siglas en inglés ROI), aumento de datos, y *transfer learning*. La CNN pre-entrenada (con un dataset de Kaggle: "Intel and MobileODT Cervical Cancer Screening") para realizar las pruebas se basó en un modelo DenseNet, que interconecta todas las capas, mejorando así el flujo de información entre capas, y la retro-propagación de gradiente. En comparación con trabajos similares en los cuales el promedio de efectividad oscilaba entre el 70 y 73%, con los ajustes realizados se alcanzó una efectividad del 76.33% [2].

En otra aplicación médica, Yadav y Jadhav [3] buscaron clasificar y capturar patrones de neumonía, con el objetivo final de probar diferentes métodos para mejorar el desempeño de la clasificación y hacer más eficiente el proceso de etiquetado de datos mediante *transfer learning*. Para esto, se desarrolló una CNN profunda capaz de recibir imágenes de radiografías torácicas. Esta red neuronal se entrenó con un dataset proporcionado por una investigación realizada por Kermnay et al. [20] que contiene 2 tipos de radiografías: normales y con neumonía, a su vez existían dos sub-clasificaciones para neumonía viral y bacteriana. La herramienta de diagnóstico utilizó diferentes técnicas y algoritmos para realizar una comparación sobre el rendimiento, entre los que se encuentran *Support Vector Machines* (SVMs) y *transfer learning*. Después de realizar varias pruebas con los diferentes algoritmos usando el mismo dataset, el estudio reveló que para el análisis y clasificación de imágenes el uso de *transfer learning* es de suma importancia para obtener el máximo desempeño y exactitud al momento de realizar el diagnóstico [3].

Por otro lado, Sairam Tabibu et al. [4] integraron la clasificación de carcinoma renal junto con la esperanza de supervivencia del paciente basado en imágenes histopatológicas, mediante la clasificación de células malignas en tejidos renales y abstracción de patrones para predecir las esperanza de supervivencia. El modelo utilizó una arquitectura RESNET18 y RESNET34 pre-entrenadas en ImageNET, logrando una efectividad de la clasificación de los subtipos de cáncer renal del 92.62% y 91.96%, respectivamente. Este algoritmo, haciendo uso de información relacionada con la forma del tumor, como el perímetro y como la forma del núcleo en los tejidos, puede calcular la tasa de supervivencia del paciente utilizando el modelo Lasso-Regularized COX. A pesar de que la tasa de efectividad de las pruebas en general está en el rango del 83%, los resultados sugieren que existen clases mal identificadas al final del análisis. Además, los resultados revelaron que el desequilibrio existente en los datos produce clasificaciones erróneas en algunos subtipos de tumores.

En este sentido, en el caso de bases de datos médicas, dos problemas importantes son el desbalanceo de clases (más muestras de pacientes típicos que atípicos) y limitada variabilidad de pacientes. Para combatir estos problemas, una estrategia viable es el uso de ensambles de diferentes modelos o arquitecturas [21]. Por ejemplo, Lopez-Martin et al. [22] presentan un clasificador binario basado en un ensamble de sub-modelos idénticos que comparten parámetros, enfocados a la detección temprana de Alzheimer, alcanzando resultados que exceden el desempeño de algoritmos de *machine learning* o *deep learning* solos. En relación con cáncer de seno, Sara Hosseinzadeh Kassani et al. [23] presentan un modelo novedoso basado en un ensamble de tres arquitecturas diferentes pre-entrenadas de CNNs (VGG19, MobileNet, DenseNet) que permiten la extracción de patrones a partir de imágenes histopatológicas de senos, que posteriormente son clasificados binariamente mediante una capa tipo *fully connected*, alcanzando una efectividades superiores a alternativas basadas en un solo algoritmo de *machine learning* de hasta 98.13%, 95.00%, 94.64%, y 83.10% en las bases de datos BreakHis, ICIAR, PathCamelyon, y Bioimaging, respectivamente.

TABLE I
TABLA COMPARATIVA DE TRABAJOS EN EL ESTADO DEL ARTE

Trabajo	Tipo Imagen	Hardware	Costo	Exactitud
Automatic Detector of Skin Cancer	Dermatoscopia (640 X 480)	I7-49790k GTX TITAN X	\$2542	91%
Clasificación de Lesiones pre-cancerosas Cervicales	Colonoscopia (224 X 224)	I7-7700K GTX 1060ti	\$803	76.33%
Clasificación de Neumonía	Radiografías Torácicas (64 X 64)	I7-6700Hq Nvidia Tesla K80	\$1296	No reportado
Clasificación de Carcinoma en Tejidos Renales	Imágenes Histopatológicas (224 X 224)	No reportado	No reportado	91 y 92.62%
Clasificación de imágenes histopatológicas de biopsias	Imágenes Histopatológicas Tamaño depende de base de datos	I7-8700K 3.7 GHz GTX 1080Ti	No reportado	98.13%, 95.00%, 94.64%, y 83.10%

La Tabla I presenta un resumen de factores cuantitativos relevantes para realizar una comparación y análisis fácil de visualizar. Como se puede apreciar, todos los trabajos utilizan imágenes de distintas índoles donde el tamaño resultó ser un factor determinante, los proyectos donde se utilizaron imágenes más grandes obtuvieron mejores resultados, sin embargo, el tamaño del dataset y el tipo de imagen son factores importantes que deben tomarse en cuenta.

III. DEFINICIÓN DEL PROBLEMA

En el ámbito médico, la clasificación citomorfológica es una evaluación complicada que puede consumir mucho tiempo por paciente. Esto debido a que existe gran variabilidad en formas y tamaños de las células analizadas, que suelen ser difíciles de diferenciar. Este reto es particularmente complicado cuando el número de pacientes que solicitan la prueba sobrepasa al número de expertos capacitados en el área. Particularmente, este tipo de situaciones puede aumentar la carga de trabajo de los médicos especialistas capacitados para realizar estas tareas, teniendo como posible consecuencia la disminución de las capacidades cognitivas para realizar diagnósticos clínicos que pueden causar cierto nivel de error [24], [25]. Además, el tiempo transcurrido entre el escaneo, interpretación, y tratamiento pueden ser considerables, aún cuando el paciente necesita atención urgente [25].

Por lo tanto, en el contexto del diagnóstico oportuno y discriminación de células anómalas, es de gran importancia mejorar el tiempo de respuesta, así como el porcentaje de aciertos en el diagnóstico de padecimientos, para facilitar la labor de los especialistas médicos.

Derivado de la investigación del estado del arte que se ha realizado durante este trabajo de investigación, observamos que no existe un modelo o herramienta capaz de discriminar de forma binaria las células humanas pertenecientes al riego sanguíneo periférico y capaz de realizar un análisis de precisión de incidencias atípicas porcentuales. Además, los

trabajos presentados no incluyen análisis celular en baja escala, ya que éstos se centran en el uso de tejidos humanos, por lo cual se presenta un área de oportunidad para determinar la existencia de leucemia mediante el análisis celular sanguíneo periférico.

IV. SOLUCIÓN PROPUESTA

A. Red Neuronal Convolutacional

En [26], Yann LeCun et al. propusieron una alternativa a los modelos de visión computacional dominantes en la época, que buscaba la automatización del aprendizaje en lugar del desarrollo manual de las heurísticas para la solución de ciertos problemas, como el reconocimiento de caracteres. Esto fue posible mediante la extracción de características directamente de los datos de forma automática, cambiando de manera radical la forma de detección de patrones. Esto llevó al desarrollo de un paradigma llamado redes de transformadores gráficos (en inglés *Graph Transformer Networks*) que mediante el entrenamiento paulatino de los módulos optimizan el desempeño global con base en un criterio definido, generando así las redes neuronales convolucionales, en las cuales se basa el modelo de la presente investigación.

En este trabajo se presentan dos modelos, como punto de partida se diseñó un modelo con dos capas convolucionales, con *Max Pooling*, seguidas de 2 capas tipo perceptrón multicapa (MLP) o *fully connected* de 120 y 84 neuronas, respectivamente, que permiten la abstracción de las características fundamentales de los datos utilizados. Este modelo se implementa para un set de datos con imágenes de 32 x 32 píxeles que utilizan 3 canales para su representación de color (R,G,B). Como función de activación se utilizó ReLU (*Rectified Linear Unit*) [27] que es comúnmente utilizada por diversos trabajos dentro del estado del arte por su simpleza matemática y capacidad de retro-propagación del gradiente. Finalmente, se realiza un proceso de *Max Pooling* permitiendo mantener los píxeles más salientes y reduciendo la cantidad de datos de salida del modelo desarrollado. Una vez procesada la imagen, ésta pasa a través de las capas de la red neuronal MLP, como se muestra en la figura 1.

Además, se presenta un modelo basado en redes pre-entrenadas RESNET18 y RESNET34 con el objetivo de evaluar el impacto de utilizar *transfer learning*. Los modelos basados en RESNET18 y RESNET34 fueron pre-entrenados utilizando ImageNET [29], esto considerando que la mayor capacidad de abstracción de un modelo más grande puede ser transferida al problema estudiado en este trabajo a pesar de haber sido entrenado en un dominio diferente. Esto se justifica debido a que el modelo pre-entrenado abstrae patrones básicos (por ejemplo líneas, formas geométricas en dos dimensiones y colores) principalmente en capas iniciales [30], que también son comunes en otros dominios de imágenes como pueden ser las imágenes médicas estudiadas en este trabajo.

La estructura para la salida de los modelos pre-entrenados son dos capas *fully connected* que requieren recibir 512 características del modelo pre-entrenado, con 100 neuronas ocultas para la segunda capa, y finalmente una capa de salida para las clases de carácter típico y atípico, como se muestra en la Figura 1.

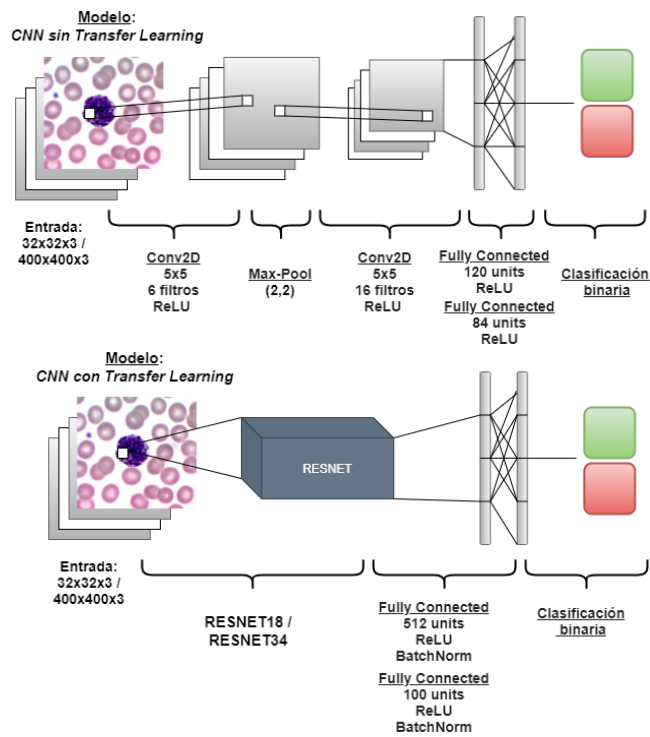


Fig. 1. Arquitectura del modelo. Arriba, modelo CNN sin Transfer Learning; abajo, modelo CNN basado en RESNET usando Transfer Learning

B. Dataset

Para el desarrollo de la investigación presentada se utilizó un dataset de 18,360 imágenes de cuerpos celulares presentes en la sangre periférica de aproximadamente 200 pacientes, este dataset es de carácter público y compartido por el Hospital de la Universidad de Munich en el repositorio de imágenes "Cancer Imaging Archive" [29]. Dentro del dataset están presentes 100 pacientes en condiciones saludables y 100 pacientes con signos anómalos detectados por especialistas de laboratorio.

El dataset se divide en 15 categorías de los diferentes tipos de células que pueden encontrarse dentro de la sangre periférica de los seres humanos. Estas muestras fueron obtenidas por punción venosa y analizada posteriormente con un microscopio de tipo M8 digital [29] mediante una técnica denominada conteo sanguíneo completo (CSC).

Para el desarrollo del modelo y la experimentación, se determinaron dentro de las 15 categorías, 2 subcategorías para las células típicas y las células atípicas encontradas generalmente en casos de leucemia [29]. En particular, las células no maduras (blastos) están relacionados con padecimientos inherentes a la leucemia. La clasificación y diversificación puede ser visualizada en la figura 2.

Tomando en cuenta lo anterior, es posible segmentar binariamente los subtipos de células y poder discriminar existencias anormales y normales dentro del grupo sanguíneo.

C. Entrenamiento y Pruebas

Para realizar el entrenamiento y pruebas del modelo desarrollado, el set de datos fue muestreado aleatoriamente conservando el 80% de las imágenes para realizar el entrenamiento y

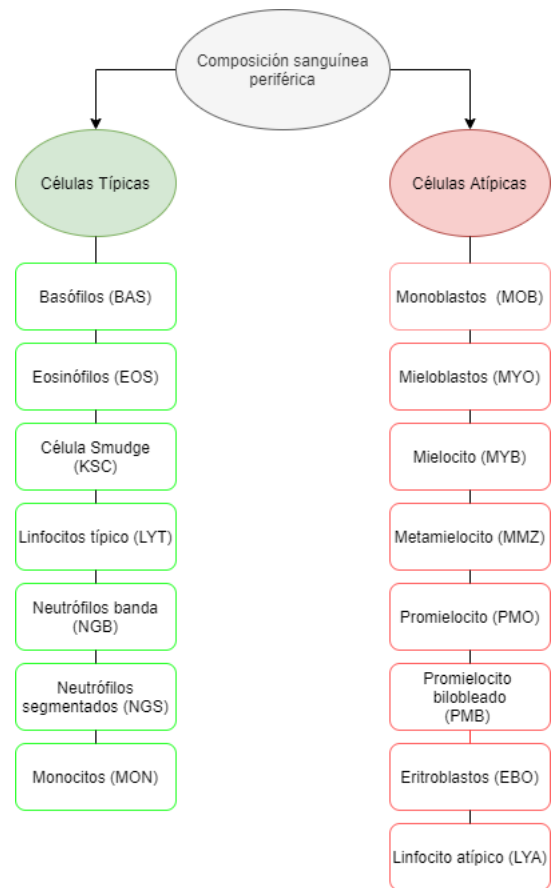


Fig. 2. Clasificación segmentada celular (Típica y Atípica)

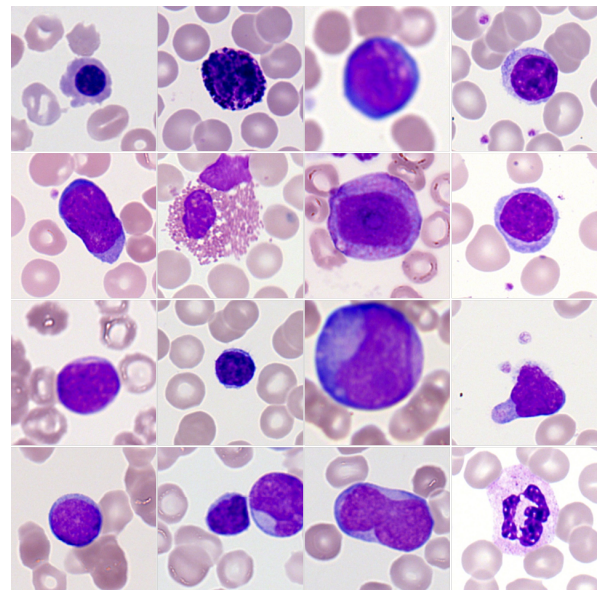


Fig. 3. Muestra de imágenes usadas para entrenamiento (primera a tercera fila) y prueba (cuarta fila). Las columnas uno y tres representan células atípicas, mientras que las columnas dos y cuatro representan células típicas.

el 20% restante para probar el modelo entrenado. Esta división del dataset fue inspirado en trabajos similares (80/20) [1].

Para evaluar la capacidad de abstracción de las CNNs, realizamos ajustes en el tamaño de las imágenes. Primero, muestreamos las imágenes del dataset a un tamaño de 32 x 32

TABLE II
EXPERIMENTOS PARA ENTRENAMIENTO Y TEST

Tamaño de imagen	Tamaño del lote	EPOCHS
32 x 32	4, 8, 32	45
400 x 400	8, 32	60

y, segundo, utilizamos la imagen de tamaño original de 400 x 400 píxeles. La figura 3 muestra ejemplos de las imágenes que integran la base de datos. El entrenamiento se realizó con diferentes tamaños de lotes (*minibatches*), y diferentes números de EPOCHS, como se muestra en la tabla II.

D. Dependencias y Herramientas Utilizadas

La red neuronal convolucional fue programada en Python 3.7.6 y Pytorch 1.4.0 como lenguaje de programación y framework, respectivamente, además también se utilizaron las librerías Matplotlib 3.1.3 y Numpy 1.18.1. Para realizar el entrenamiento de la red neuronal y las pruebas correspondientes para cada experimento, se empleó equipo de cómputo estándar con un procesador Ryzen 7 2700X, 16 GB de RAM, y una tarjeta de video NVIDIA RTX 2070 SUPER.

Para validar los resultados de la comparación binaria en los grupos creados a partir de las células consideradas normales y las anormales, se implementaron diversas técnicas estadísticas para medir su efectividad, entre ellas: efectividad general del modelo, efectividad por clase, sensibilidad, especificidad, falsos positivos, falsos negativos, verdaderos negativos, y verdaderos positivos.

V. MÉTRICAS DE EVALUACIÓN

Esta sección presenta los indicadores principales sobre los cuales se evaluó el desempeño del modelo y sus resultados respecto a otros trabajos presentados en el estado del arte con características similares.

A. Efectividad Porcentual de Discriminación Celular

Esta métrica permite evaluar de forma general la efectividad del modelo con los datos de prueba, y así calcular el porcentaje de aciertos. Para esto se utilizó la relación de $TotalAciertos/TotalDatos$. Para tener una mejor interpretación del modelo, también se evaluó la efectividad por clase.

B. Sensibilidad y Especificidad

Para determinar los tipos de errores, utilizamos los conceptos de sensibilidad y especificidad, utilizados comúnmente en investigaciones con aplicaciones médicas y en todos los trabajos estudiados en el estado del arte.

En este contexto, la sensibilidad (S) determina el porcentaje de casos con incidencias atípicas que fueron correctamente diagnosticadas, es decir, calcula la capacidad del modelo para detectar las anomalías del sujeto que efectivamente presenta un desorden atípico en la sangre periférica, mediante (1). En cambio, la especificidad (E) calcula la proporción de casos sanos o típicos que fueron clasificados por el modelo como típicos o dentro de los parámetros definidos como

normales para esta investigación, mediante (2). Este par de métricas permite evaluar el desempeño granular de los casos diagnosticados correctamente, tanto positivos como negativos.

$$S = \frac{VP}{VP+FN} \quad (1)$$

$$E = \frac{VN}{VN+FP} \quad (2)$$

C. Importancia de los Falsos Negativos

El manejo de los falsos negativos es determinante para este trabajo ya que representa un sector vulnerable de los casos analizados, debido al peligro que la herramienta representaría si la proporción de este tipo de errores es alta. Por esto, el modelo propuesto debe producir resultados que minimicen este indicador, aunque cabe recalcar que no se eliminarán en su totalidad, por lo que los resultados deben ser monitoreados en todo momento. En este contexto, definimos un falso negativo como: "un resultado de una prueba que indica que una persona no tiene cierta enfermedad o afección cuando en realidad sí la tiene" [31].

VI. EXPERIMENTACIÓN Y RESULTADOS

El proceso de experimentación permite evaluar el desempeño de nuestro modelo. En su versión sencilla con dos capas convolucionales y la versión basada en modelos pre-entrenados RESNET18 y RESNET34 utilizando *transfer learning*. Cabe destacar que la carga de los datos se realiza de manera aleatoria para cada modelo entrenado. Al final, se realiza una comparación de los resultados respecto a los trabajos descritos en el estado del arte para verificar el desempeño del modelo contra otros trabajos, así como entre las variantes de experimentos realizados para este trabajo.

Debido a la naturaleza del problema descrito y a las características de trabajos anteriormente revisados, para fines de esta investigación se establece como objetivo alcanzar una precisión general superior al 90% en efectividad total, sensibilidad y especificidad, tomando en cuenta proyectos con resultados similares [1], [4]. Finalmente, se espera que la proporción de falsos negativos se encuentre entre el 5-10% de los elementos analizados. Este porcentaje es el número estándar de falsos negativos en las pruebas de laboratorio en diversos estudios análisis [32], [33], sin embargo éstos pueden variar de un proyecto a otro.

La tabla III presenta los resultados de los experimentos para imágenes de 32 x 32 píxeles del primer modelo descrito. Puede observarse que los indicadores de desempeño se mantuvieron relativamente constantes, existiendo pequeñas fluctuaciones en el porcentaje de falsos negativos que no superan el 2.5% de los datos analizados para el set de prueba, manteniendo de esta forma un comportamiento consistente en los resultados obtenidos.

De manera similar, la Tabla IV presenta los resultados de los experimentos con imágenes de tamaño original (400 x 400 píxeles) que utiliza el mismo modelo descrito. Aquí, es posible observar que también se mantienen resultados consistentes similares a los presentados en la Tabla III. Puede observarse

TABLE III
RESULTADOS SOBRE INDICADORES DE DESEMPEÑO PARA
IMÁGENES 32X32

ID	Lote	EPOCHS	Precisión	Sensibilidad	Especificidad	%FN
1	4	45	95%	0.97	0.87	2.02%
2	8	45	95%	0.97	0.88	2.45%
3	34	45	95%	0.98	0.88	1.69%
4	8	60	95%	0.97	0.89	2.26%
5	34	60	95%	0.97	0.86	2.07%

que el porcentaje de falsos negativos es un poco mayor teniendo un máximo de 3.49% al reportado por las imágenes de 32 x 32 con un máximo de 2.45%. Este resultado es relevante ya que muestra la capacidad de abstracción de la CNN aún con imágenes con baja resolución, que en contraste con el desempeño de un experto humano, dicha disminución no va en detrimento del desempeño de la red. En ambos casos, el entrenamiento fue realizado utilizando *Stochastic Gradient Descent* (SGD) con *learning rate* (*lr*) de 0.001 y *momentum* de 0.9, donde *lr* fue seleccionado de una búsqueda de rangos entre 0.1 y 0.000001, el valor para *momentum* fue dejado en su valor predeterminado.

Finalmente, se utilizó *transfer learning* con modelos pre-entrenados. RESNET18 y RESNET34 fueron utilizados para diferentes tamaños de imagen, como se muestra en las tablas V y VI. Así, se puede observar una mejora significativa en ciertos casos para la efectividad general y una disminución dentro de los falsos negativos en comparación con el modelo anteriormente descrito, teniendo como mínimo 1.20% de falsos negativos para las imágenes de 32 x 32 y un 0.87% para las imágenes de tamaño original. Es importante notar, que en contraste con el modelo anterior, dado que ahora se utiliza un modelo con mayor capacidad de abstracción, la eficiencia en la discriminación de células típicas y atípicas es comparable para las imágenes muestreadas y las imágenes de tamaño original. La optimización de los parámetros para este modelo fue realizada utilizando ADAM con *lr* de 5×10^{-4} , obtenido de una búsqueda en el intervalo $[1 \times 10^{-5}, 1 \times 10^{-2}]$, y regularización tipo *weight decay* de 3×10^{-3} , obtenido de una búsqueda en el intervalo $[1 \times 10^{-5}, 1 \times 10^{-2}]$.

A. Indicadores Específicos de Desempeño por Clase

Respecto a los resultados presentados anteriormente sobre los indicadores generales desempeño, se toma el mejor modelo de cada grupo, tomando en cuenta la precisión general del modelo y el porcentaje total de error en dicho experimento y se presentan los resultados para cada clase, al igual que la proporción de aciertos y de errores en cada uno.

De forma similar a los resultados obtenidos previamente, los resultados específicos por clase para los modelos seleccionados se mantuvieron constantes en la detección de casos

TABLE IV
RESULTADOS SOBRE INDICADORES DE DESEMPEÑO PARA
IMÁGENES 400X400

ID	Lote	EPOCHS	Precisión	Sensibilidad	Especificidad	%FN
6	34	34	94%	0.98	0.80	1.28%
7	4	45	94%	0.97	0.85	2.07%
8	8	45	94%	0.97	0.86	2.64%
9	102	50	94%	0.96	0.87	3.49%

TABLE V
RESULTADOS SOBRE INDICADORES DE DESEMPEÑO PARA
IMAGENES 32X32 (RESNET)

ID	Modelo	EPOCHS	Precisión	Sensibilidad	Especificidad	%FN
10	Resnet18	30	96%	0.99	0.86	1.20%
11	Resnet18	50	96%	0.97	0.93	2.21%
12	Resnet34	30	96%	0.98	0.91	1.74%
13	Resnet34	50	96%	0.98	0.89	1.55%

TABLE VI
RESULTADOS SOBRE INDICADORES DE DESEMPEÑO PARA
IMÁGENES 400X400 (RESNET)

ID	Modelo	EPOCHS	Precisión	Sensibilidad	Especificidad	%FN
14	Resnet18	30	97%	0.98	0.93	1.33%
15	Resnet18	50	96%	0.97	0.93	2.04%
16	Resnet34	30	95%	0.99	0.83	0.87%
17	Resnet34	50	96%	0.98	0.93	1.82%

atípicos. Sin embargo, las mejoras para detectar a los pacientes efectivamente sanos en modelos que utilizaron *transfer learning* son más notorias en comparación con el primer modelo, esto es constante para ambos tamaños de imagen (5% de mejora para imágenes 32 x 32 y 8% para las de tamaño original), particularmente para la clase típica en los dos casos, como se muestra en la figura 4.

B. Indicadores Específicos de Desempeño de Clasificación en Aciertos y Errores

Respecto al desempeño de clasificación de los tipos de aciertos (Verdaderos Positivos y Verdaderos Negativos) y tipos de errores (Falsos Positivos y Falsos Negativos) realizados durante las predicciones, se mantuvo el patrón de los resultados específicos al igual que los resultados generales anteriormente presentados, como se muestra en la figura 5, y en la Tabla VII, donde se muestra la matriz de confusión para los mejores modelos con *transfer learning*. Se puede ver una clara disminución de los errores de manera general en los que utilizaron *transfer learning* y por lo tanto una mejor capacidad de predicción en los indicadores de precisión general y por clase. Además, también se puede observar que la proporción de error general (representada por la suma de falsos positivos y falsos negativos divididos entre los casos totales) es de apenas 3.37% (ID:12,32x32) y 2.66% (ID:14,400x400) para los modelos que utilizaron *transfer learning*. De igual forma, la tasa de aciertos corresponde a la efectividad general anteriormente mencionada del 96 % (ID:12,32x32) y 97% (ID:14,400x400), demostrando que *transfer learning* es una técnica que impacta positivamente este proceso de clasificación.

C. Comparación del Desempeño contra Trabajos en Aplicaciones Médicas

Esta sección presenta una comparación entre los mejores modelos obtenidos como resultado de esta investigación con

TABLE VII
MATRIZ DE CONFUSIÓN PARA MODELOS CON *Transfer Learning*

	ID 12		ID 14	
	Predicción		Predicción	
	1	0	1	0
Real	1	2915	64	49
	0	60	633	640

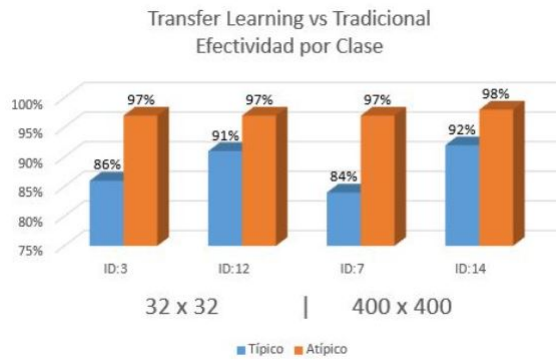


Fig. 4. Resultados de la efectividad por clase y tamaño de imagen en comparación entre el modelo tradicional y *transfer learning*

respecto a los indicadores de desempeño de otros trabajos presentados en el estado del arte. Es importante mencionar que se reportan únicamente los trabajos que utilizan las mismas métricas de evaluación para realizar una comparación adecuada evitando aquellos que no reportaron todos los resultados necesarios para su comparación. Además, debe hacerse notar que esta comparación es sólo para fines cualitativos, ya que los trabajos contra los que se está cotejando son aplicados en dominios diferentes de cáncer con bases de datos con características distintas a los datos utilizados en este trabajo. No obstante, permiten demostrar que nuestro modelo arroja resultados con porcentajes de efectividad y error comparables a los obtenidos por modelos en el estado del arte en aplicaciones médicas relacionadas con cáncer.

La Tabla VIII presenta los resultados de los indicadores de desempeño de diagnóstico de cada trabajo. Como puede observarse, los resultados presentados por la actual investigación (Discriminador celular automático, ID:14) demuestran que el modelo utilizado fue efectivo en la detección de incidencias anómalas cuando realmente estaban presentes en el individuo, indicando menor incidencia de falsos negativos. La especificidad demuestra que la proporción de individuos diagnosticados con incidencias típicas correctamente identificados fue 1% menor que el mejor de los trabajos en el estado del arte, indicando un porcentaje mayor de falsos positivos, sin embargo, los resultados muestran el mismo número de falsos negativos como de falsos positivos. A pesar de ello, el

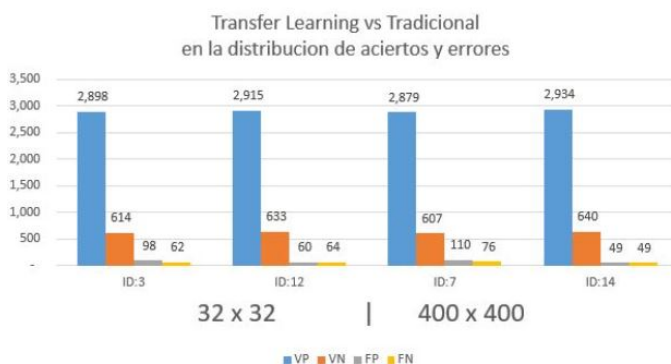


Fig. 5. Distribución de aciertos y errores en los modelos seleccionados tradicionales y *transfer learning*

TABLE VIII

TABLA COMPARATIVA DE LOS INDICADORES DE DESEMPEÑO

Trabajo	Precisión	Sensibilidad	Especificidad
Skin Cancer Detector	91%	0.95	0.92
Carcinoma en Tejidos Renales	91-92 %	0.88	0.94
Discriminador celular automático	97%	0.98	0.93

porcentaje de errores por falsos positivos no supera el 1.5% .

Finalmente, puede observarse que el trabajo actual superó el objetivo propuesto y fue competitivo respecto a trabajos similares en sus respectivos sets de datos, cumpliendo el objetivo por medio de los indicadores de desempeño propuestos para esta investigación.

VII. CONCLUSIONES

Realizar diagnóstico médico oportuno puede resultar una tarea difícil debido a la naturaleza multi-factorial que se desprende del ambiente de trabajo, conocimiento, área de experiencia, carga de trabajo, así como la falta de especialistas capacitados. Por esto, la problemática identificada específicamente para la detección de la leucemia gira entorno a que no existe una herramienta con capacidades de clasificación citomorfológica para realizar una discriminación binaria y determinar una afección celular.

Fue a partir de estos problemas que se planteó como una posible solución el desarrollo de un modelo de redes neuronales convolucionales para poder clasificar y brindar la información relevante anteriormente mencionada. Para conseguirlo, se desarrolló un modelo con imágenes ajustadas a un tamaño de 32x32 y 400x400 píxeles capaz de presentar indicadores de su desempeño al realizar detecciones tanto típicas como atípicas.

Después de presentar los resultados, se observa que el modelo resultó especialmente efectivo al realizar clasificaciones de la constitución del grupo sanguíneo periférico en los seres humanos, teniendo una efectividad por encima del 90% y un error general de apenas el 2.66% (ID:14). Además, la efectividad alcanzada en la clasificación de individuos ciertamente enfermos presentó disminución en las incidencias de falsos negativos a aproximadamente el 1.33%, haciéndolo una herramienta eficaz al realizar un diagnóstico respecto a otras herramientas presentadas con una proporción de error superior tanto en lo general como en lo específico.

REFERENCES

- [1] L. Zhang, H. Gao, J. Zhang and B. Badami, "Optimization of the Convolutional Neural Networks for Automatic Detection of Skin Cancer", *Open Medicine*, vol. 15, no. 1, pp. 27-37, 2020. Available: 10.1515/med-2020-0006.
- [2] T. Zhang et al., "Cervical precancerous lesions classification using pre-trained densely connected convolutional networks with colposcopy images", *Biomedical Signal Processing and Control*, vol. 55, p. 101566, 2020. Available: 10.1016/j.bspc.2019.101566.
- [3] S. Yadav and S. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis", *Journal of Big Data*, vol. 6, no.1, 2019. Available: 10.1186/s40537-019-0276-2
- [4] S. Tabibu, P. Vinod and C. Jawahar, "Pan-Renal Cell Carcinoma classification and survival prediction from histopathology images using deep learning", *Scientific Reports*, vol. 9, no. 1, 2019. Available: 10.1038/s41598-019-46718-3.
- [5] "Lunit Inc.", Lunit.io, 2020. [Online]. Available: <https://www.lunit.io/en/>. [Accessed: 27- Mar- 2020].

- [6] A. Suarez Leon and J. Nunez Alvarez, "1D Convolutional Neural Network for Detecting Ventricular Heartbeats", *IEEE Latin America Transactions*, vol. 17, no. 12, pp. 1970-1977, 2019. Available: <https://latam.ieceer9.org/index.php/transactions/article/view/2850/351>. [Accessed 5 September 2020].
- [7] L. Evangelista and E. Guedes, "Ensembles of Convolutional Neural Networks on Computer-Aided Pulmonary Tuberculosis Detection", *IEEE Latin America Transactions*, vol. 17, no. 12, pp. 1954-1963, 2019. Available: <https://latam.ieceer9.org/index.php/transactions/article/view/2813/349>. [Accessed 5 September 2020].
- [8] de Oliveira Lima, L. Silva de Araujo Filho, F. Santos da Silva and C. Serodio Figueiredo, "Pigmented Dermatological Lesions Classification Using Convolutional Neural Networks Ensemble Mediated by Multilayer Perceptron Network", *IEEE Latin America Transactions*, vol. 17, no. 11, pp. 1902-1908, 2019. Available: <https://latam.ieceer9.org/index.php/transactions/article/view/1948/341>. [Accessed 5 September 2020].
- [9] "Project InnerEye - Medical Imaging AI to Empower Clinicians - Microsoft Research", Microsoft Research, 2020. [Online]. Available: <https://www.microsoft.com/en-us/research/project/medical-image-analysis/>. [Accessed: 27- Mar- 2020].
- [10] R. Pedro da Silva Neto and A. Oseas de Carvalho Filho, "Automatic classification of breast lesions using Transfer Learning", *IEEE Latin America Transactions*, vol. 17, no. 12, pp. 1964-1969, 2019. Available: <https://latam.ieceer9.org/index.php/transactions/article/view/2837/350>. [Accessed 5 September 2020].
- [11] "What Is Cancer?", National Cancer Institute, 2020. [Online]. Available: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>. [Accessed: 18- Feb- 2020].
- [12] "¿Qué es el cáncer?", Cancer.org, 2020. [Online]. Available: <https://www.cancer.org/es/cancer/aspectos-basicos-sobre-el-cancer/que-es-el-cancer.html>. [Accessed: 18- Feb- 2020].
- [13] "Leucemia: MedlinePlus enciclopedia médica", Medlineplus.gov, 2020. [Online]. Available: <https://medlineplus.gov/spanish/ency/article/001299.htm>. [Accessed: 18- Feb- 2020].
- [14] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification", *Neurocomputing*, vol. 321, pp. 321-331, 2018. Available: 10.1016/j.neucom.2018.09.013.
- [15] Y. Zhou et al., "A Radiomics Approach With CNN for Shear-Wave Elastography Breast Tumor Classification", *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1935-1942, 2018. Available: 10.1109/tbme.2018.2844188.
- [16] Bernal et al., "Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review", *Artificial Intelligence in Medicine*, vol. 95, pp. 64-81, 2019. Available: 10.1016/j.artmed.2018.08.008 [Accessed 19 May 2020].
- [17] D. Zhang et al., "Panoptic Segmentation with an End-to-End Cell R-CNN for Pathology Image Analysis", *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018*, pp. 237-244, 2018. Available: 10.1007/978-3-030-00934-2_27 [Accessed 19 May 2020].
- [18] L. Zhang, F. Yin and J. Cai, "A Multi-Source Adaptive MR Image Fusion Technique for MR-Guided Radiation Therapy", *International Journal of Radiation Oncology*Biophysics*Physics*, vol. 102, no. 3, p. e552, 2018. Available: 10.1016/j.ijrobp.2018.07.1537.
- [19] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm", *Advances in Engineering Software*, vol. 95, pp. 51-67, 2016. Available: 10.1016/j.advengsoft.2016.01.008 [Accessed 19 May 2020].
- [20] Kermany D, Goldbaum M. Labeled optical coherence tomography (oct) and chest X-ray images for classification. In: Mendeley data. 2018. p. 2.
- [21] Y. Cao, T. A. Geddes, J. Y. H. Yang, and P. Yang, "Ensemble deep learning in bioinformatics," *Nature Machine Intelligence*, vol. 2, no. 9, Art. no. 9, Sep. 2020, doi: 10.1038/s42256-020-0217-y.
- [22] M. Lopez-Martin, A. Nevado, and B. Carro, "Detection of early stages of Alzheimer's disease based on MEG activity with a randomized convolutional neural network," *Artificial Intelligence in Medicine*, vol. 107, p. 101924, Jul. 2020, doi: 10.1016/j.artmed.2020.101924.
- [23] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Classification of histopathological biopsy images using ensemble of deep learning networks," in *Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering*, USA, Nov. 2019, pp. 92-99, Accessed: Jan. 07, 2021. [Online].
- [24] C. Matek, S. Schwarz, K. Spiekermann and C. Marr, "Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks", *Nature Machine Intelligence*, vol. 1, no. 11, pp. 538-544, 2019. Available: 10.1038/s42256-019-0101-9.
- [25] De Fauw et al., "Clinically applicable deep learning for diagnosis and referral in retinal disease", *Nature Medicine*, vol. 24, no. 9, pp. 1342-1350, 2018. Available: 10.1038/s41591-018-0107-6.
- [26] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998. Available: 10.1109/5.726791.
- [27] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines.," in *ICML*, 2010, pp. 807-814.
- [28] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," arXiv:1409.0575 [cs], Jan. 2015. Accessed: Jan. 07, 2021. [Online]. Available: <http://arxiv.org/abs/1409.0575>.
- [29] Matek, C., Schwarz, S., Marr, C., & Spiekermann, K. (2019). A Single-cell Morphological Dataset of Leukocytes from AML Patients and Non-malignant Controls [Data set]. The Cancer Imaging Archive. <https://doi.org/10.7937/tcia.2019.36f5o9ld>
- [30] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps," arXiv:1312.6034 [cs], Apr. 2014, Accessed: Jan. 07, 2021. [Online]. Available: <http://arxiv.org/abs/1312.6034>.
- [31] "Importancia del cálculo de la sensibilidad, la especificidad y otros parámetros estadísticos en el uso de las pruebas de diagnóstico clínico y de laboratorio", *Docs.bvsalud.org*, 2020. [Online]. Available: <http://docs.bvsalud.org/biblioref/2018/05/883697/importancia-calculo-sensibilidad-y-especificidad.pdf>. [Accessed: 15- Apr- 2020].
- [32] M. Sánchez Nava, A. Olivares Montano, N. Contreras Carreto and M. Díaz Suárez, "Certeza diagnóstica de la colposcopia, citología e histología de las lesiones intraepiteliales del cérvix", *Rev Invest Med Sur Mex*, vol. 20, no. 2, pp. 95-99, 2020. Available: <https://www.medigraphic.com/pdfs/medsur/ms-2013/ms132b.pdf>. [Accessed 19 May 2020].
- [33] E. Horvath, M. Galleguillos and V. Schonstedt, "¿Existen cánceres no detectables en la mamografía?", *Revista Chilena de Radiología*, vol. 13, no. 2, pp. 84-89, 2020. Available: <https://scielo.conicyt.cl/pdf/rchradiol/v13n2/art07old.pdf>. [Accessed 15 April 2020].



operation in production. claudio.soto.ayala@gmail.com

Luis Claudio Soto-Ayala received the B.Sc. degree in Computer Science from the Tecnológico de Monterrey, Querétaro, Mexico, in 2020. His degree thesis focused on the classification of blood cells' anomalies using Deep Learning, in particular Convolutional Neural Networks. Currently, he is working as an engineer and software developer in the financial group Santander, Mexico, Querétaro, in the CSA department, where he develops applications and web solutions for the optimization and automation of critical tasks, as well as the validation of their correct



Jose A. Cantoral-Ceballos (Ph.D., M.Sc., MIEEE) received the B.Sc. (Excellence Hons.) degree in electronic and communications engineering from the Tecnológico de Monterrey, Queretaro, Mexico, in 2005, and the M.Sc. (Distinction) and Ph.D. degrees from the University of Manchester, Manchester, U.K., in 2009 and 2013, respectively, with a focus on algorithms for tomography imaging in embedded systems. During 2013-2014 he was a Post-Doctoral Research Associate at the University of Manchester where he worked on the design and development of

a tomography data acquisition system, as well as imaging algorithms from limited views. From 2015 to 2020 he was at the Advanced Technology Center (CIATEQ A.C.), Mexico, where his research efforts focused on the study of novel Deep Learning methods to solve Digital Signal Processing and Energy problems. Currently, he is a Researcher Professor at the Tecnológico de Monterrey, Queretaro, in the department of Computer Science, where he continues his research on Deep Learning solutions to different problems, in particular to the study of neurological signals. Dr. Cantoral-Ceballos was the recipient of the National Instruments Prize for his M.Sc. dissertation in 2009 and the Best Technical Presentation Award at the Seventh World Congress of Industrial Process Tomography (WCIPT), Krakow, in 2013. joseantonio.cantoral@tec.mx