

Modelo de clasificación de señales de electroretinograma para diagnóstico de retinopatía diabética temprana

Ing. Carlos Oliver  
Hernández Montejano

2022



# Universidad Autónoma de Querétaro

## Facultad de Ingeniería

Modelo de clasificación de señales de electroretinograma para diagnóstico de retinopatía diabética temprana

### Tesis

Que como parte de los requisitos para obtener el grado de

### Maestro en Ciencias en Inteligencia Artificial

Presenta:

Ing. Carlos Oliver Hernández Montejano

Dirigido por:

Dr. Saúl Tovar Arriaga

Co-Dirigido por:

Dra. Stéphanie Colette Thébault

Querétaro, Qro. a Septiembre 2022



Dirección General de Bibliotecas y Servicios Digitales  
de Información



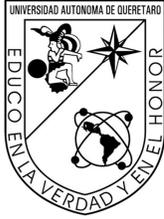
Modelo de clasificación de señales de  
electrorretinograma para diagnóstico de retinopatía  
diabética temprana

**por**

Carlos Oliver Hernández Montejano

se distribuye bajo una [Licencia Creative Commons  
Atribución-NoComercial-SinDerivadas 4.0  
Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/).

**Clave RI:** IGMAC-300613



Universidad Autónoma de Querétaro

Facultad de Ingeniería

Maestría en Ciencias en Inteligencia Artificial

**Modelo de clasificación de señales de electroretinograma para diagnóstico de retinopatía diabética temprana**

Tesis

Que como parte de los requisitos para obtener el grado de Maestro en Ciencias en Inteligencia Artificial

Presenta:

**Ing. Carlos Oliver Hernández Montejano**

Dirigido por:

**Dr. Saúl Tovar Arriaga**

Co-Dirigido por:

**Dra. Stéphanie Colette Thébault**

Dr. Saúl Tovar Arriaga

Presidente

Dra. Stéphanie Colette Thébault

Secretario

Dr. Marco Antonio Aceves Fernández

Vocal

Dr. Sebastián Salazar Colores

Suplente

Dra. Danjela Ibrahim

Suplente

Centro Universitario Querétaro, QRO

Septiembre 2022

México

## **1. Dedicatoria**

A mi familia, mi madre, mi padre y mi hermana por el apoyo, por guiarme y por ser las bases que me ayudaron a llegar hasta aquí.

A Fatima, por su amor, creencia e incondicional apoyo hacia mi persona.

A todos los profesores del programa de Maestría en Ciencias en Inteligencia Artificial por transmitirnos el conocimiento y las ganas de seguir superándonos.

## 2. Agradecimientos

En primer lugar a mi familia, por estar conmigo en cada paso y faceta de mi vida y por proporcionarme las herramientas necesarias para poder llegar aquí y lograr mis objetivos.

A Fatima, por su paciencia, creencia, apoyo y cariño demostrado. Por ser mis ganas de levantarme cada mañana para convertirme en mejor persona y superarme a mi mismo día con día.

Agradezco a mi asesor Dr. Saúl Tovar por su tiempo, esfuerzo y paciencia dedicados a este proyecto y acompañarme a lo largo de esta experiencia.

A la Dra. Stephanie Thébault por darme la confianza de contribuir en este trabajo, por su gran experiencia y consejos que fueron la clave para la realización de este proyecto.

Al Dr. Sebastián Salazar por su constante seguimiento y apoyo técnico, por la contribución de sus grandes ideas que me ayudaron a aterrizar este trabajo.

A todos los profesores del programa de maestría, por sus enseñanzas y compromiso de formar futuros investigadores.

Agradezco al Consejo Nacional de Ciencia y Tecnología por proveer los fondos utilizados para este proyecto.

# Índice

<b>1. Dedicatoria</b>	<b>3</b>
<b>2. Agradecimientos</b>	<b>4</b>
<b>3. Abreviaturas y siglas</b>	<b>10</b>
<b>4. Resumen</b>	<b>13</b>
<b>5. Abstract</b>	<b>14</b>
<b>6. Introducción</b>	<b>15</b>
6.1. Planteamiento del problema . . . . .	16
<b>7. Justificación</b>	<b>17</b>
<b>8. Antecedentes</b>	<b>18</b>
8.1. Inteligencia Artificial . . . . .	18
8.2. Machine Learning y Deep Learning . . . . .	18
8.3. Redes Neuronales Convolucionales . . . . .	19
8.3.1. Parámetros de las CNN . . . . .	22
8.3.2. AlexNet . . . . .	25
8.3.3. HyperResNet . . . . .	26
8.3.4. CNN Paralela . . . . .	29
8.3.5. Métricas . . . . .	30
8.3.6. Retinopatía Diabética . . . . .	34
8.3.7. Electrorretinograma . . . . .	39
8.3.8. Transformada Wavelet . . . . .	46
8.3.9. Estado del arte . . . . .	50
<b>9. Hipótesis</b>	<b>54</b>

<b>10. Objetivos</b>	<b>55</b>
10.1. Objetivo general . . . . .	55
10.2. Objetivos específicos . . . . .	55
<b>11. Metodología</b>	<b>56</b>
11.1. Base de datos . . . . .	56
11.2. Escalogramas . . . . .	61
11.3. Modelos . . . . .	67
11.3.1. AlexNet Modificada . . . . .	67
11.3.2. HyperResNet AutoML . . . . .	68
11.3.3. CNN Paralela . . . . .	70
<b>12. Resultados</b>	<b>74</b>
<b>13. Conclusiones</b>	<b>80</b>
13.0.1. Trabajo futuro . . . . .	83
<b>14. Anexos</b>	<b>94</b>
14.1. Anexo 1 Carta de autorización para uso de base de datos . . . . .	94
14.2. Anexo 2 Registro de pruebas de modelos . . . . .	95
14.3. Anexo 3 Dictamen de aprobación de artículo en revista Mecamex . . . . .	96

## Índice de tablas

1.	Definición de tipos de Positivos y Negativos. . . . .	31
2.	Escala Internacional de Severidad de la Retinopatía Diabética (ETDRS) [44]. . . . .	37
3.	Estado del arte. . . . .	50
4.	Categorías originales señales ffERG espontáneas. . . . .	56
5.	Nuevas categorías señales ffERG espontáneas. . . . .	57
6.	Estructura de dataset final de señales ffERG espontáneas. . . . .	61
7.	Comparativa de métricas de los modelos implementados. . . . .	76

## Índice de figuras

1.	Convolución de una imagen aplicando kernel y función de activación RELU [21]. . . . .	21
2.	Ejemplo de imagen de entrada y kernel [21]. . . . .	22
3.	Ejemplo de <i>subsampling Max-pooling 2x2</i> [21]. . . . .	23
4.	Ejemplo de convolución <i>stride=1</i> y <i>stride=2</i> respectivamente [22]. . . . .	24
5.	Topología original AlexNet [23]. . . . .	26
6.	Topología original ResNet [25]. . . . .	28
7.	Ejemplo de CNN Paralela para el procesamiento de imágenes [30] . . . . .	30
8.	Estructura de una matriz de confusión. . . . .	32
9.	Ejemplos de Curvas ROC, diagnóstico perfecto, diagnóstico bueno y sin valor diagnóstico respectivamente. . . . .	35
10.	Escala de gravedad de RD propuesta por Sun y col. en 2020 [44]. . . . .	39
11.	Ejemplo típico de ERG humano [38]. . . . .	40
12.	Diferencias en ERG evocado de campo completo en persona sana y personas con diabetes sin retinopatía aparente [63]. . . . .	45
13.	A) Montaje fotográfico de fondo de ojo. B) Mapa de tiempos implícitos de ERG multifocal evocado [54]. . . . .	45
14.	Tres wavelets de Morlet en el dominio del tiempo (fila superior) y en el dominio de la frecuencia (fila inferior) con dos frecuencias y diferentes parámetros de compensación de tiempo-frecuencia [68]. . . . .	49
15.	Histograma de tiempos de duración de señales ffERG espontáneas en segundos. . . . .	58
16.	Señales ffERG espontáneas omitidas para el estudio. . . . .	59
17.	Muestras de señales ffERG espontáneas originales y normalizadas respectivamente . . . . .	60
18.	Muestras de señales ffERG espontáneas y escalogramas respectivos obtenidos con la librería Librosa. . . . .	63

19. Muestras de señales ffERG espontáneas y escalogramas respectivos obtenidos con CWT. . . . .	65
20. Extracción de escalogramas de secciones de la señal más grande de acuerdo a la ventana de tiempo de la señal más pequeña. . . . .	67
21. Extracción de escalograma de la señal más pequeña. . . . .	67
22. Modelo propuesto CNN AlexNet modificado . . . . .	68
23. Modelo propuesto HyperResNet obtenido a través de la herramienta KerasTuner. . . . .	69
24. Representación de entrada de los escalogramas de la señal más larga a la CNN Paralela. . . . .	71
25. Representación de entrada del escalograma de la señal más pequeña a la CNN Paralela. . . . .	71
26. Diagrama de metodología. . . . .	72
27. Gráficas de entrenamiento del modelo AlexNet, exactitud y pérdida respectivamente. . . . .	74
28. Gráficas de entrenamiento del modelo HyperResNet, exactitud y pérdida respectivamente. . . . .	75
29. Gráficas de entrenamiento del modelo CNN Paralela, exactitud y pérdida respectivamente. . . . .	76
30. Comparativa de matrices de confusión de los modelos implementados.	76
31. Curvas de análisis ROC modelo Alexnet. . . . .	77
32. Curvas de análisis ROC CNN Paralela. . . . .	78
33. Curvas de análisis ROC HyperResNet . . . . .	78

### 3. Abreviaturas y siglas

- **IA:** Inteligencia Artificial
- **ERG:** Electrorretinograma
- **ffERG:** Full-Field Electroretinogram (Electrorretinograma de campo completo)
- **mfERG :** Multifocal Electroretinogram (Electrorretinograma multifocal)
- **PERG :** Pattern Plectroretinogram (Electrorretinograma de patrón)
- **TI :** Tiempos Implícitos
- **ML:** Machine Learning (Aprendizaje automático)
- **DL:** Deep Learning (Aprendizaje profundo)
- **ANN:** Artificial Neural Network (Red Neuronal Artificial)
- **CNN:** Convolutional Neural Network (Red Neuronal Convolutacional)
- **ALM:** Algoritmo Levenberg-Marquardt
- **AQV:** Aprendizaje de Quantificación Vectorial
- **FBR:** Función de Base Radial
- **MAE:** Máquina de Aprendizaje Extremo
- **MRL:** Modelo de Regresión Logística
- **ANN-PSO:** Optimización de enjambre de partículas basada en redes neuronales artificiales
- **DM:** Diabetes Mellitus
- **OMS:** Organización Mundial de la Salud

- **DM1:** DM Tipo 1
- **DM2:** DM Tipo 2
- **ETDRS:** Early Treatment Diabetic Retinopathy Study (Estudio de tratamiento temprano de la retinopatía diabética)
- **RD:** Retinopatía Diabética
- **RDNP:** Retinopatía Diabética No Proliferativa
- **RDP:** Retinopatía Diabética Proliferativa
- **SIN RD:** Sin Retinopatía Diabética
- **WFT** Wavelet Fourier Transform (Transformada Wavelet de Fourier)
- **CWT:** Continuous Wavelet Transform (Transformada Continua Wavelet)
- **AutoML:** Auto Machine Learning (Auto Aprendizaje Automático)
- **AlexNet:** CNN AlexNet
- **HyperResNet:** AutoML CNN Residual Network (Red Neuronal Convolutacional Residual AutoML)
- **Keras-tuner:** Framework para búsqueda y optimización de hiperparámetros.
- **VC:** Visión por computadora
- **LRN:** Local Response Normalization (Normalización de respuesta local)
- **FC:** Fully Connected (Completamente conectada)
- **SGD:** Stochastic Gradient Descent (Descenso de Gradiente estocástico)
- **KNN:** K-Nearest Neighbour (Clasificador de Vecinos más cercanos)
- **VP:** Verdadero Positivo

- **FP:** Falso Positivo
- **VN:** Verdadero Negativo
- **FN:** Falso Negativo.
- **TPC:** Total de predicciones correctas
- **TP:** Total de predicciones
- **E:** Exactitud
- **P:** Precisión
- **S:** Sensibilidad
- **F1:** F1-Score
- **ROC:** Receiver Operating Characteristic (Característica Operativa del Receptor)
- **AUC:** Área Bajo la Curva

## 4. Resumen

La RD es una enfermedad causada por la diabetes que en la actualidad se diagnóstica principalmente a través del estudio de imágenes de fondo de ojo y análisis clínicos, sin embargo estas técnicas de diagnóstico categorizan la gravedad de la RD con base en los daños que existen en la retina. Se ha demostrado que técnicas como las señales de ERG permiten la identificación de la presencia de RD incluso antes de que los daños en la retina sean visibles. El estudio en los patrones de estas señales de manera espontánea puede ser la clave para la detección temprana de la RD y detener su progreso antes de que los daños en la retina sean graves. En este trabajo se implementó la comparativa de tres modelos de CNN para la clasificación de imágenes de escalograma obtenidos a partir de señales de ffERG espontáneos en dos categorías RD y NoRD. Los modelos implementados fueron los siguientes: Modificación a CNN Alexnet, ResNet implementada con AutoML (HyperResNet) y modelo de CNN paralelo a nueve capas. Las topologías presentaron exactitudes de 86.7 %, 79 % y 78 % respectivamente en la identificación de lo que es RD y NoRD.

**Palabras clave—, Retinopatía diabética, diagnóstico automatizado, redes neuronales convolucionales, señales de electrorretinograma, electrorretinograma espontáneo, escalogramas.**

## 5. Abstract

DR is a disease caused by diabetes that is currently diagnosed mainly through the study of fundus images and clinical tests; however, these diagnostic techniques categorize the severity of DR based on the damage that exists in the retina. Techniques such as ERG signals have been shown to allow identification of the presence of DR even before retinal damage is visible. Studying the patterns of these signals spontaneously may be the key to early detection of DR and stopping its progress before retinal damage becomes serious. In this work, the comparison of three CNN models for the classification of scalogram images obtained from spontaneous ffERG signals in two categories RD and NoRD was implemented. The implemented models were the following: Modification to CNN Alexnet, ResNet implemented with AutoML (HyperResNet) and nine-layer parallel CNN model. The topologies presented accuracy of 86.7 %, 79 % and 78 % respectively in the identification of what is RD and NoRD.

**Keywords—, Diabetic retinopathy, automatic diagnosis, convolutional neural networks, electroretinogram signals, spontaneous electroretinogram, scalograms.**

## 6. Introducción

La Inteligencia Artificial ha traído beneficios sin precedentes a la sociedad, por lo que es de suma importancia investigar de qué manera lograr una maximización de estos beneficios y explotar la materia al grado de evitar peligros potenciales [1].

Actualmente nos encontramos en una era dónde la Medicina Personalizada y la Medicina de Precisión tienen como objetivo mejorar la calidad asistencial e incrementar la sostenibilidad de los sistemas públicos de salud. Todo parece indicar que la IA por su potencial predictivo y clasificador en entornos del mundo real, tendrá un papel principal en todos los pasos del proceso asistencial: prevención, diagnóstico, tratamiento, control y seguimiento, así también como implantar la cultura de la salud [2].

Los beneficios potenciales de la IA en la medicina han sido medulares en la última década del siglo XX y lo serán aún más en el presente siglo. Es notable el impacto en los equipos médicos y del software especializado en el desarrollo de la IA, sobre todo en áreas como la Radiología y las Neurociencias. La IA ha sido vital para la creación de software que permiten la modelación y simulación de procesos biológicos fundamentales, como la respiración, infartos, actividad eléctrica cardiaca, conductividad del cuerpo humano, visión, entre otros [3].

En el desarrollo de este proyecto de investigación se propone la utilización de técnicas de IA para hacer frente a una de las consecuencias causada por una de las enfermedades más comunes en la actualidad, la Diabetes Mellitus. Los daños a la visión causados por la diabetes es un problema que enfrentan las personas que la padecen, por esta razón y a lo largo de este proyecto se profundiza en el estudio de técnicas electrofisiológicas que en conjunto con modelos de IA puedan ser usados para la detección temprana de RD.

## 6.1. Planteamiento del problema

La Retinopatía Diabética (RD) es una enfermedad común entre los pacientes con Diabetes. Aparece durante los primeros años y evoluciona mientras progresa la enfermedad. La RD causa daños severos a la visión y no es reversible, pero si es posible prevenirla y detener su progreso a través de un estricto control glucémico [4] y la fotocoagulación con láser [5], respectivamente. Desde un enfoque médico, la situación a la que se enfrenta la oftalmología contra la RD es que no existen terapias o tratamientos para atacar directamente la causa de la enfermedad, esto se debe principalmente a la ausencia de herramientas capaces de valorar la integridad de la funcionalidad retiniana.

Al momento de un diagnóstico el médico siempre se basa en la presencia de signos o síntomas a nivel clínico y paciente. Los signos que aparecen en las imágenes de fondo de ojo son los que un médico utiliza para evaluar la gravedad de la RD. La aparición de estos signos ocurre en etapas avanzadas de la enfermedad por tal razón se llega a que el 10 % de los pacientes con diabetes padece limitación visual y hasta el 2 % ceguera total. Estos porcentajes comparado de acuerdo al número de personas con diabetes en el mundo el número es exorbitante.

En México, el acceso a los aparatos que miden la actividad eléctrica retiniana es limitado por lo cual el estudio con ERG no se utiliza de manera frecuente. Además un estudio de ERG lleva tiempo ya que requiere diversas actividades previas como la dilatación de pupila, adaptación a la oscuridad, toma de registros, contacto de los electrodos, entre otros. La interpretación y análisis de los datos requiere personal altamente capacitado no solo en medicina, sino también en oftalmología e incluso electrofisiología.

## 7. Justificación

Al detectar de forma temprana la RD, en un estadio que actualmente se considera como subclínico, el porvenir del paciente mejorará significativamente. Entre más temprana sea la detección, más aplican las estrategias preventivas y al disponer de un método para detección temprana, se podrán desarrollar nuevas estrategias terapéuticas para estadios tempranos de la enfermedad.

Hoy en día existen diversos métodos para detectar RD, el más común es el tratamiento de imágenes de fondo de ojo, procedimiento para el cual ya existen algunos sistemas avanzados y métodos que son capaces de detectar enfermedades de la retina [6–13]. Así mismo se encuentra el ERG, el cual es capaz de detectar la actividad eléctrica de las células en la retina y permite el diagnóstico de diversas enfermedades oculares [14], o neurodegenerativas [15].

El uso del ERG en diagnóstico automatizado es escaso debido a la complejidad de su estudio, además se ha utilizado de manera evocada manteniendo la dependencia del equipo médico adecuado y protocolos largos para la toma de muestras. Los trabajos descritos en el estado del arte comienzan a apuntar su uso en modelos inteligentes para la detección de zonas retinianas con daños causados por la RD.

Contar con un sistema automatizado de clasificación de señales ERG espontáneo utilizando técnicas de inteligencia artificial, permitiría detectar de forma temprana la RD, ofreciendo oportunidades de diagnóstico cuantitativo, fácil de usar, rápido, de bajo costo [6] y nuevas estrategias terapéuticas para estadios tempranos de la enfermedad. En comparación con el estado del arte, donde los resultados son solamente clasificación entre “sanos” y “enfermos”, la principal ventaja lograr un diagnóstico más específico y detectarlo en etapas mínimas o incluso antes de que existan daños visibles.

## **8. Antecedentes**

### **8.1. Inteligencia Artificial**

La Inteligencia Artificial (IA) es una rama de la informática dedicada al desarrollo de sistemas inteligentes. Su finalidad es emular la inteligencia humana mediante algoritmos computacionales, que otorgan a un sistema la capacidad de aprender y mejorar el análisis de información.

En la época actual la IA ha tenido un impacto impresionante en el día a día, casi cualquier dispositivo electrónico expone este tipo de tecnología a la palma de la mano, computadoras, teléfonos, relojes, asistentes, sistemas de luces, sistemas de seguridad, sistemas de salud, automóviles y la lista es amplia. Sistemas capaces de procesar cantidades tan grandes de datos que hace unos años la humanidad solo lo vería en ciencia ficción, hoy en día son posibles gracias a las técnicas de IA como ML y DL.

### **8.2. Machine Learning y Deep Learning**

El aprendizaje automático, conocido comúnmente como Machine Learning (ML), es un subconjunto de la IA cuyo objetivo es el entrenamiento de máquinas a partir de datos ya conocidos. Una vez que la información es procesada, la máquina es capaz de tomar decisiones o realizar predicciones [16].

Dentro del subconjunto de ML, se encuentra el aprendizaje profundo o Deep Learning (DL), el cual es un conjunto de algoritmos y modelos computacionales que imita la arquitectura de las redes neuronales en el cerebro [5] A este conjunto de algoritmos y modelos se le conoce como redes neuronales artificiales (ANN). El funcionamiento de una ANN es similar al de un cerebro, cuando recibe nueva in-

formación, esta es comparada con la información que ya conoce y es etiquetada o clasificada en varias categorías [5] Las ANN se componen de un conjunto de nodos computacionales interconectados denominados neuronas, las cuales trabajan entre sí para aprender de forma colectiva desde la entrada de datos hasta optimizar la salida final.

El término DL se refiere al número de capas de neuronas que se implementan en una ANN. Estas capas se clasifican en tres: Capa de entrada (recibe los datos a procesar), capa de salida (arroja el resultado del procesamiento) y capa oculta (procesa los datos). Una ANN profunda puede contener múltiples capas ocultas, lo cual le permite realizar tareas más complejas, mientras que una ANN superficial contiene solamente una [5].

### **8.3. Redes Neuronales Convolucionales**

Las Redes Neuronales Convolucionales (CNN) mantienen similitud con las ANN tradicionales. Cada “neurona”, está formada por pesos que se pueden aprender de los datos, recibe entradas de datos y realiza una operación de producto escalar [17]. También tienen una función de pérdida con respecto a la capa anterior y pueden usar una función de no linealidad [18] [17]. La diferencia más notable entre las CNN y ANN es que las CNN reciben comúnmente imágenes como entrada, y esto permite programar algunas características dentro de la red que puedan reducir la cantidad de parámetros [18].

Las CNN difieren de las ANN principalmente en el tipo de capas ocultas que se pueden incluir. Una CNN organiza las neuronas considerando tres dimensiones: ancho, alto y profundidad, excluyendo la capa oculta. Cada capa transforma la entrada 3D en una salida 3D utilizando funciones de activación. Por ejemplo, una imagen que entra a la CNN, sus dimensiones  $(n \times m)$  son el alto  $(n)$  y el ancho  $(m)$ , y la

profundidad es un valor tres, debido a las capas RGB que conforman la imagen [18].

Generalmente, las CNN se conforman de cuatro capas principales:

- **Capa de entrada:** Contiene los datos de la imagen (píxeles).
- **Capa de convolución:** Es una pila de mapas de características. Su principal objetivo es extraer las características de la imagen de entrada, esta capa ejecuta la mayor parte de los cálculos de la red. La función de activación ReLU es la función más utilizada en las CNN [17].
- **Capa de agrupación:** La capa de agrupación toma la capa de convolución como entrada y al ser una pila de mapas de características, puede ser de una dimensión muy alta, por ejemplo: una imagen de  $6016 \times 6016$  píxeles, si se le aplica una convolución de  $1080 \times 1080$ , la entrada a la capa de agrupación sería de 1,166,400 píxeles, por lo que esto indica trabajar con más parámetros. La capa de agrupación reduce el tamaño de la convolución, reduciendo así los parámetros y cálculos.
- **Capa completamente conectada:** Esta capa produce puntuaciones (resultados) que serán utilizados para realizar la clasificación de las imágenes.

Las CNN realizan operaciones de convolución a diferencia de las ANN que realizan multiplicaciones de matrices lo cual facilita el trabajo con imágenes. En comparación con una CNN la ANN requeriría un número elevado de neuronas y capas para poder procesar imágenes lo cual repercute directamente el tiempo de entrenamiento y complejidad computacional. Las CNN aprovechan todo un volumen de entrada y lo transforman en un volumen de salida, de esta manera una neurona se encarga de procesar cierta región de la imagen, así almacena menos parámetros reduciendo el consumo de recursos y mejorando la eficacia, por lo tanto realiza menos operaciones [19] [20].

Las capas de convolución producen múltiples mapas de características este mapa de características está dado por la siguiente fórmula:

$$h_{i,j}^k = \tanh(w^k x)_{i,j} + b_k \quad (1)$$

Dónde  $w_k$  es el peso,  $b_k$  es el sesgo,  $x$  es el valor del píxel específico y  $\tanh$  es para la no linealidad de los datos. Los subíndices  $i, j$  se refieren a la entrada de la matriz que representa el mapa de características. El peso  $w_k$  es lo que conecta los píxeles en el mapa a la capa anterior.

A manera general la capa de convolución va apilando los mapas de características que se obtienen durante la operación anterior. El mapa de características se coloca en la capa de agrupación y se calcula el tamaño espacial del volumen de salida de la siguiente manera:

$$SpatialSize_{output} = \frac{W - F + 2P}{S + 1} \quad (2)$$

Dónde  $W$  es el tamaño del volumen de entrada,  $F$  es el tamaño del cuerpo receptivo en la capa de convolución,  $P$  es el relleno de valor 0 y  $S$  es el paso.

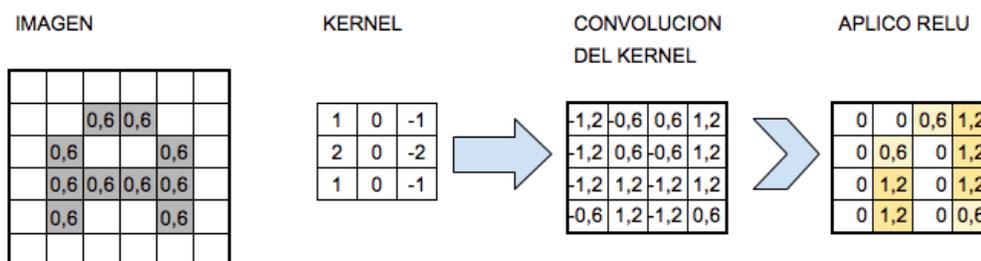


Figura 1: Convolución de una imagen aplicando kernel y función de activación RELU [21].

### 8.3.1. Parámetros de las CNN

Las CNN constan de funciones y parámetros generales que se encuentran en casi todas las topologías. A continuación se muestran las funciones más utilizadas en las CNN.

#### 8.3.1.1 Kernel

Es un parámetro utilizado en la capa de convolución para aplicar las operaciones convolutivas a los datos de entrada. Este consta de dos parámetros, el paso (stride) y el tamaño de kernel. El paso define el tamaño de la salida con base en los datos de entrada, si es igual a uno, la salida será del mismo tamaño que la entrada, si es igual a dos la salida será correspondiente a la mitad del tamaño de entrada. El kernel puede tomar cualquier dimensión rectangular o cuadrangular.

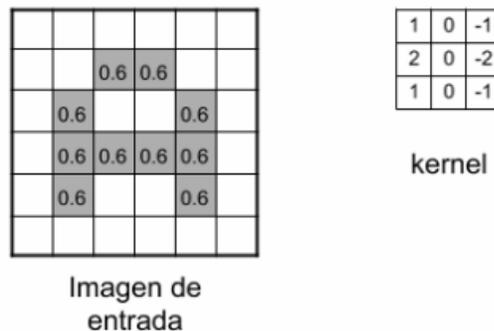


Figura 2: Ejemplo de imagen de entrada y kernel [21].

#### 8.3.1.2 Subsampling Max-Pooling

El subsampling es un paso en el que se reduce la cantidad de neuronas antes de entrar en una nueva capa de convolución, esto se realiza debido a que después de una convolución el número de neuronas se expande de manera exponencial. Supongamos que se tiene una imagen en blanco y negro de  $28 \times 28$  píxeles, de esta imagen se obtiene una capa de entrada de 784 neuronas y tras aplicar una

convolución de 32 filtros se obtiene una capa oculta de 25,088 neuronas las cuales corresponden a 32 capas de características. Si se aplicara una nueva convolución a partir de esta capa el número de neuronas sería demasiado grande, por lo tanto esto requiere mayor número de recursos, cálculos y tiempo de procesamiento.

El proceso de subsampling reduce el tamaño de las imágenes filtradas asegurándose de prevalecer las características importantes que se detectaron en cada filtro. El *Max-pooling* es una técnica de *subsampling* que funciona de la siguiente manera: Retomando el ejemplo anterior y suponiendo que se aplicará un *Max-pooling* de  $2 \times 2$  significa que se recorre cada una de las 32 capas de características pero en lugar de tomar un solo píxel se toma de  $2 \times 2$  (4 píxeles) y se preserva el valor más alto de esos 4 píxeles, de esta manera la imagen resultante se reduce a la mitad. Al finalizar se tendrán 32 mapas de características de  $14 \times 14$  píxeles y el número resultante de neuronas se reduce a 6,272.

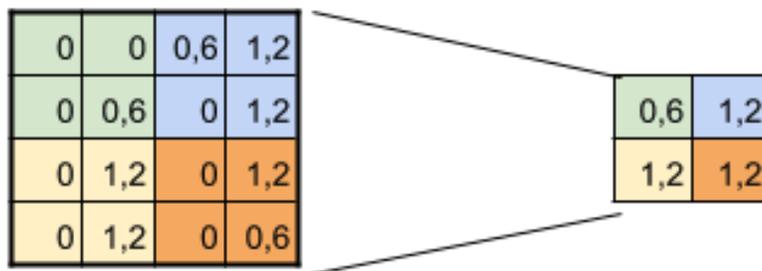


Figura 3: Ejemplo de *subsampling Max-pooling*  $2 \times 2$  [21].

### 8.3.1.3 Stride

Es *stride* es otra forma de reducir las salidas de una CNN, los *strides* controlan las zonas dónde se aplicarán las convoluciones, ya que en ocasiones no es necesario que se aplican en todas las posiciones.

Por ejemplo, si un *stride* de una CNN se establece en 1, el filtro se moverá un solo píxel o unidad a la vez. El tamaño del filtro afecta el volumen de salida codificado, por lo que el *stride* a menudo se establece en un número entero. Si por ejemplo se tiene un filtro de tamaño  $3 \times 3$  píxeles, los nueve píxeles contenidos en la matriz se convertirán en 1 en la capa de salida. A medida que aumenta el *stride* la salida se reduce aún más.

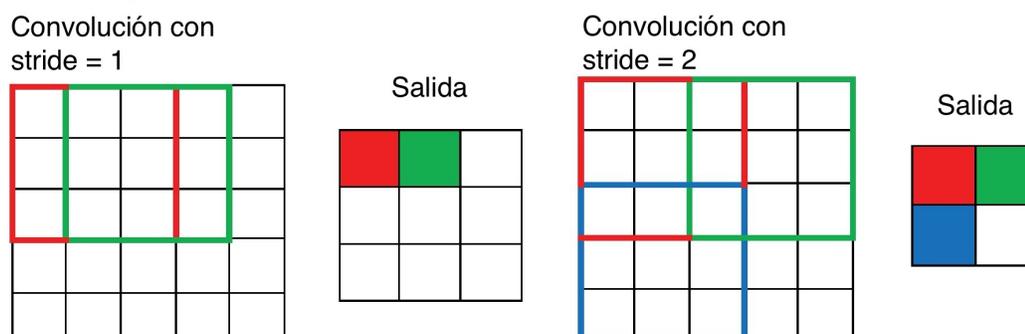


Figura 4: Ejemplo de convolución *stride=1* y *stride=2* respectivamente [22].

#### 8.3.1.4 Dropout

Cuando las ANN o las CNN majean más de una capa con la intención aproximarse a un determinado objetivo, pueden llegar a conducirse a un sobreajuste. Para esto existen algunas técnicas como el Dropout en dónde una serie de neuronas que seleccionadas aleatoriamente son ignoradas durante el proceso de entrenamiento. Al estar desactivadas queda temporalmente eliminada su contribución a las neuronas siguientes y las actualizaciones de esta neurona no se aplican en la propagación hacia atrás.

El efecto final es que la red será menos sensible a los pesos calculados por las neuronas y esto resultará en una mejor generalización de los datos y reduciendo el sobreajuste.

### 8.3.1.5 Salidas (Clasificadores)

En la gran mayoría de los casos de clasificación se buscan dos opciones, clasificación binaria (Si o No) y clasificación multiclase (1, 2, 3). Para cubrir este tipo de salidas o clasificadores existen diferentes métodos de clasificación.

- **One Hot Encoder:** Esta técnica es comúnmente utilizada para cuando se maneja un número bajo de clases, por ejemplo para una clasificación con solo dos clases el codificador arrojaría in vector de salida con dos parámetros  $[0, 1]$  y  $[1, 0]$ , para casos con tres clases devolvería tres vectores con tres parámetros cada uno  $[0, 0, 0]$ ,  $[0, 1, 0]$  y  $[1, 0, 0]$  y así sucesivamente para cada número de clases.
- **Softmax:** Es más utilizada para cuando se manejan multiclases. Esta función asigna probabilidades decimales a cada clase en un problema de varias clases. Estas probabilidades deben sumar 1, esto ayuda también a que el entrenamiento converja más rápidamente que con alguna otra técnica.
- **Categorical crossentropy:** La entropía cruzada categórica es una función que se utiliza en tareas de clasificación multiclase. Formalmente está diseñado para cuantificar la diferencia entre dos distribuciones de probabilidad. Esta función es muy buena medida de cuan distinguibles son dos distribuciones de probabilidad entre sí.

### 8.3.2. AlexNet

Alex Krizhevsky y col. en el año 2012 propusieron un enfoque de modelo de CNN amplio y profundo con el cual ganaron el desafío *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) [23], este modelo bautizado como AlexNet logró una precisión de reconocimiento impresionante en comparación con otros enfoques tradicionales de ML y VC. Fue un avance muy significativo en el campo de ML y VC para tareas de clasificación y reconocimiento de imagen, a partir de este entonces

el interés por el DL aumentó rápidamente.

La topología de la red AlexNet se muestra en la Figura 5. La primera capa convolucional realiza una operación de convolución y *Max Pooling* con LRN dónde se utilizan 96 filtros receptivos diferentes de  $11 \times 11$ . Las operaciones *Max Pooling* se realizan de  $3 \times 3$  con un *stride* de 2. Las mismas operaciones se realizan en la segunda capa con filtros de  $5 \times 5$ . En las capas tres, cuatro y cinco se utilizan filtros de  $3 \times 3$  con 384, 384 y 296 mapas de características respectivamente. Utiliza dos capas FC con *Dropout* seguida por una capa *Softmax* al final [24].

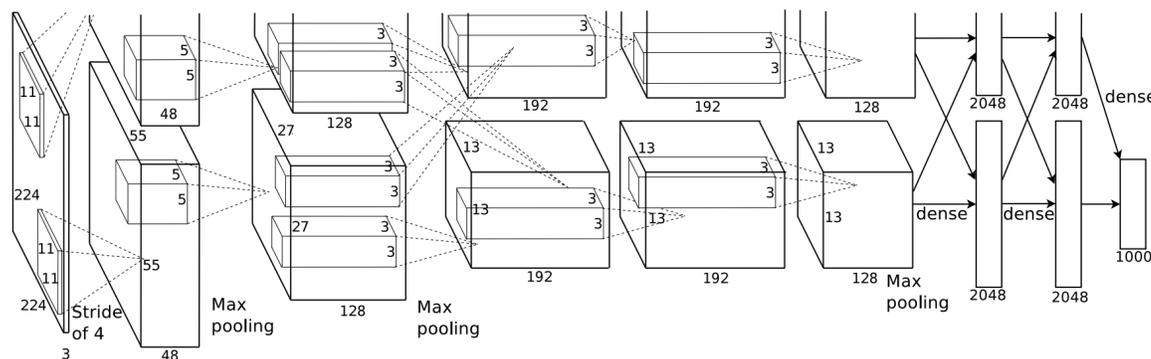


Figura 5: Topología original AlexNet [23].

### 8.3.3. HyperResNet

Las redes residuales (ResNet) fueron propuesta de Kaiming y col. En 2015 y están inspiradas en el hecho biológico de que algunas neuronas están conectadas con otras en capas dónde no necesariamente estén contiguas, realizando saltos de capas intermedias. Son redes comúnmente utilizadas en tareas de VC. Al igual que una de sus predecesoras la red AlexNet, ganó la competencia de ILSVRC en 2015. Tuvo un gran impacto debido a que este paradigma permitió por primera vez entrenar profundas con más de 100 capas, controlando exitosamente el problema de desvanecimiento de gradiente [25]. La topología de ResNet propuesta por Kaiming

se muestra en la Figura 6.

El rendimiento de la ResNet de Kaiming es impresionante, sin embargo las arquitecturas de ResNet tienen varias limitaciones potenciales: las conexiones implementadas en la ResNet conducen a la acumulación de una combinación de niveles de representaciones de características en cada capa, y en una red profunda algunas características aprendidas por capas anteriores podrían no proporcionar información útil en capas posteriores [26].

Actualmente existen herramientas interesantes que ayudan a mejorar y adaptar algunos modelos de ML de acuerdo a los datos que están recibiendo de entrada. Una de estas herramientas es *KerasTuner* [27] una librería que elige el conjunto óptimo de hiperparámetros para un modelo ResNet desarrollado con *TensorFlow*. El proceso de selección del conjunto correcto de hiperparámetros se denomina ajuste de hiperparámetros o *hypertuning*.

Los hiperparámetros son las variables que gobiernan el proceso de entrenamiento y la topología de un modelo de ML. Estas variables permanecen constantes durante el proceso de entrenamiento e impactan directamente en el rendimiento del modelo. Los hiperparámetros son de dos tipos:

- **Hiperparámetros del modelo:** Los que influyen en la selección del modelo, como el número y el ancho de las capas ocultas.
- **Hiperparámetros del algoritmo:** Influyen en la velocidad y la calidad del algoritmo de aprendizaje, como la tasa de aprendizaje para el SGD y un KNN.

*KerasTuner* adapta ambos tipos de hiperparámetros para encontrar la mejor relación exactitud-rendimiento de un modelo ResNet.

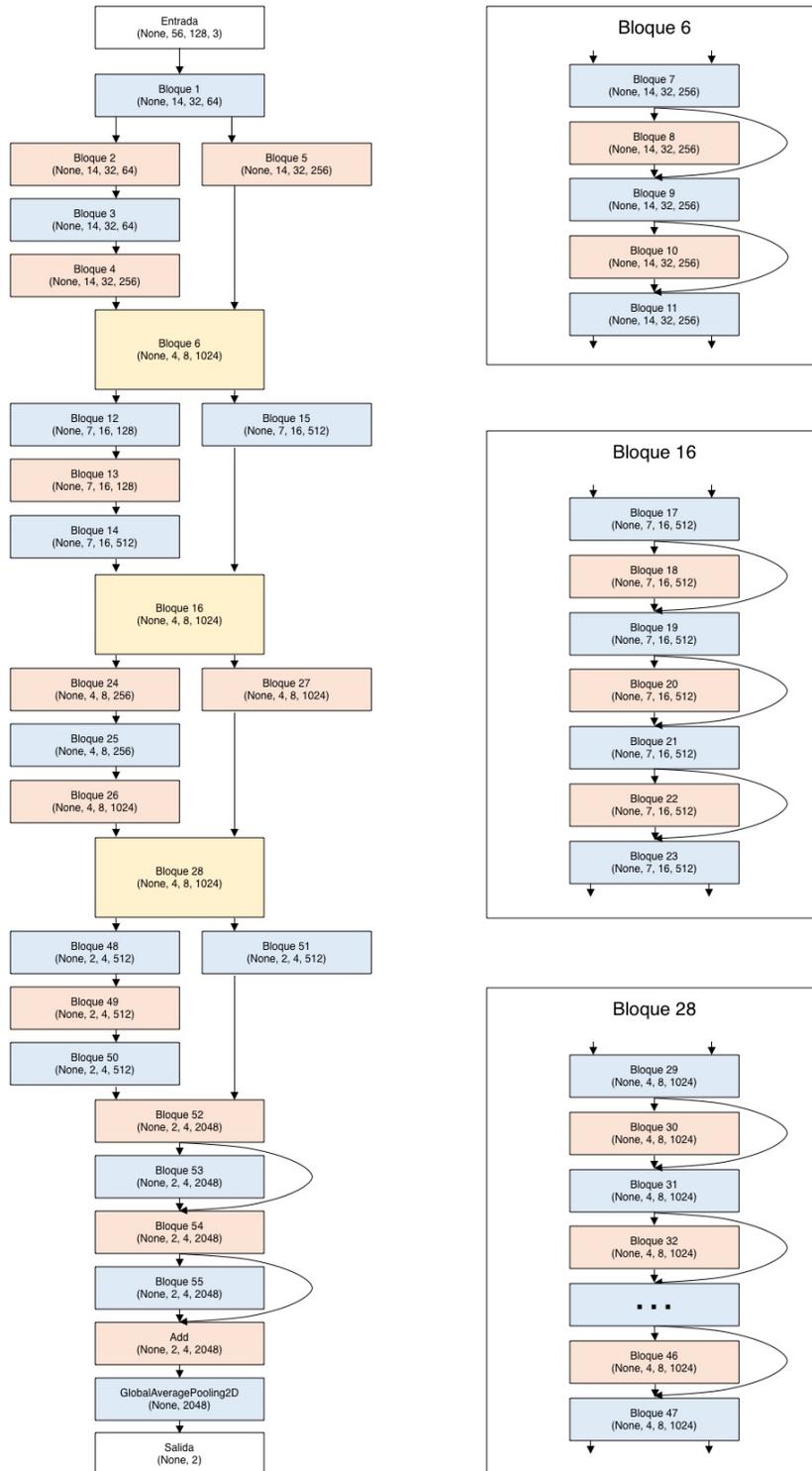


Figura 6: Topología original ResNet [25].

#### 8.3.4. CNN Paralela

Las CNN se han aplicado numerosas veces en los últimos años en tareas para el reconocimiento y clasificación de imagen, video y voz. Aunque han logrado resultados realmente impresionantes con el pasar del tiempo han llegado a ser cada vez más complejas al incluir más capas y convoluciones con el objetivo de mejorar el rendimiento y exactitud. Es bien sabido que el entrenamiento de las CNN es una tarea computacionalmente intensa por lo que realizar tareas en paralelo se ha vuelto una técnica interesante para completar el entrenamiento en tiempos aceptables.

Es verdad que las CNN paralelas pueden ocupar mayor número de memoria en comparación con CNN lineales, sin embargo sus resultados en cuanto al mayor aprovechamiento de recursos y tiempos ha logrado llamar la atención para ser implementadas en tareas que requieren una gran complejidad computacional y por supuesto mejorar su rendimiento y maximización de recursos.

Por ejemplo Sunwoo Lee y col. en 2017 [28] propuso una estrategia de paralelización que maximiza la superposición de la comunicación entre procesos con la computación. Lidy en 2016 [29] realiza una comparativa entre una red secuencial y una paralela para la detección de género musical, estado de ánimo y compositor con espectrograma de Mel. Entre otras aplicaciones más [30–32].

Las CNN paralelas se caracterizan por tener cierto número de capas actuando (valga la redundancia) de manera paralela. Es como si a partir de una entrada de datos estos se separaran de tal manera que cada secuencia de capas convolutivas procese cierta parte de los datos y al finalizar sus operaciones de convolución combinen las salidas de cada una ofreciendo un resultado final. Cada capa paralela funciona como si fuese una CNN colaborando con otras CNN al mismo tiempo para al final entregar un resultado en conjunto. Esta técnica aprovecha los recursos de

hardware para reducir el tiempo de entrenamiento. En la Figura 7 se muestra un ejemplo de una CNN Paralela.

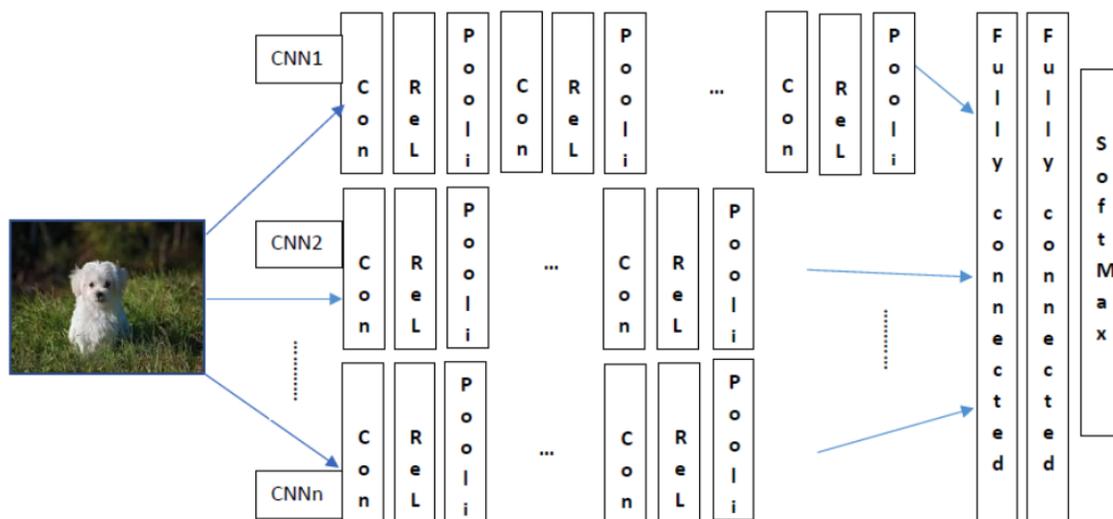


Figura 7: Ejemplo de CNN Paralela para el procesamiento de imágenes [30]

### 8.3.5. Métricas

Cualquier modelo de ML o DL es medible en cuanto a su rendimiento y exactitud, información como el tiempo de entrenamiento, complejidad computacional y exactitud de la predicción son algunos de los aspectos más buscados en la construcción de modelos inteligentes, sin embargo existen otras métricas que pueden ayudar a observar que tan bien está funcionando un clasificador.

Antes de adentrarse en las métricas y en la forma de calcularlas es importante tener en mente cuatro conceptos que todas estas métricas utilizan.

#### 8.3.5.1 Verdaderos/Falsos Positivos y Negativos

Se considera que un clasificador binario clasifica instancias de datos como positivas o negativas:

- **Positivo:** La instancia se clasifica como miembro de la clase que el clasificador intenta identificar. Por ejemplo: Si intenta clasificar la imagen de un gato y el clasificador la identifica como gato, es positivo.
- **Negativo:** La instancia se clasifica como no miembro de la clase que el clasificador intenta identificar. Por ejemplo: Si intenta clasificar la imagen de un gato y el clasificador la identifica como perro, es negativo.

La base de las métricas como exactitud, precisión, sensibilidad y *F1-Score* provienen de los conceptos de VP, FP, VN y FN que se explican en la siguiente tabla, con el valor 1 como una predicción positiva.

Predicción	Valor real	Tipo	Definición
1	1	Verdadero Positivo	Predicción positiva y fue positivo.
0	0	Verdadero Negativo	Predicción negativa y fue negativo.
1	0	Falso Positivo	Predicción positiva pero fue negativa.
0	1	Falso Negativo	Predicción negativa pero fue positiva.

Tabla 1: Definición de tipos de Positivos y Negativos.

### 8.3.5.2 Matriz de confusión

Una matriz de confusión (Figura 8) se utiliza para ilustrar el rendimiento de un clasificador en función de los cuatro valores VP, FP, VN y FN. Estos se grafican el uno contra el otro para crear una matriz que exprese los aciertos en cada una de estas categorías.

		Valores reales	
		Positivo	Negativo
Valores predicción	Positivo	VP	FP
	Negativo	FN	VN

Figura 8: Estructura de una matriz de confusión.

### 8.3.5.3 Exactitud

La métrica básica y más utilizada para la evaluación de un modelo es la exactitud, describe el número de predicciones correctas sobre todas las predicciones:

$$E = \frac{(VP + VN)}{(VP + VN + FP + FN)} = \frac{TPC}{TP} \quad (3)$$

### 8.3.5.4 Precisión

Mide cuantas de las predicciones positivas realizadas son correctas.

$$P = \frac{VP}{(VP + FP)} \quad (4)$$

### 8.3.5.5 Sensibilidad / Recall

Es una medida de cuántos de los casos positivos predijo correctamente el clasificador, sobre todos los casos positivos en los datos.

$$S = \frac{VP}{(VP + FN)} \quad (5)$$

### 8.3.5.6 F1-Score

Esta medida combina la precisión y sensibilidad. Generalmente se describe como media armónica de los dos. La media armónica es solo otra forma de calcular un “promedio” de valores, generalmente descrito como más adecuado para proporciones (como precisión y sensibilidad) que la media aritmética tradicional.

$$F1 = 2 * \frac{(P * S)}{(P + S)} \quad (6)$$

La idea es proporcionar una métrica única que pondere las dos proporciones (precisión y sensibilidad) de manera equilibrada, lo que requiere que ambas tengan un valor más alto para que aumente el valor de la puntuación  $F1$ .

Algunas ventajas de utilizar  $F1$ -Score son:

- Una precisión o sensibilidad muy pequeña darán como resultado una puntuación general más baja. Por lo tanto, ayuda a equilibrar las dos métricas.
- Si elige su clase positiva como la que tiene menos muestras, la puntuación  $F1$  puede ayudar a equilibrar la métrica entre muestras positivas/negativas.
- Combina muchas de las otras métricas en una sola, capturando varios aspectos a la vez.

### 8.3.5.7 Curva ROC

La curva ROC es una gráfica que representa la sensibilidad frente a la especificidad de un sistema clasificador binario. Para producir una curva ROC primero se tabulan las sensibilidades y especificidades para diferentes valores de una medida de prueba continua. Como resultado se obtiene una lista de valores de prueba y la correspondiente sensibilidad y especificidad de la prueba en ese valor. La curva

ROC e gráfica trazando la sensibilidad (tasa de VP) en el eje de las  $y$  contra la especificidad (tasa de FP) en el eje  $x$  para los diversos valores tabulados [33].

Una curva ROC que sigue la línea diagonal  $y = x$  produce resultados falsos positivos al mismo ritmo que verdaderos positivos. Por lo tanto, se espera que una prueba diagnóstica con una precisión razonable tenga una Curva ROC en el triángulo superior izquierdo sobre  $y = x$  lo cual se toma como línea de referencia, esto se representa en la Figura 9.

El área bajo la curva ROC (AUC) es una medida global es una medida global que se utiliza para medir la capacidad discriminativa de un modelo de ML o DL. Un AUC de 0,5 representa una prueba sin capacidad de discriminar (es decir, no mejor que el azar), mientras que un AUC de 1,0 representa una prueba discriminativa perfecta. De manera general se podría categorizar los resultados de AUC de la siguiente manera:

- 0,5: Sin valor discriminativo, como lanzar una moneda.
- 0,5 a 0,6: Test malo.
- 0,6 a 0,75: Test regular.
- 0,75 a 0,9: Test bueno.
- 0,9 a 0,97: Test muy bueno.
- 0,97 a 1,0: Test excelente.

### **8.3.6. Retinopatía Diabética**

La Diabetes Mellitus (DM), es definida por la Organización Mundial de la Salud (OMS) como una enfermedad metabólica que se caracteriza por los altos niveles de

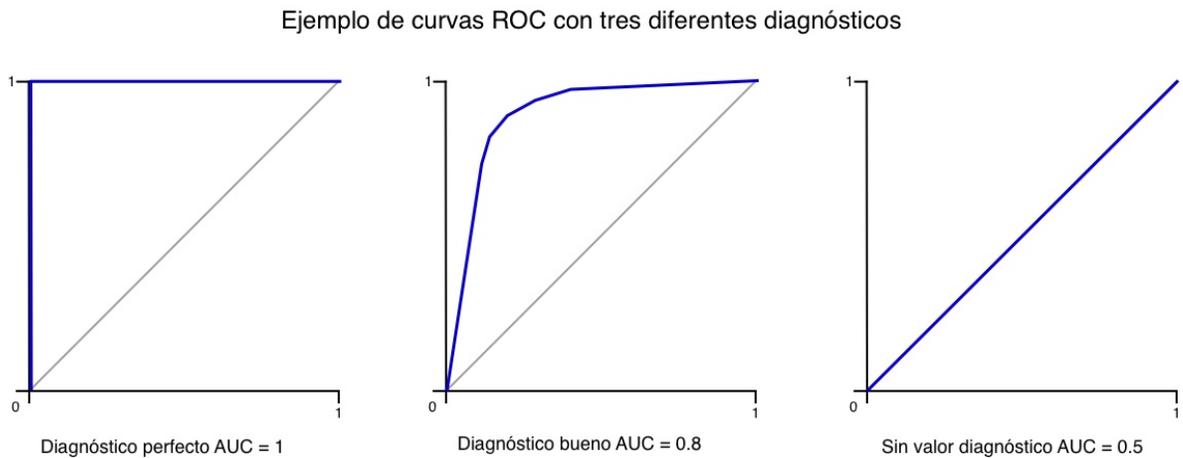


Figura 9: Ejemplos de Curvas ROC, diagnóstico perfecto, diagnóstico bueno y sin valor diagnóstico respectivamente.

azúcar en la sangre y la alteración al metabolismo de los hidratos de carbono, grasas y proteínas a causa de los defectos en la acción y/o secreción de insulina [35].

La DM es clasificada principalmente en el tipo 1 (DM1), donde no existe una secreción de insulina a causa de una destrucción autoinmune de las células beta-pancreáticas que producen la hormona, y el tipo 2 (DM2), donde se observa una resistencia a la acción de insulina y/o una secreción de insulina inadecuada. La DM2 se relaciona con otros factores como la hipertensión, los desniveles de colesterol y un aumento del riesgo cardiovascular [35].

La diabetes causa complicaciones relacionadas con los desniveles de glucemia, como daño renal, pérdida de sensibilidad periférica y pérdida de la visión [36, 37]. La Retinopatía Diabética (RD) es una de las principales causas de pérdida de visión asociada con la diabetes, sin embargo no es la única. La diabetes aumenta el riesgo de padecer cataratas [38] y glaucoma [39], que se relacionan también con ceguera. La RD es una de las causas más comunes de nuevos casos de pérdida de visión a nivel mundial entre los adultos de 20 a 74 años [40]. Casi todos los pacientes con

diabetes tipo 1 y aproximadamente el 60 % de los pacientes con diabetes tipo 2 padecen esta enfermedad [40].

Más del 80 % de las personas que han padecido diabetes por más de 10 años presentan algún grado de RD [41]. Para los pacientes con DM2, se estima que hasta el 21 % padezcan RD al momento de diagnosticar la diabetes [41]. Estudios demuestran que las terapias para control intensivo de glucemia y un seguimiento continuo, mayor a tres años, puede proteger la retina en personas con diabetes. Además, lesiones de RD equivalentes o menos graves que el grado de RDNP moderada [42], se pueden controlar solo si la hemoglobina glucosilada A1c en pacientes con DM1 y DM2 se reduce al menos en un 0.8 % [42]. Si la gravedad de las lesiones de RD es más avanzada que el grado de RDNP moderada, es posible que el control glucémico no produzca beneficios. La detección temprana de RD permite actuar de tal manera que sea posible detener su desarrollo, pero desgraciadamente no corregir el daño ya causado [42].

En la clínica, se consideran varios grados de RD según el daño vascular. La RD no proliferativa (RDNP) que puede ser mínima, leve, moderada, severa y muy severa [40]. El otro grupo es la RD proliferativa (RDP) de estadio moderada, de alto riesgo y avanzada [38]. El enfoque vascular se debe en gran parte a que el deterioro de los vasos se puede identificar por inspección visual y en etapas avanzadas, correlacionada con la pérdida de la visión [42, 43]. Desde hace ya varios años, el ETDRS (Early Treatment Diabetic Retinopathy Study) ha servido como el principal sistema de estadificación para la atención e investigación de la RD [44] (Tabla ??).

<b>Nivel de severidad</b>	<b>Hallazgos en la oftalmoscopia bajo midriasis</b>
Sin retinopatía aparente	Ninguna anomalía.
Retinopatía no proliferativa leve	Microaneurismas (1 a 3).
Retinopatía no proliferativa moderada	Microaneurismas (más de 3) pero menos signos que en el grado severo.
Retinopatía no proliferativa severa	Más de 20 hemorragias en cada cuadrante, irregularidades venosas en dos cuadrantes o anomalías vasculares en uno. Sin signos de retinopatía proliferativa.
Retinopatía proliferativa	Uno o más de los siguientes: <ul style="list-style-type: none"> <li>• Neovasos</li> <li>• Hemorragia vítrea o prerretiniana</li> </ul>

Tabla 2: Escala Internacional de Severidad de la Retinopatía Diabética (ETDRS) [44].

Sin embargo, desde el punto de vista de la estadificación, la ETDRS no se desarrolló para evaluar la efectividad de los tratamientos que mejoran la RD y tampoco define adecuadamente los estadios de gravedad ya que no contempla la neurodegeneración [44]. Muy recientemente, Sun y col [44] han propuesto una escala de gravedad de RD que considera la patología retiniana neural y vascular y sus afectaciones en la función visual relacionadas a la diabetes (Figura 10, modificado de [44]). El sombreado de fondo representa el estado de salud sistémica. La etiqueta A, representa un ojo con excelente visión y anomalías retinianas, neurales o vasculares mínimas o leves en un paciente con buena salud sistémica. La etiqueta B, representa

un ojo con pérdida moderada de la visión, por ejemplo, causado por edema macular diabético y enfermedad neurovascular en un paciente con algunos problemas de salud sistémica. La etiqueta C, muestra un ojo con enfermedad avanzada, por ejemplo, desprendimiento de retina por tracción macular y pérdida de visión severa en un paciente con mala salud sistémica. La escala anterior, puede proporcionar más información que no se resuelve con la ETDRS y guiar a los médicos en la toma de mejores decisiones terapéuticas para los pacientes [44].

La RD es una enfermedad que difícilmente se detecta en etapas tempranas, debido a que el paciente no tiene síntomas hasta que comienza la pérdida de visión, y en el transcurso de pocos años puede desarrollarse rápidamente [14]. Además, existen estrategias terapéuticas que contribuyen a detener su progreso, pero no corrigen el problema [14]. Uno de ellos es la fotocoagulación con láser, la cual consiste en quemar la retina neural alrededor de las zonas de hemorragias, para disminuir la demanda energética de las células retinianas y por ende, la hipoxia y la consecuente formación de nuevos vasos sanguíneos, con el fin de retrasar el progreso de la RD [5]. Cabe mencionar que esa estrategia se usa también para tratar el glaucoma y tumores retinianos [5]. Se usa también la vitrectomía, que consiste en una pequeña incisión en el ojo para extraer el vitreo con su sangrado y el tejido cicatrizal que tira de la retina, y reemplazarlo con gel de silicón [45].

Algunas herramientas de imageneología avanzada, como la tomografía de coherencia óptica (OCT) de dominio espectral, la angiografía y la oftalmoscopia permiten valorar la retina periférica y central e inferir sobre la integridad de la unidad neurovascular de la retina [6]. Las imágenes de fondo de ojo son el estándar de oro para la detección de la RD [46]. Permite contar el número de microneurismas, hemorragias, exudados y su extensión física. Cabe señalar que se ha investigado el uso de las herramientas de Inteligencia Artificial (IA) para la clasificación automatizada de estas imágenes, con el fin de mejorar las opciones diagnósticas, la toma de decisión

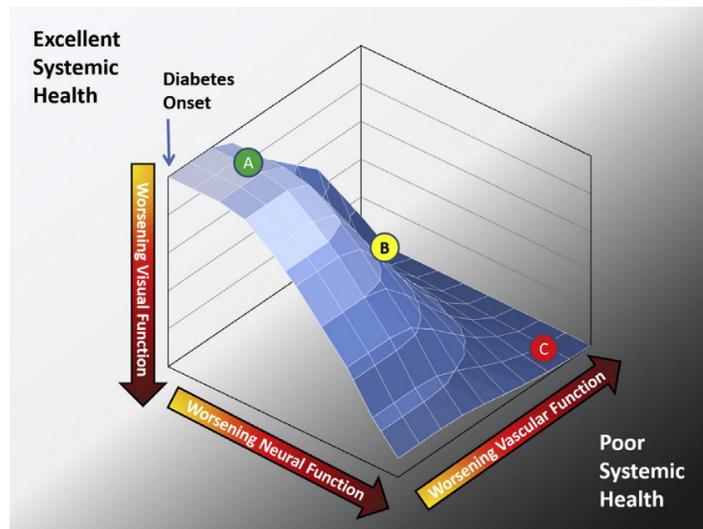


Figura 10: Escala de gravedad de RD propuesta por Sun y col. en 2020 [44].

terapéuticas y el monitoreo de tratamiento para la RD [6] [47, 48]. Sin embargo, los modelos existentes carecen de la capacidad para predecir la RD en etapas tempranas, debido a que se enfocan en el análisis de imágenes de la retina para detectar signos clínicos de la RD como las hemorragias [49].

### 8.3.7. Electrorretinograma

Existen técnicas electrofisiológicas para medir la función eléctrica de la retina y que son de gran interés para el diagnóstico de RD. El electroretinograma (ERG), obtiene registros de las señales eléctricas emitidas por las células de la retina, en respuesta a un estímulo luminoso. Más específicamente, la luz produce un cambio eléctrico bifásico en la córnea en relación con la retina en la parte posterior del ojo [50].

Las respuestas de un ERG varían en amplitud y patrón dependiendo de la especie (ratones, humanos, etc.) en la que se aplican los destellos de luz, debido a la proporción de bastones y conos que son específicos para cada especie, la duración e intensidad de la estimulación y los métodos de registro [14]. Sin embargo, la ma-

yoría de las especies de vertebrados comparten características en el ERG, como una onda *a* negativa y una onda *b* positiva.

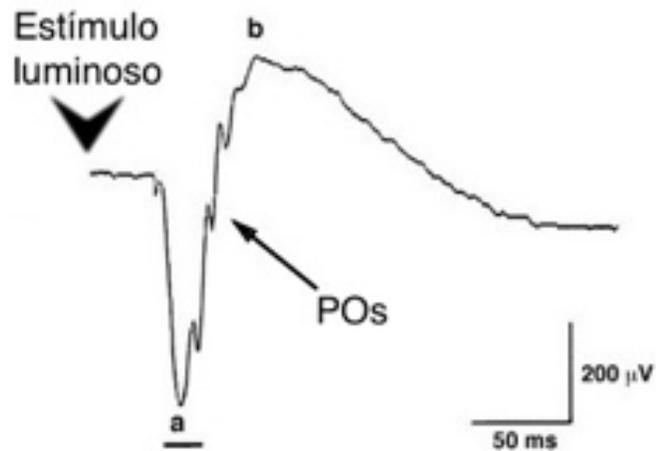


Figura 11: Ejemplo típico de ERG humano [38].

La onda *a* es una onda negativa que aparece primero después del destello de luz, seguida de una onda *b* positiva [14]. Todo lo anterior fue propuesto por Gotch en 1903 y mejorado por Einthoven y Jolly en 1908 [14]. Estos principios son los que definieron las bases para analizar señales de ERG.

Una característica importante que aparece en las señales de ERG son los Potenciales Oscilatorios (PO). Son las pequeñas depresiones que se observan en la fase ascendente de la onda *b* (Figura 11, modificado de [14].) Para valorar estos POs, es necesario aplicar un filtro digital de paso alto (mayor a 60 Hz) y amplificar la señal [14]. La carencia o baja manifestación de los POs, indica que existe un problema de procesamiento de la información visual [51]. No se sabe con absoluta certeza, pero evidencias apuntan a las neuronas amacrinas, que forman circuitos inhibidores en la capa plexiforme interna de la retina, como generadoras de los POs [52].

La amplitud de onda *a* y *b* y su tiempo implícito, es decir el tiempo que separa el estímulo luminoso de amplitud máxima de la onda, son también indicadores de la integridad anatómica y funcional de las células que las originan. Los fotorreceptores generan la onda *a* y las células de Müller y células bipolares la onda *b* [51]. De forma notable, estudios en ratones obesos por dieta inducida alta en grasa (modelos de prediabetes) desarrollaron a los doce meses enfermedad microvascular y a los seis meses intolerancia a la glucosa y disfunción electroretinográfica, manifestando tiempos implícitos aumentados y amplitudes reducidas de sus POs [53], estas anomalías electroretinográficas se correlacionaron con la intolerancia a la glucosa [53]. Estos resultados muestran que la enfermedad de la retina diabética presenta anomalías electroretinográficas e inflamatorias antes del desarrollo de lesiones vasculares que son comunes en la RD, estableciendo un modelo para evaluar nuevas intervenciones para tratar la enfermedad ocular [53].

Con los avances de la tecnología y el paso del tiempo, fueron apareciendo diversos tipos de ERG. El ERG multifocal (mfERG) es de los más eficientes hoy día, debido a que produce registros de más de 100 sectores retinianos diferentes [54] y permite registrar la respuesta de los conos [55]. Consiste en la aplicación de estímulos luminosos circunscritos en hexágonos de distintos tamaños según la zona de la retina en una secuencia temporal aleatoria y la reconstrucción de la respuesta asociada en cada hexágono [56]. Se conforma principalmente de tres ondas, la onda N1 que es una onda bifásica negativa inicial, seguida de un pico positivo denominado P1 y una segunda onda negativa final conocida como N2 [56]. La onda N1 se produce al igual que la onda *a* por los fotorreceptores, mientras que la P1 está formada básicamente por las células bipolares, al igual que la onda *b* [56]. La onda *s* [57] es otro tipo de onda positiva que aparece en el mfERG al momento de que descende la primera onda P1, la cual se aprecia mejor en periodos más largos de estimulación. Las propiedades de esta onda en ojos sanos sugieren que probablemente se originan en las células ganglionares de la retina [57], esta sugerencia se

respalda por su ausencia en ojos con neuritis óptica [57].

El ERG de patrón (PERG) registra la respuesta de la retina central con un patrón geométrico de estimulación luminosa, de tipo damero o a rayas, reversible, blanco y negro, con características de contraste y luminancia controladas con un punto de fijación central [55]. Permite la diferenciación electrofisiológica entre maculopatías y neuropatías periféricas y también facilita la interpretación de los potenciales visuales evocados [58]. En las señales de ERG-P de pacientes sin problemas de visión se pueden apreciar al menos tres ondas. La primera, una onda negativa pequeña denominada N35 se presenta a los 35 ms aproximadamente, la segunda, una onda positiva denominada P50, alcanza un pico aproximadamente a los 50 ms y la última es una onda negativa llamada N95, aproximadamente a los 95 ms [59].

Una gran ventaja de utilizar el ERG es que no causa efectos secundarios y debido a que los electrodos se colocan en la piel cercana a los ojos o sobre la córnea, detectan la señal eléctrica de las células de forma no invasiva [60]. Durante el proceso de registro de señales, el sujeto de estudio debe estar mirando fijamente un estímulo luminoso, la señal de salida es amplificada y digitalizada [60]. Esta señal se interpreta en función de la amplitud del voltaje y el tiempo en el que se manifiesta, como descrito previamente.

La alta frecuencia de la aparición de RD al momento del diagnóstico de diabetes tipo 2, es debida a que esta enfermedad comienza cuatro o cinco años antes del diagnóstico clínico [4]. Al llegar a los 10 años de DM2, el número de pacientes que presenta RD es del 35 al 40 % [4]. Por lo tanto, se sugiere establecer criterios para su detección temprana, dónde se incluye la utilización del mfERG [44].

La realización de los estudios de ERG, así como su análisis e interpretación en el contexto clínico son complicados. De acuerdo con los estándares de ISCEV (Socie-

dad Internacional de Electrofisiología Clínica de la Visión) [61]., la evaluación clínica de los registros ERG se basa en la medición de tiempos implícitos y amplitudes de ondas particulares. Existe mucha variabilidad entre los registros de pacientes, por lo que este tema genera dependencia en la experiencia del personal experto para reconocer los picos con precisión, lo cual es una complicación en el desarrollo de sistemas automatizados para el análisis de señales ERG [62].

No obstante, existen algunos métodos que pueden facilitar la identificación de características en las señales de un ERG y distinguir entre casos normales y anormales con mejor exactitud, aunque no se han aplicado en el campo de la RD. Los métodos para análisis se dividen en tres grupos generales [62]. El Grupo 1 se centra en la exactitud para determinar los parámetros (amplitud y tiempo) de las señales del ERG. El Grupo 2 permite la reducción de dimensionalidad y actuar como preprocesador de algoritmos de clasificación ya sean aplicados a imágenes, señales o cualquier tipo de información que requiera ser clasificada. El Grupo 3 realiza una clasificación automática de ERG usando criterios de cálculo propios no legibles por humanos. Entre algunos de los métodos de clasificación podemos encontrar los siguientes:

- **Funciones discriminantes y análisis de regresión (Clasificadores Bayesianos):** Abreviado como D&RA, es un algoritmo de clasificación que permite identificar funciones lineales o cuadráticas capaces de distinguir entre dos muestras de aprendizaje. Requiere una suposición sobre la distribución normal de características. Resulta en la obtención de un conjunto reducido de parámetros importantes para el diagnóstico [62].
- **Transformada de onda continua (CWT):** Mejora en la precisión para medición de tiempos y amplitudes de la forma de onda en el dominio de coeficientes CWT [62].
- **Transformada de onda discreta (DWT):** Es un preprocesador de clasificado-

res enfocado en la eliminación de ruido de la señal y reducir dimensionalidad de la información de entrada. Mejora el rendimiento del algoritmo de clasificación aumentando la proporción de muestras de aprendizaje en la tercera dimensión [62].

- **Redes Neuronales Artificiales (ANN):** También conocido como algoritmo de clasificación, es usado para distinguir entre señales normales y anormales ofreciendo una clasificación acertada y rápida [62].
- **Análisis de Componentes Principales (PCA):** Se usa principalmente para reducción de dimensionalidad de datos de entrada. Puede ser usado como preprocesador para algoritmos de clasificación o como un método de visualización en dos dimensiones de información multidimensional [62].
- **Análisis de componentes independientes (ICA):** Se utiliza para eliminación de la señal y recuperación óptima de la grabación de ERG [62].

Desde hace algunos años se ha demostrado a través del análisis de señales de ERG, como es que las alteraciones en las oscilaciones de las señales ERG alertan sobre la aparición de daños en la retina. Dado es el caso de la Figura 12 [63] dónde se observan alteraciones en el ffERG evocado de una persona normal y dos casos de pacientes sin retinopatía aparente o sea sin daños visibles. Se puede observar claramente como los potenciales oscilatorios que aparecen en la señal del paciente normal disminuyen significativamente tanto en el Caso 1 como en el Caso 2 y existe una ralentización en los tiempos implícitos del ffERG.

Otro ejemplo muy distintivo es el caso de la Figura 13 dónde en A se puede ver un montaje de una imagen de fondo de ojo dónde se resaltan las zonas con hemorragias o vasos sanguíneos rotos, en B se ve una representación de un mfERG, el cual secciona la retina en varias partes y en cada una de estas secciones hexagonales el mfERG detecta la actividad eléctrica, los números que aparecen en

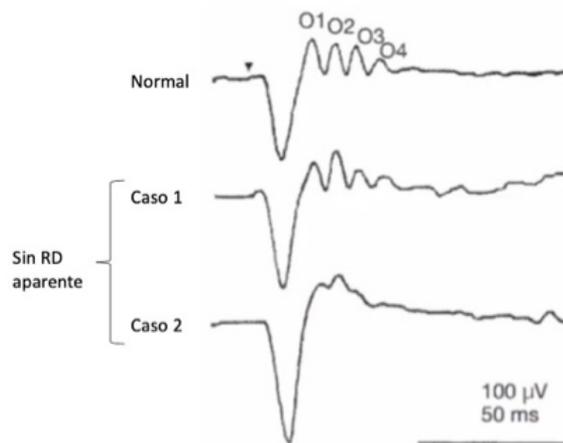


Figura 12: Diferencias en ERG evocado de campo completo en persona sana y personas con diabetes sin retinopatía aparente [63].

los hexágonos es el retraso en milisegundos existente entre los tiempos implícitos promedio de mfERG de un paciente sano contra ese paciente con RD. Es notorio a simple vista como existe una relación entre las zonas con hemorragia y las secciones del ERG dónde el retraso es mayor.

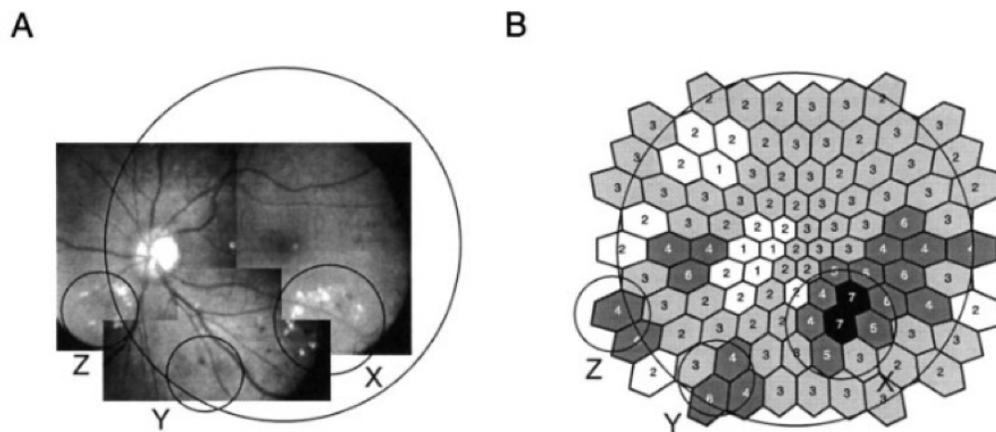


Figura 13: A) Montaje fotográfico de fondo de ojo. B) Mapa de tiempos implícitos de ERG multifocal evocado [54].

Sin embargo también se puede apreciar retrasos menores en zonas alrededor

de dónde se encuentran los daños, las alteraciones en ocurridas en los potenciales oscilatorios indican que la actividad eléctrica comienza a tener retrasos antes incluso de que exista un problema identificable a simple vista en la imagen de fondo de ojo, ya que en los dos casos de la Figura 12 el diagnóstico del oftalmólogo es que aún no existe retinopatía diabética.

Estos estudios han sido las bases para el desarrollo de sistemas de IA con la intención de utilizar las señales de ERG para la detección y predicción de enfermedades en la retina a manera de mejorar y agilizar los diagnósticos de forma temprana.

### **8.3.8. Transformada Wavelet**

Una *Wavelet* son señales o formas de onda que mantienen una duración limitada y un promedio de cero [64]. Pueden ser del tipo irregulares o asimétricas, estas características les permiten adaptarse mejor en el análisis de señales que la transformada de Fourier.

El catálogo de *wavelets* es extenso, en el podemos encontrar los tipos *Haar*, *Daubechies*, *Mexican Hat*, *Symlet*, entre otros. Elegir una familia *wavelet* depende del tipo y naturaleza de la señal que se desea analizar o de la información que se quiere obtener de ella [64]. Uno de los criterios más comunes al momento de escoger una *wavelet* es la forma de la señal, se puede comenzar por escoger las familias de *wavelets* que mantengan la mayor similitud, y a través de la realización de pruebas se elige la que obtiene mejores resultados.

La *wavelet* elegida para la implementación de la transformada se le conoce como *wavelet* madre, esta recibe una serie de modificaciones como expansiones o compresiones a lo largo de la señal, este proceso permite obtener los detalles y componentes globales de la señal que se está procesando [64]. Matemáticamente

la *wavelet* madre se define por la siguiente fórmula considerando los parámetros de escalamiento y desplazamiento:

$$\psi_{u,s}(t) := \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right), u \in \mathbb{R}, s > 0. \quad (7)$$

Dónde  $s$  es el escalamiento y  $u$  es el desplazamiento. Descrito de manera general, la transformada *wavelet* descompone una señal utilizando versiones escaladas y desplazadas de la *wavelet* madre. Esta técnica actúa como un filtro pasa banda que solamente permite el paso de ciertas características de la señal en una frecuencia determinada [64].

La CWT es un tipo de transformada *wavelet* que está definida como la sumatoria del producto de una señal continua y la *wavelet* madre en su forma desplazada y escalada  $\psi_{u,s}(t)$ . Sea  $f \in L^2(\mathbb{R})$  la CWT de  $f$  al tiempo  $u$  y escala  $s$  está definida como:

$$Wf(u, s) := \langle f, \psi_{u,s}^*(t) \rangle dt \quad (8)$$

devuelve los componentes de frecuencia de  $f$  correspondiente a la escala  $s$  en el tiempo  $u$  [64]. La teoría *wavelet* proviene del hecho de que los dos parámetros (tiempo  $t$  y escala  $s$ ) de la CWT en la fórmula anterior hacen posible el estudio de una señal en ambos dominios de tiempo y frecuencia simultáneamente, con una resolución que depende de la escala de interés [64]. De acuerdo con estas consideraciones la CWT provee una descomposición de tiempo-frecuencia de una señal llamado plano de tiempo-frecuencia [65]. Este método es más exacto y eficiente que otras técnicas como la ventana de la transformada de Fourier (WFT) [66]

El escalograma de  $f$  está definido por la función

$$\mathcal{S}(f) := ||Wf(s, u)|| = \left( \int_{-\infty}^{\infty} |Wf(s, u)|^2 du \right)^{\frac{1}{2}} \quad (9)$$

representando la energía de  $Wf$  en la escala  $s$ . Los valores que se obtienen en

$S(f)$  son mayor o igual a 0 para todas las escalas  $s$ , y si  $S(f) > 0$  se dice que la señal  $f$  tiene detalles en la escala  $s$ . Por lo tanto, el escalograma permite la detección de las frecuencias más representativas de una señal, esto es, las escalas que contribuyen a la mayoría o al total de energía de una señal. Si existe un intervalo de interés específico  $[t_0, t_1]$ , se puede definir la ventana de escalograma correspondiente como [64]:

$$S_{[t_0, t_1]}(s) := ||Wf(s, u)||_{[t_0, t_1]} = \left( \int_{t_0}^{t_1} |Wf(s, u)|^2 du \right)^{\frac{1}{2}} \quad (10)$$

Según A. Mohammad, *Harr, Daubechies, Symlet, Mexican Hat* y *Morlet* son algunas de las wavelets que son eficientes para el análisis de señales ERG [67]. En el caso de este proyecto se utilizó la familia Morlet para generar los escalogramas. Una *wavelet Morlet* está definida como una onda sinusoidal adelgazada en forma Gaussiana. Para el análisis de tiempo-frecuencia, se utiliza una *wavelet* de Morlet compleja en la que la *wavelet* Gaussiana disminuye una onda sinusoidal compleja. La *wavelet* de Morlet compleja se convoluciona con la señal de la serie de tiempo y el resultado de la convolución es una señal de valor complejo de la cual se puede extraer potencia y fase de cada punto de tiempo. En otras palabras, el resultado de la convolución es una serie temporal de “similitudes” entre la señal y la *wavelet* [68].

Una *wavelet* de *Morlet* compleja  $w$  está definida como el producto de una onda sinusoidal compleja y una ventana Gaussiana.

$$w = e^{2i\pi ft} e^{-\frac{t^2}{2\sigma^2}} \quad (11)$$

Dónde  $i$  es el operador imaginario ( $i = \sqrt{-1}$ ),  $f$  es la frecuencia en  $Hz$ ,  $t$  es el tiempo en segundos y  $\sigma$  es el ancho de la Gaussiana, definido como:

$$\sigma = \frac{n}{2\pi f} \quad (12)$$

El parámetro que define la compensación de precisión de tiempo-frecuencia es

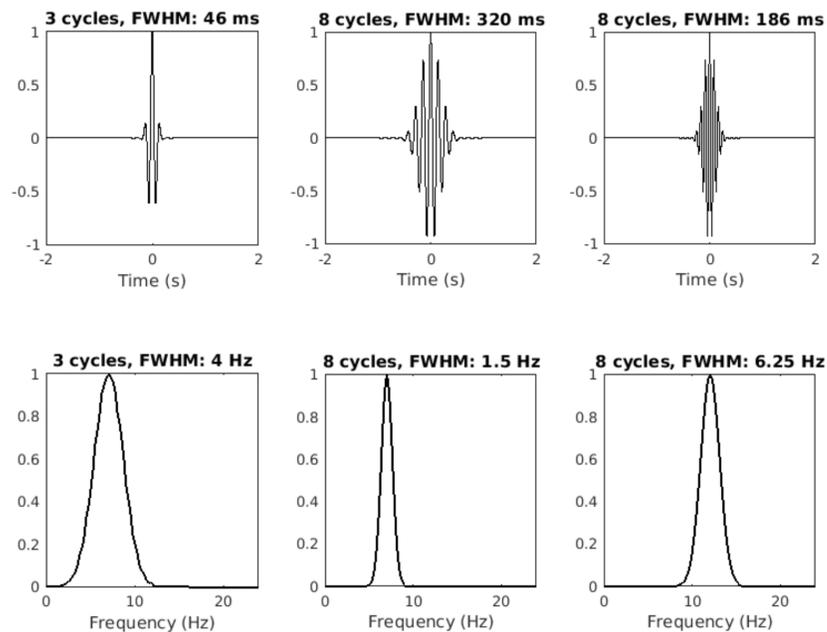


Figura 14: Tres wavelets de Morlet en el dominio del tiempo (fila superior) y en el dominio de la frecuencia (fila inferior) con dos frecuencias y diferentes parámetros de compensación de tiempo-frecuencia [68].

el número de ciclos ( $n$ ) el cual dependerá de la aplicación que se le requiera dar a la *Morlet*.

Algunas ventajas que ofrece la *wavelet Morlet* para el análisis de tiempo-frecuencia son las siguientes:

1. Su forma Gaussiana en el dominio de frecuencia evita la aparición de bordes afilados lo cual minimiza los efectos de onda que pueden malinterpretarse como oscilaciones.
2. Los resultados de la convolución *wavelet* conservan la resolución temporal de la señal original.
3. Es más eficiente desde el punto de vista computacional, implica menor número de cálculos que se implementan con la transformada rápida de Fourier.

Estado del arte			
Objetivo	Técnica	Precisión	Autores
Apoyo en diagnóstico de problemas maculares con ERG-P.	ANN Multicapa y ALM	80 %	Güven et al. 2006
Diagnóstico de enfermedad del nervio óptico con ERG-P.	ANN Multicapa y ALM	94.2 %	Kara et al. 2006
Diagnóstico de enfermedad del nervio óptico con ERG-P.	ANN Multicapa y AQV	92 %	Kara et al. 2007
Diagnóstico temprano de Glaucoma con mfERG.	ANN con FBR y MAE	98 %	Boquete et al. 2012
Análisis por CWT en señales de mfERG para diagnóstico de Glaucoma (Detección de sectores glaucomatosos).	ANN con FBR y coeficientes CWT.	Sensibilidad: 89 % Especificidad: 84 %	Jiménez et al. 2015
Predicción de desarrollo de lesiones de RDNP en áreas retinianas locales en un periodo de 3 años en pacientes con RD existente.	MRL basado en retrasos de TI del mfERG	88 %	Jason et al. 2008
Predicción local de la aparición de RD en pacientes sin retinopatía previa.	MRL considerando el desarrollo de RD y variables como z-score del TI del mfERG, sexo, duración de la diabetes, glucemia, hemoglobina glucosilada, edad y tipo de diabetes.	80 %	Harrison et al. 2011
Clasificador híbrido inteligente para estimar niveles de obesidad con señales ERG de campo completo.	ANN-PSO y DWT Daubechies 4 (db4)  Reducción de dimensionalidad por Características estadísticas	95.74 %	Yapici et al. 2021

Tabla 3: Estado del arte.

### 8.3.9. Estado del arte

Algunos de los trabajos más significativos con ERG se muestran en la Tabla 3, dando seguimiento a como ha ido evolucionando el tratamiento y uso de estas señales para la detección de enfermedades visuales como glaucoma, nervio óptico, problemas maculares, RD y obesidad.

Con el pasar del tiempo estos estudios se van acercando cada vez más a la detección temprana de enfermedades visuales.

Güven y col. en el 2006 [69] propusieron un procedimiento para simplificar el

diagnóstico de la RD y apoyar a los médicos en emitir un juicio sobre la existencia de problemas maculares, derivados de patologías como traumatismos, infecciones, degeneración, problemas vasculares o inflamaciones, a través del análisis de señales de PERG, dónde se pueden observar diferencias en amplitud de ondas N35, P50 y N95 y tiempo de respuesta entre las señales de pacientes sanos y pacientes con problemas maculares [69]. Implementaron una red neuronal artificial (ANN) de retroalimentación multicapa y un algoritmo de aprendizaje por retropropagación Levenberg-Marquart, obteniendo una sensibilidad del 96 %, una especificidad del 100 % y una clasificación correcta del 98 % [69]. La red neuronal clasifica las señales como “sanos” y “enfermos” y los resultados de las pruebas de clasificación fueron confirmados por expertos médicos [69].

Kara y col. en el 2006 [59], realizaron un estudio orientado al diagnóstico de enfermedad del nervio óptico, a partir de señales PERG divididas en dos categorías “ojo sano” y “ojo enfermo” (con problemas del nervio óptico). Realizaron la implementación de una ANN de retroalimentación multicapa y algoritmo de retropropagación de Levenberg-Marquart (LM) que puede identificar las características (amplitud, latencia y tiempo de respuesta) de las ondas N35, P50 y N95 y clasificarlas en “sanos” y “enfermos”. Obtuvieron una sensibilidad del 96.4 %, especificidad del 90.4 % y una predicción positiva del 94.2 %. Los resultados de las pruebas de clasificación fueron evaluados por expertos médicos, quienes confirman una clasificación exitosa de las señales por parte de la ANN.

En 2007, Kara y col. [70], utilizaron una ANN supervisada de cuantificación de vectores para la clasificación de señales de PERG, tomando en cuenta las características (amplitud y tiempo de respuesta) de las ondas N35, P50 y N95, con la intención de diagnosticar problemas en el nervio óptico clasificándolas como “ojo sano” y “ojo enfermo”. Obtuvieron un 94 % de sensibilidad, 90.32 % de especificidad, 5.94 % de falso negativo, 9.67 % de falso positivo y una clasificación correcta

del 92 %. Los resultados de la ANN fueron evaluados por expertos médicos, quienes confirman una clasificación exitosa.

Asimismo, Boquete y col. en el 2012 [71], utilizaron un método de clasificación a partir de una ANN de función de base radial (RBF) y un algoritmo de Máquinas de aprendizaje extremo, el cual analiza las señales entregadas por un mfERG con la finalidad de detectar sectores glaucomatosos. Basados en métodos actuales de medición de amplitudes de señales mfERG [72, 73], latencias [74] o ambos [75] y en los análisis de ondas S de señales mfERG cuya amplitud se relaciona con el desarrollo de glaucoma [57] concluyeron que era posible determinar el estado glaucomatoso o saludable de la retina. La información de entrada a la ANN fue el vector de las 13 características que conforman cada registro de forma de onda del mfERG [71]. Cada uno de los componentes en el vector se integra en una de dos categorías: características posicionales ( $c_1 \dots c_6$ ), las cuales marcan la ubicación temporal y la amplitud máxima de cada onda y las características de transición ( $c_7 \dots c_{13}$ ), que representan transiciones entre ondas y cambios morfológicos durante su respuesta temporal [71]. Este método produjo una sensibilidad y especificidad superior al 80 %, permitiendo el diagnóstico temprano de glaucoma a través de la identificación de señales [71].

Jiménez y col. en 2015 [76] con el propósito de investigar la aplicación del análisis *wavelet* en señales de mfERG para el diagnóstico de glaucoma, aplicó CWT en las señales utilizando *wavelet Morlet* real. Los coeficientes obtenidos de la CWT de la escala de máxima correlación se utilizaron como entrada a una red neuronal con función de base radial que actúa como clasificador. Obteniendo una sensibilidad de 89 % y una especificidad de 84 % refleja una detección fiable de sectores sanos y glaucomatosos, mejorando así los hallazgos previos en otros proyectos relacionados a la detección de glaucoma.

En 2008 Jason y col. [77] utilizó un modelo de regresión logística dejando como

entrada al modelo los retrasos en los tiempos implícitos del mfERG con la intención de predecir áreas de lesiones retinianas en pacientes con RD dando seguimiento en un periodo de tres años durante el desarrollo de la enfermedad. Obtuvo un área bajo la curva ROC de 0.95 con una sensibilidad de 88 % y una especificidad de 98 %.

Posteriormente Harrison y col. en 2011 [78] utilizaron el modelo de regresión logística de Jason considerando el desarrollo de RD y otras variables del paciente como el sexo, duración de la diabetes, glucemia, edad y tipo de diabetes para predecir la aparición de RD en pacientes sin retinopatía previa. Con esta técnica se obtuvo un 80 % de exactitud en la predicción de áreas con lesiones retinianas.

Yapici y col. en 2021 [79] utilizó un modelo híbrido de perceptrón multicapa con algoritmo de optimización de enjambre de partículas para encontrar la mejor topología e hiperparámetros de la red. Utilizó los coeficientes de la transformada discreta Wavelet para después aplicar reducción de dimensionalidad por características estadísticas, utilizó 6 de estas características, las cuales sirvieron como entrada a la red neuronal. Su red fue relativamente simple constando de solamente una capa oculta variando entre 8 y 20 neuronas en esta capa. Con esta técnica obtuvo 95.74 % de exactitud en la clasificación para estimar niveles de obesidad utilizando señales ffERG.

## 9. Hipótesis

Un modelo basado en redes neuronales convolucionales, que utilice regiones destacadas de imágenes de escalograma obtenidas a partir de señales de electroretinograma espontáneos de campo completo, sea un marcador para valorar y monitorear la funcionalidad retiniana en la fase asintomática, con la intención de detectar la retinopatía diabética en etapas tempranas.

## **10. Objetivos**

### **10.1. Objetivo general**

Desarrollar un modelo inteligente que a partir de técnicas de inteligencia artificial sea posible identificar las características de las señales de electroretinograma espontáneos con diferentes grados de retinopatía diabética.

### **10.2. Objetivos específicos**

- Adquirir la base de datos con señales de ERG espontáneos que será utilizada para el desarrollo del proyecto.
- Diseñar y desarrollar el método de IA que sea capaz de identificar las características definidas en las señales ERG espontáneos con retinopatía diabética y clasificarlas.
- Realización de un diseño de experimentos para validación de los resultados de la investigación.

## 11. Metodología

### 11.1. Base de datos

Para el desarrollo de esta investigación se utilizó una base de datos de señales de ffERG espontáneos proporcionada por la Dra. Stéphanie Colette Thébault del Instituto de Neurobiología de la Universidad Nacional Autónoma de México. La base de datos está conformada por 278 archivos .csv con información de las señales en un rango desde 10 segundos hasta 18 minutos etiquetadas en diversas categorías. Además se incluye un archivo .xlsx dónde se encuentra información médica relevante de cada paciente, dicha información es anónima y no presenta datos personales.

Las categorías bajo las cuales están clasificadas las señales se muestran en la Tabla 4.

<b>Categoría</b>	<b>Número de señales</b>
RD MODIFICADA POR FOTOCOAGULACIÓN	1
RDNP	13
RDNP LEVE	8
RDNP MODERADA	16
RDNP SEVERA	5
RDP	15
RDP AVANZADA	2
SIN RD	218

Tabla 4: Categorías originales señales ffERG espontáneas.

A simple vista se puede observar que el número de señales por cada categoría

es muy bajo para utilizarse en modelos de ML o DL, por lo cual se decidió separar las categorías de retinopatía en dos. El primer conjunto engloba cualquier señal con algún grado de retinopatía (RDNP, RDNP LEVE, RDNP MODERADA, RDNP SEVERA, RDP, RDP AVANZADA y RD MODIFICADA POR FOTOCOAGULACIÓN) el otro grupo incluye las señales categorizadas como sano o SIN RD. En la Tabla 5 se muestra el número de señales resultante tras la reagrupación.

<b>Categoría</b>	<b>Número de señales</b>
RD	60
SIN RD	218

Tabla 5: Nuevas categorías señales ffERG espontáneas.

El número de datos en cada señal es variable dependiendo del tiempo de duración de cada una, algunos archivos incluso se encuentran vacíos por lo que no son útiles para los modelos, otros contienen muy poca información casi por debajo de los 10 segundos de duración. Se decidió verificar la duración de las señales en un histograma para identificar la cantidad de señales que existen entre ciertos rangos de duración, al ver que solamente existen cinco por debajo de 104,86 segundos se decidió omitir estas señales ya que no parecen aportar suficiente información relevante para los modelos.

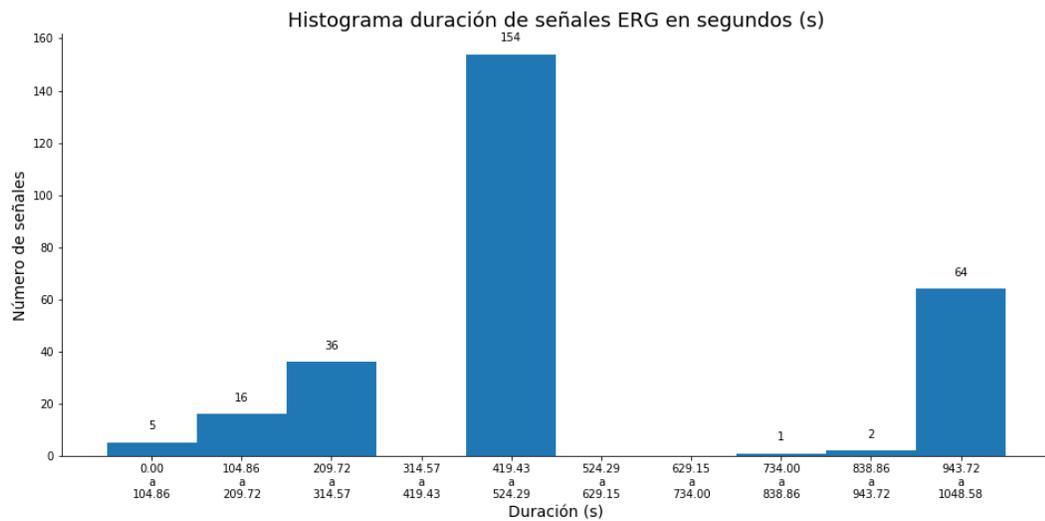


Figura 15: Histograma de tiempos de duración de señales ffERG espontáneas en segundos..

Para facilitar el manejo de la información se concentraron todas las señales en un solo archivo .csv dónde se indica el id del paciente y la categoría en formato binario (Sin RD: 0 y RD: 1). En cuanto a los datos de las señales se optó por aplicar un proceso de normalización ya que existe una diferencia bastante notable en la amplitud de cada una. La normalización es una técnica de escalado, mapeo o preprocesamiento de datos dónde se pretende imponer un rango de valores estandarizado a partir de un rango ya existente [80, 81]. Aplicar estas técnicas son muy útiles para tareas de clasificación y predicción [80].

## Muestras ERG menores a 116s

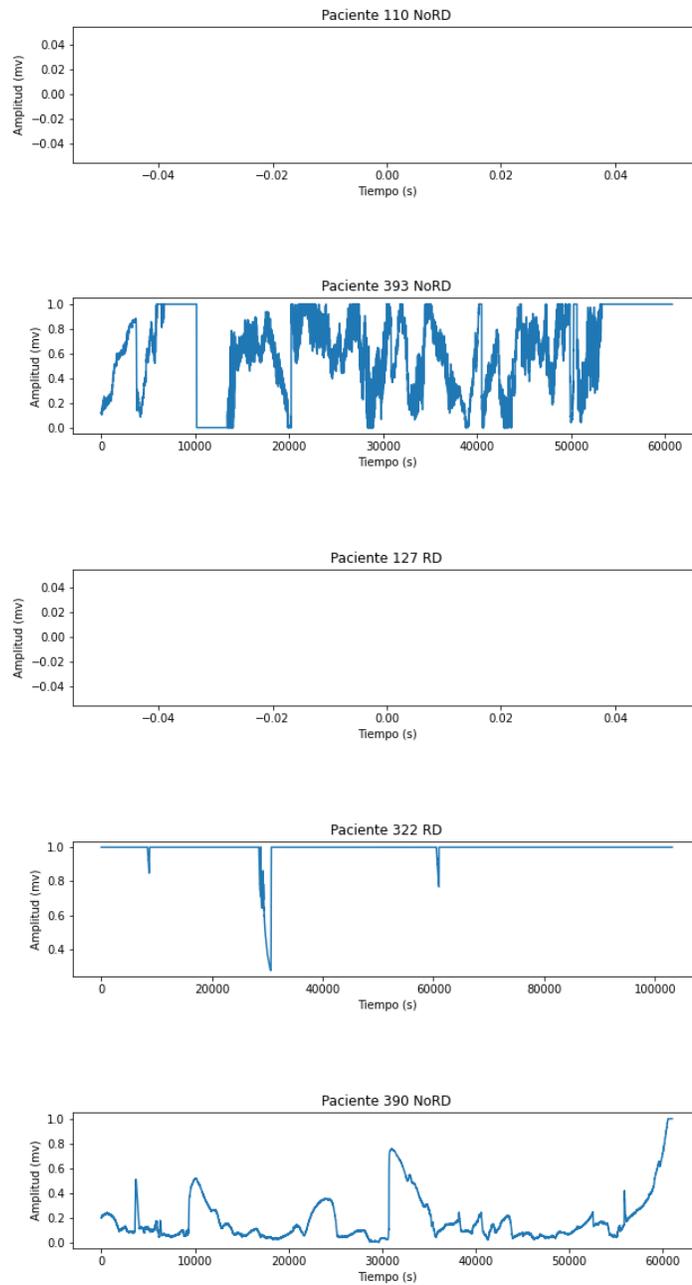


Figura 16: Señales ffERG espontáneas omitidas para el estudio.

Existen varias técnicas de normalización entre las que encontramos como la *Min-Max*, *Z-score* o *Decimal*. Para este caso se aplicó la normalización *Min-Max*,

una técnica de normalización muy sencilla dónde se ajustan los datos en un límite predefinido [80]. Esta técnica de normalización está definida por la siguiente fórmula:

$$A' = \frac{A - \min(A)}{\max(A) - \min(A)} * (D - C) + C \quad (13)$$

Dónde  $A'$  contiene los datos normalizados con la técnica *Min-Max*,  $C$  es el límite mínimo definido,  $D$  es el límite máximo definido y  $A$  es el conjunto de datos original [80]. Esto permite mantener las señales en el mismo rango de amplitud como se puede observar en la Figura 17, dónde las señales de la izquierda son las señales con sus valores de amplitud originales y las de la derecha muestran la misma señal pero normalizada en el rango de 0 a 1.

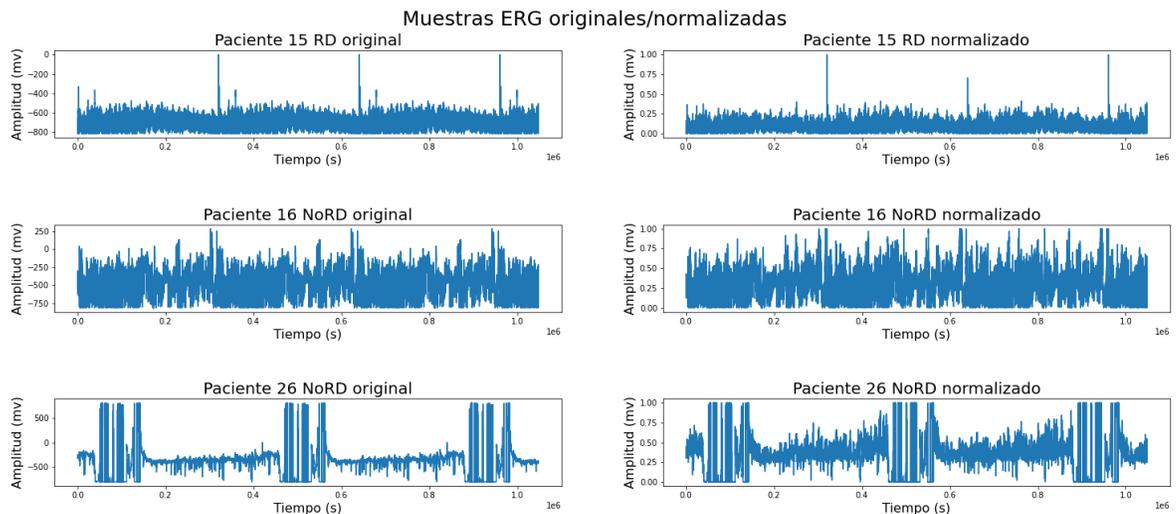


Figura 17: Muestras de señales ffERG espontáneas originales y normalizadas respectivamente

Una vez que las señales fueron normalizadas se concentró toda la información en un archivo `.csv` dónde se almacenó la información de cada señal, el id del paciente al que pertenece y su categoría correspondiente. Todo esto se realizó con la intención de facilitar el acceso a la información de cada señal, en la Tabla 6 se

observa una muestra de la estructura de este archivo.

	pac	cat	2	3	4	5	6	7	8	9	...	1048565	1048566	1048567	1048568	1048569	1048570	1048571	1048572	1048573	1048574	
57	110.0	0.0	0.397400	0.394649	0.398199	0.405759	0.414982	0.423541	0.430914	0.436559	...	NaN	NaN									
102	131.0	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	NaN	NaN									
136	138.0	1.0	0.205960	0.206280	0.206067	0.206274	0.205771	0.205604	0.205462	0.205049	...	NaN	NaN									
158	202.0	0.0	0.212701	0.213081	0.213540	0.214083	0.213841	0.213589	0.213113	0.212399	...	NaN	NaN									
182	302.0	0.0	0.205018	0.205246	0.204726	0.204485	0.204427	0.204353	0.204223	0.203790	...	NaN	NaN									
187	298.0	0.0	0.206706	0.207123	0.207054	0.206463	0.205879	0.205789	0.206068	0.205965	...	NaN	NaN									
219	446.0	1.0	0.265368	0.265738	0.265431	0.265175	0.264690	0.264421	0.264177	0.263859	...	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
231	334.0	1.0	0.196465	0.196427	0.195005	0.193734	0.194361	0.195556	0.194814	0.193964	...	NaN	NaN									
233	340.0	0.0	0.203846	0.204309	0.203863	0.203183	0.202505	0.202521	0.202275	0.202014	...	NaN	NaN									
236	344.0	0.0	0.204474	0.204368	0.203886	0.203617	0.203576	0.203865	0.203868	0.203743	...	NaN	NaN									

Tabla 6: Estructura de dataset final de señales ffERG espontáneas.

## 11.2. Escalogramas

En este proyecto se consideró trabajar con imágenes de escalograma, las cuales son una representación única de cada señal. Un escalograma es el valor absoluto de la CWT de una señal representada en función de su tiempo-frecuencia. Es una técnica útil para analizar señales del mundo real con características que ocurren en diferentes escalas.

Para la creación de los escalogramas de las señales ERG se aplicó la CWT con la familia Morlet compleja (`'cmor1-1.5'`) a 660 escalas. Se realizaron diferentes pruebas considerando la longitud de las señales.

1. Un solo escalograma por cada señal considerando el tiempo total de cada señal.
2. Un escalograma por cada señal con el tiempo truncado a 1 minuto.
3. Varios escalogramas por cada señal considerando una ventana de tiempo en función de la señal más pequeña.

En el primer caso se utilizó la librería “Librosa” en Python, esta librería es especialmente utilizada para el procesamiento de señales de audio, sin embargo, debido

a que posee un excelente manejo de memoria para el cálculo de escalogramas, se decidió utilizarla para realizar algunas pruebas. Con esta técnica se consideró el 100 % del tiempo de duración de cada señal. Un ejemplo de las imágenes obtenidas con esta técnica puede visualizarse en la Figura 18. A continuación se presenta el pseudocódigo para la generación de imágenes de escalograma utilizando la librería *Librosa*.

---

**ALGORITMO 1**

Pseudocódigo para la extracción de escalogramas *Librosa*.

---

```
while sizeof(signalsArray) != 0 do
    paciente ← signal.idPaciente
    categoria ← signal.Categoria
    signal ← signal.data
    if categoria = 0 then
        clase = NoRD
    else
        clase = RD
    nombreImagen ← path + clase + paciente
    coeficientes ← librosa.scaleogram(signal)
    decibels ← librosa.hzTodb(coeficientes)
    escalograma ← librosa.display.specshow(decibels)
    escalograma.save(nombreImagen)
```

---

La segunda técnica fue la aplicación de CWT por medio de la librería de *Py-Wavelets* la cual aplica el procesamiento de CWT a cada señal de acuerdo a los parámetros especificados de escala y *wavelet* madre (*'cmor1-1.5'*). Para estos escalogramas se truncó la duración de las señales a un minuto, esto debido a la complejidad computacional que representa la implementación. En la Figura 19 se pueden apreciar las muestras de algunas señales y su escalograma equivalente resultante. En seguida se muestra el pseudocódigo para esta técnica.

## Muestras ERG Señales/Escalogramas librosa

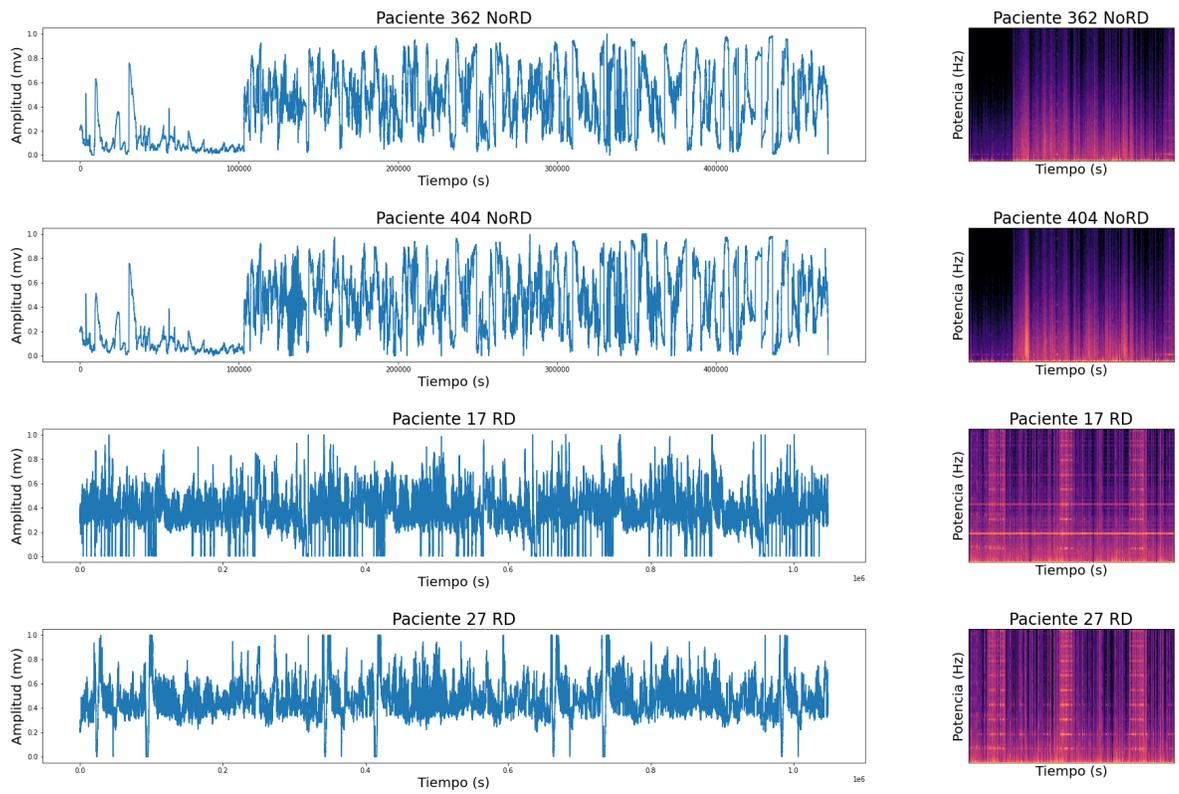


Figura 18: Muestras de señales ffERG espontáneas y escalogramas respectivos obtenidos con la librería Librosa.

**ALGORITMO 2****Pseudocódigo para la extracción de escalogramas CWT.**

---

```
minutos ← 1
escalas ← 660
limite ← minutos * 6000
while sizeof(signalsArray) ≠ 0 do
    paciente ← signal.idPaciente
    categoria ← signal.Categoria
    signal ← signal.data
    if categoria = 0 then
        clase = NoRD
    else
        clase = RD
    firstNull ← signalPrimerNaN
    if firstNull > 0 then
        signal = signal[firstNull]
    else
        signal = signal[limite]
    nombreImagen ← path + clase + paciente
    coeficientes ← CWT.Morlet(signal)
    escalograma ← scaleogram(coeficientes)
    escalograma.save(nombreImagen)
```

---

## Muestras ERG Señales/Escalogramas CWT

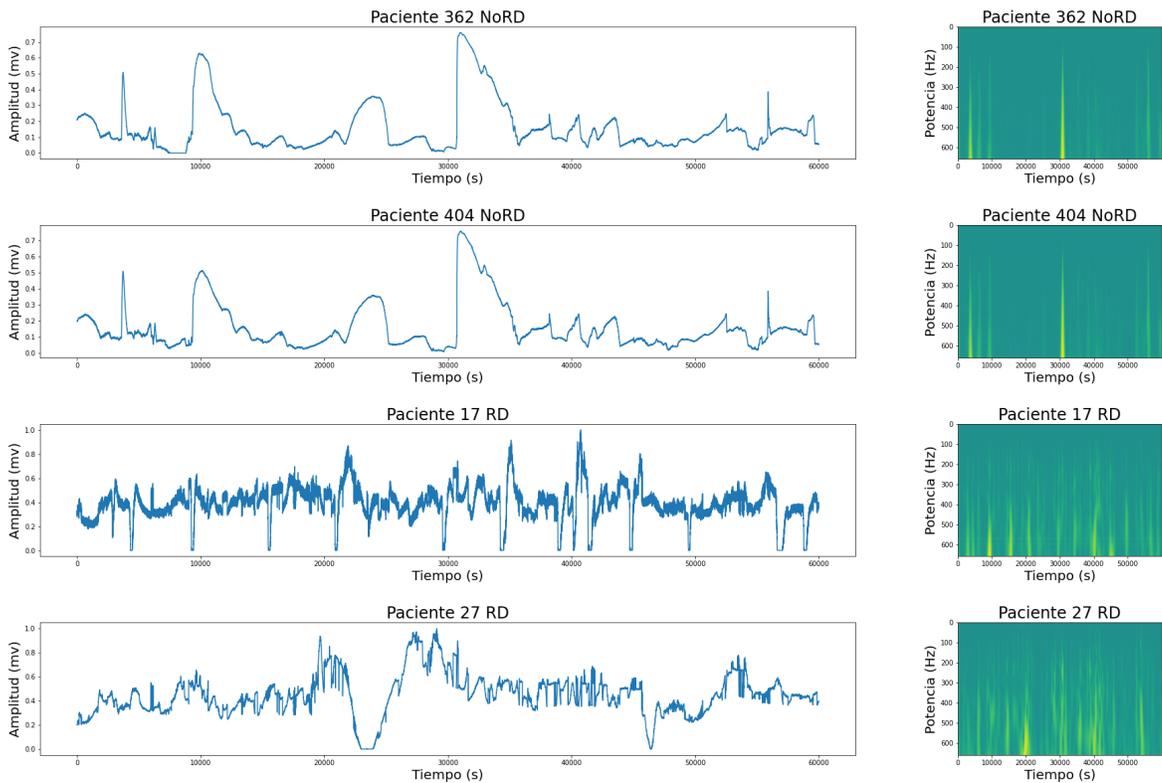


Figura 19: Muestras de señales ffERG espontáneas y escalogramas respectivos obtenidos con CWT.

Con la tercera técnica se realizó una comparación entre la señal de duración más pequeña (1,95 minutos) y la más grande (17,4 minutos). La señal más grande se dividió en secciones con ventanas del tamaño de la señal más pequeña. Esto da como resultado un máximo de 9 escalogramas en el caso de la señal más grande (Figura 20) y un solo escalograma con la señal más pequeña (Figura 21). Con esta técnica se utilizaron los escalogramas en escala de grises. En seguida se muestra el pseudocódigo para la extracción de las secciones y el escalograma de cada sección:

**ALGORITMO 3**

Pseudocódigo escalogramas secciones de señal CWT.

---

```

escalas ← 660
minimo ← sizeOf(signal.minima)
while sizeof(signalsArray) ≠ 0 do
    paciente ← signal.idPaciente
    categoria ← signal.Categoria
    firstNull ← signalPrimerNaN
    if firstNull > 0 then
        signal = signal[firstNull]
    else
        signal = signal.data
    if categoria = 0 then
        clase = NoRD
    else
        clase = RD
    nSecciones ← sizeOf(signal)/sizeOf(signal.minima)
    if nSecciones > 1 then
        i ← 1
        while nSecciones ≠ 0 do
            puntosDeCorte ← i * minimo
        else
            puntosDeCorte ← nSecciones * minimo
        secciones ← cortarSignal(puntosDeCorte)
        while secciones > 0 do
            nombreImagen ← path + clase + paciente + seccion
            coeficientes ← CWT.Morlet(seccion)
            escalograma ← scaleogram.gray(coeficientes)
            escalograma.save(nombreImagen)

```

---

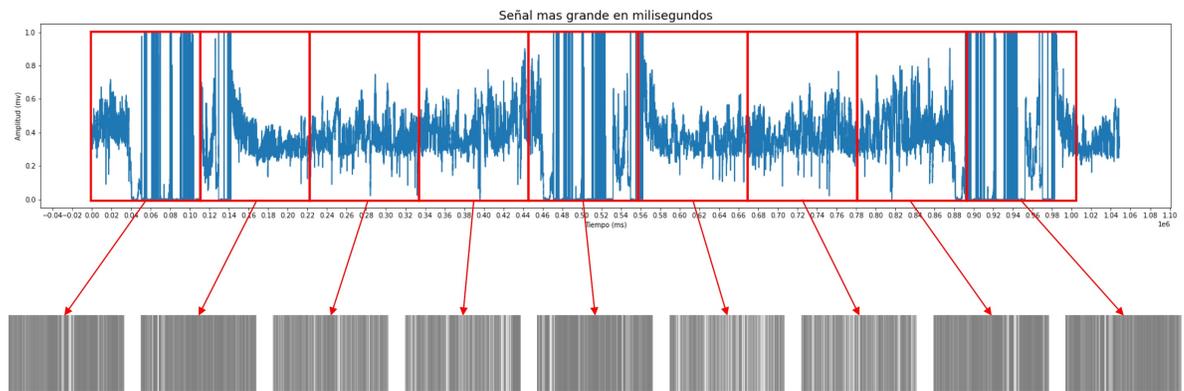


Figura 20: Extracción de escalogramas de secciones de la señal más grande de acuerdo a la ventana de tiempo de la señal más pequeña.

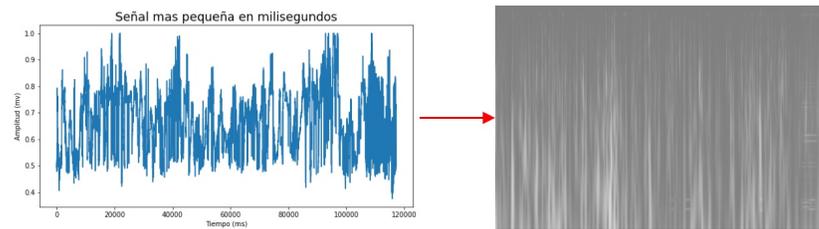


Figura 21: Extracción de escalograma de la señal más pequeña.

### 11.3. Modelos

Para este proyecto se implementaron tres topologías de CNN a manera de comparativa. CNN AlexNet, HyperResNet y una implementación propia de una red paralela a la cual se le denominó ChannelNet.

#### 11.3.1. AlexNet Modificada

La CNN AlexNet se implementó con sus parámetros originales tan solo adaptando el tamaño de la capa de entrada, la capa de salida, la función de pérdida (`categorical_crossentropy`) y los filtros intermedios, con la intención de que acepte los escalogramas como entrada y devuelva como resultado una respuesta binaria (NoRD: 0, RD: 1). La red AlexNet trabajó con los escalogramas generados con la

librería Librosa, las imágenes se generaron con un tamaño de  $300 \times 300 \times 3$ , se consideró el 80 % de los datos como entrenamiento y 20 % como pruebas.

### Topología CNN AlexNet ajustada

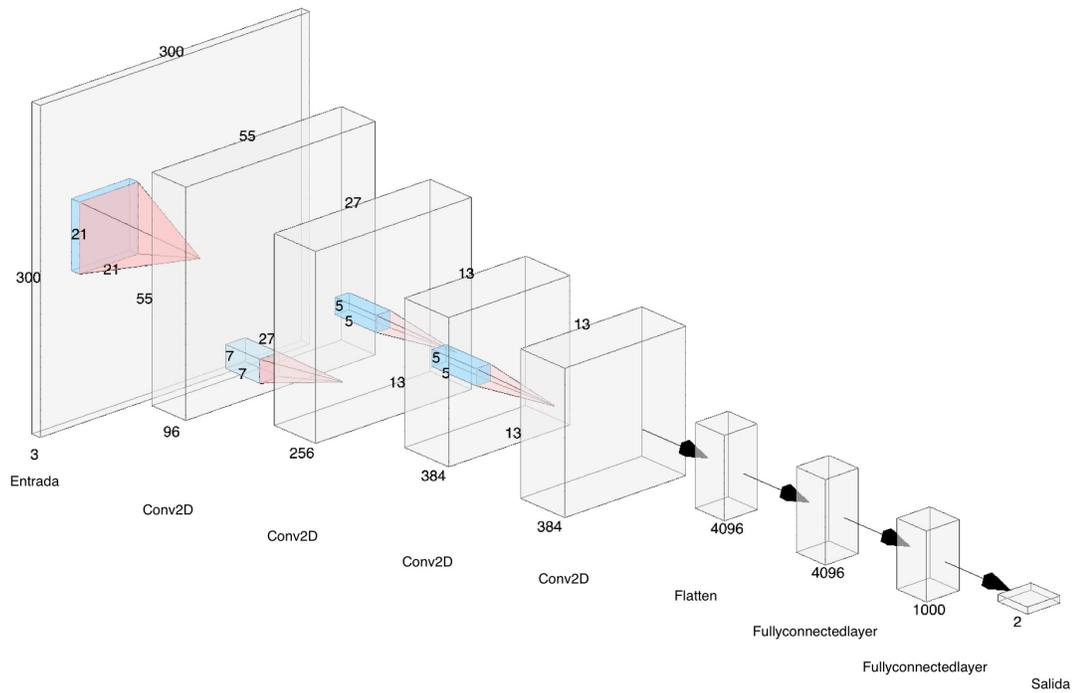


Figura 22: Modelo propuesto CNN AlexNet modificado

#### 11.3.2. HyperResNet AutoML

A través de la librería “KerasTuner” se ajustó el modelo de HyperResNet, gracias al proceso de AutoML que ofrece la librería se encontraron los hiperparámetros adecuados para el conjunto de datos. Este modelo trabaja con los escalogramas generados por CWT en un tamaño de  $56 \times 128 \times 3$ , se consideró el 80 % como datos de entrenamiento y el 20 % como datos de pruebas (Figura 23).

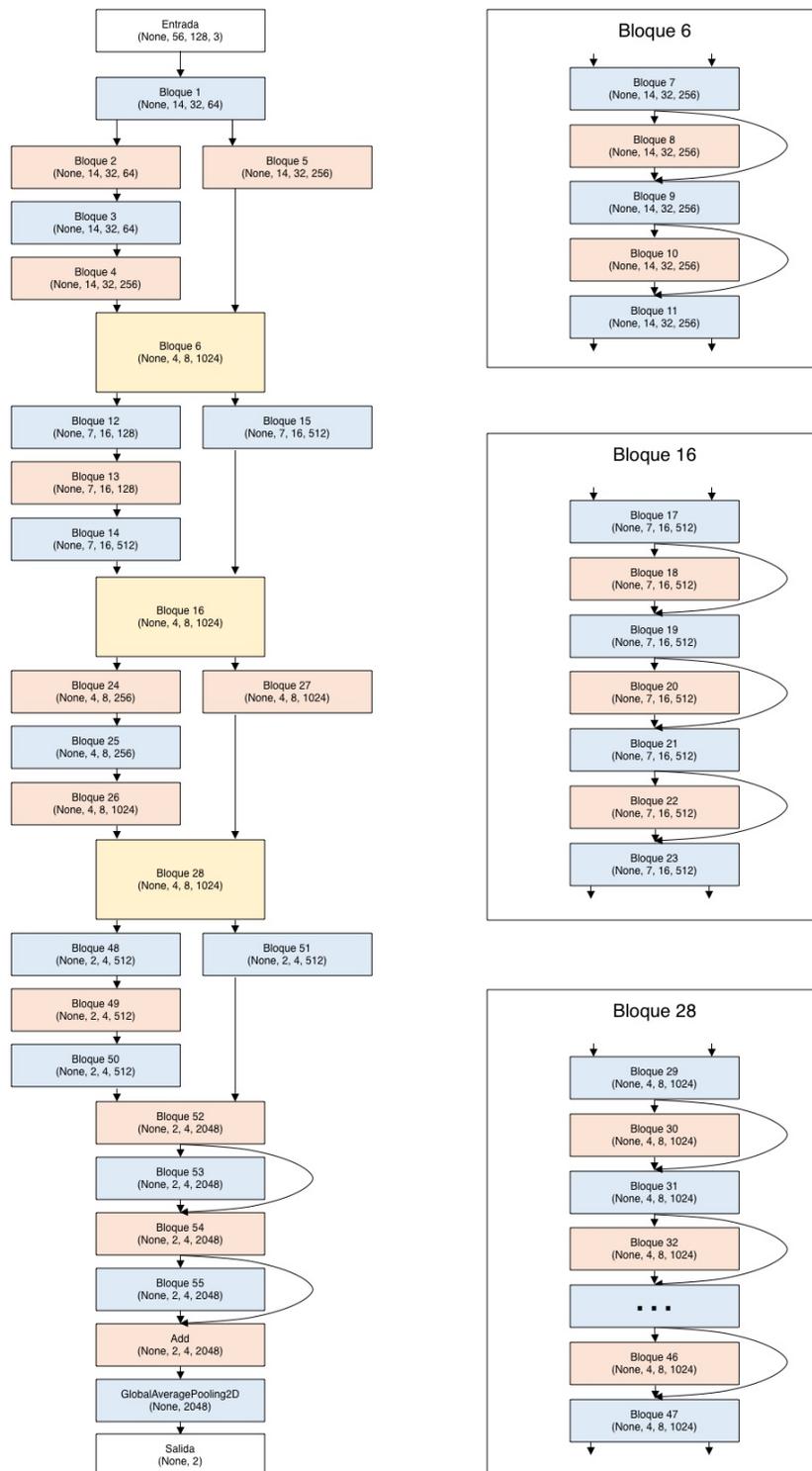


Figura 23: Modelo propuesto HyperResNet obtenido a través de la herramienta KerasTuner.

### 11.3.3. CNN Paralela

El tercer modelo implementado está basado en la idea de que las redes pequeñas con los hiperparámetros correctos también funcionan en el análisis de señales [79] y pueden llegar a competir con redes más complejas y profundas. Partiendo de una red neuronal pequeña con tan solo dos capas de convolución y con el objetivo de encontrar la mejor combinación de exactitud, maximización de los datos y complejidad computacional, se propuso un modelo de red convolucional de capas paralelas. La intención de este modelo de capas paralelas es considerar la mayor información posible de la señal ERG, aceptando desde la señal más pequeña hasta la señal más grande.

El modelo está construido a partir de nueve capas paralelas, cada capa acepta un escalograma de tamaño  $432 \times 288 \times 1$  el cual corresponde a una sección de 1,95 minutos de la señal (Figura 21). Cada una de las imágenes pertenecientes a una sección de la señal entra por cada capa de la red, el modelo procesa de manera paralela cada una de las imágenes y al finalizar realiza una predicción de la clase a la que pertenece el conjunto de imágenes.

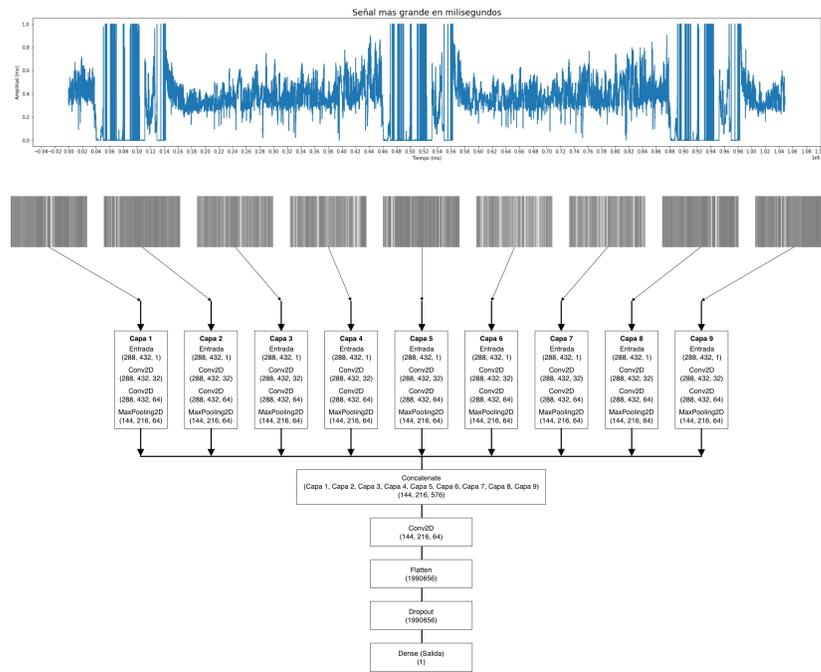


Figura 24: Representación de entrada de los escalogramas de la señal más larga a la CNN Paralela.

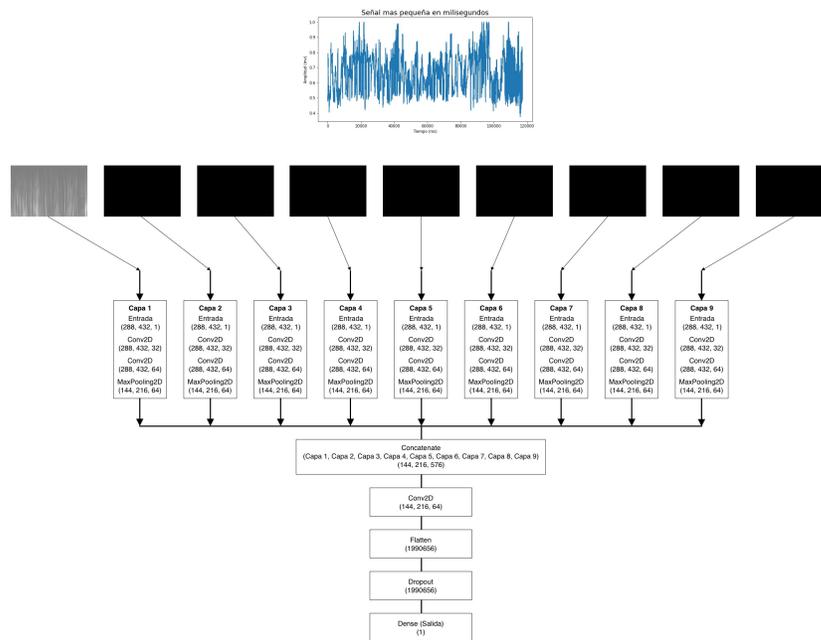


Figura 25: Representación de entrada del escalograma de la señal más pequeña a la CNN Paralela.

Para el caso de las señales más pequeñas (las cuales generan menos escalogramas) el resto de las capas se rellena con imágenes de valor cero (color negro).

En la Figura 26 se muestra el diagrama completo de metodología, para encontrar los hiperparámetros adecuados de los modelos se pasó por una serie de pruebas y ajustes con los resultados de los entrenamientos.

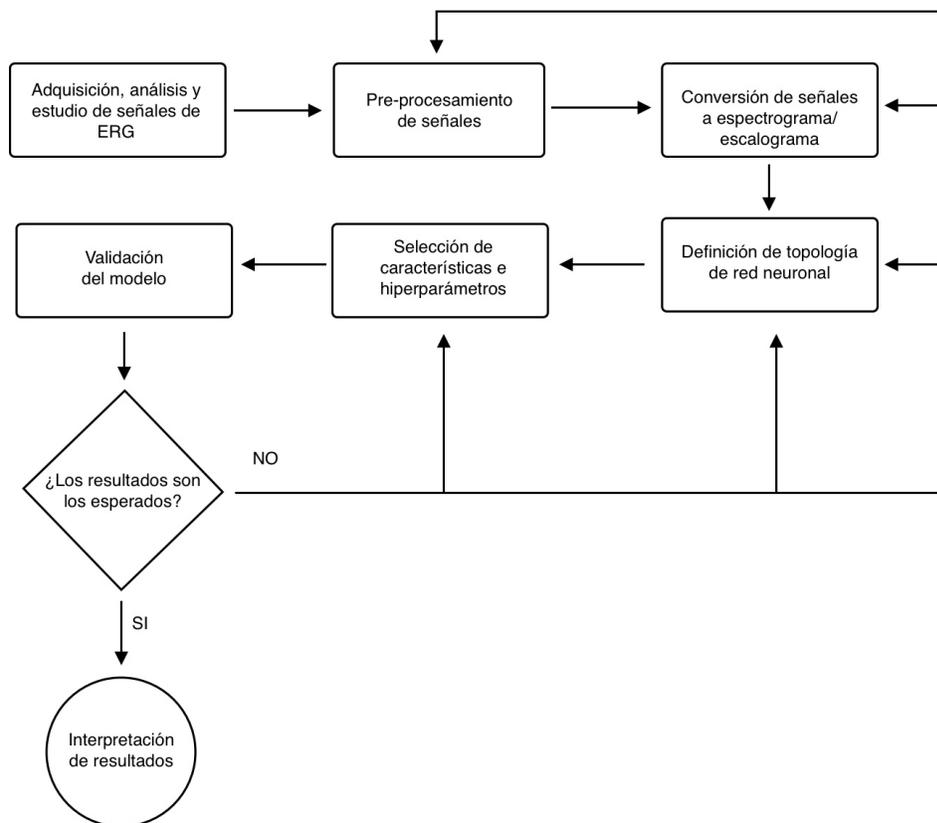


Figura 26: Diagrama de metodología.

Durante los entrenamientos se implementaron diferentes técnicas y modelos, con base en los resultados que se obtenían de cada ejecución se ajustaban los hiperparámetros (*epoch*, *strides*, *filters*, *loss function*, etc.), la topología (*depth*, *layers*, *model types*, *architecture*) y los escalogramas (escalas, librerías, técnicas *wavelet*, familias *wavelets*, tiempo de la señal).

La base de datos de señales ERG ha estado en constante alimentación, a medida que el tamaño de la base de datos fue creciendo, los modelos comenzaron a ofrecer mejores resultados, por lo que el tamaño de los grupos (entrenamiento, validación y pruebas) que se utilizaban durante los primeros modelos fueron muy pequeños. En algunos casos se implementaron técnicas para balanceo de datos como el aumento artificial de señales por DWT utilizando familias como *Symlet* y *Daubechies*.

En otros casos más particulares se utilizaron algunas técnicas de balanceo más modernas como el *Stratified K-fold*, una técnica para evaluación de rendimiento de modelos basado en *folds* balanceados. Esta técnica permite mantener los *folds* con el mismo número de datos en cada clase tomando como base el tamaño de la clase más pequeña, la selección de datos para la creación de los *folds* es de manera aleatoria por lo tanto mantiene la seguridad de que los datos serán variados en cada *fold*. Dentro del entrenamiento de cada uno se puede especificar el tamaño de los conjuntos de entrenamiento, pruebas y validación con los porcentajes que se deseen.

De manera general se siguen los pasos en la metodología propuesta, regresando a cada uno de los pasos anteriores y ejecutándolos nuevamente desde ese punto tras evaluar e interpretar los resultados de los modelos.

## 12. Resultados

Para todos los modelos se utilizó el mismo porcentaje del conjunto de datos, 80 % para entrenamiento y 20 % para pruebas. Se implementaron las mismas métricas para la evaluación de los modelos, exactitud, precisión, sensibilidad y *F1-Score*.

Todas estas métricas se pueden obtener a partir de la matriz de confusión de cada uno de los modelos, esta matriz evalúa el desempeño general de los modelos que realizan tareas de clasificación.

En la Figura 27 se muestra el resultado del entrenamiento del modelo AlexNet, se puede observar cómo poco a poco mejora la pérdida del modelo, sin embargo la exactitud se mantiene baja en las validaciones durante el entrenamiento.

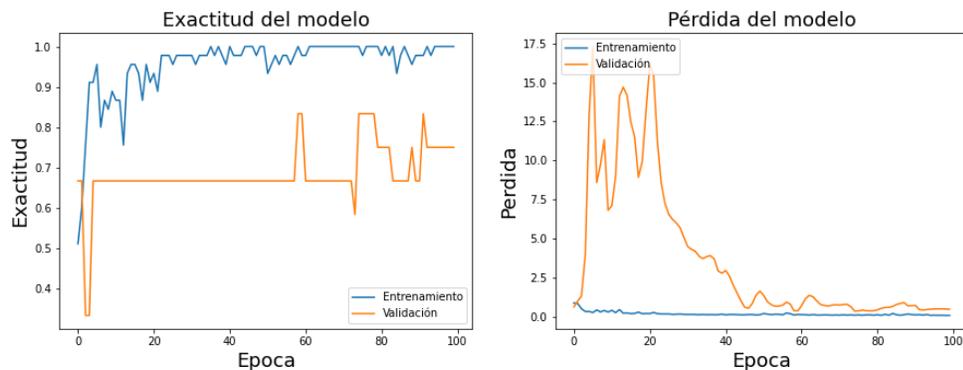


Figura 27: Gráficas de entrenamiento del modelo AlexNet, exactitud y pérdida respectivamente.

Las gráficas de entrenamiento de la HyperResNet muestran un comportamiento interesante ya que desde la primera época logra alcanzar exactitudes de entre el 65 % y el 70 %. Sin embargo a lo largo de la ejecución se llegan a tener bajas de hasta el 30 %. De manera muy general el modelo mantiene exactitudes cercanas al

conjunto de entrenamiento con el conjunto de validación. La pérdida es una situación similar comenzando con no más del 70% y teniendo picos de dispersión de más del 100% en algunas ocasiones.

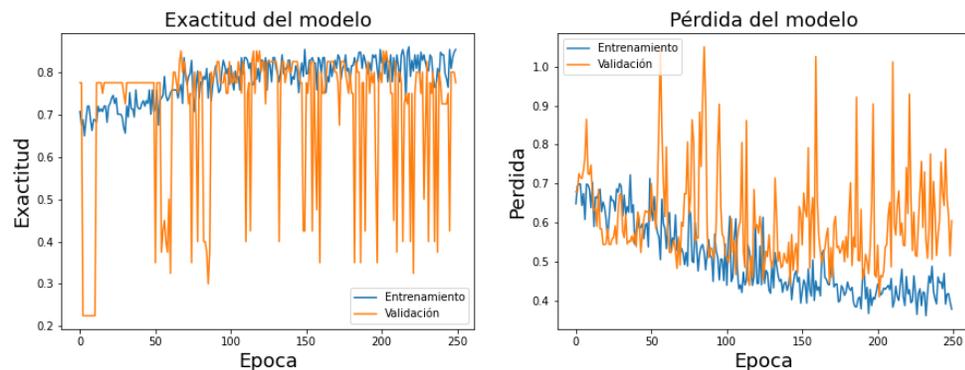


Figura 28: Gráficas de entrenamiento del modelo HyperResNet, exactitud y pérdida respectivamente.

Para el caso de la red paralela tanto la exactitud como la pérdida van muy de la mano entre el entrenamiento y la validación, a partir de la sexagésima época se puede observar como la exactitud no incrementa y la pérdida comienza a separarse gradualmente. También se puede observar cómo desde el inicio se encuentran exactitudes de más del 75% pero comienza a estancarse sin sobrepasar 85%, la pérdida se dispersa poco a poco en relación a la gráfica de exactitud, cuando comienza a estancarse la exactitud la diferencia en la pérdida comienza a ser cada vez mayor.

En la Tabla 7 se muestran los resultados obtenidos y en la Figura 30 la matriz de confusión de cada uno de los modelos.

Es evidente que la red AlexNet obtuvo los mejores resultados, una exactitud de 86.7% y 86% en cuanto a precisión, sensibilidad y media armónica. Es un tipo de CNN mejor estructurada y estable por lo que es lógico que sus resultados sean al-

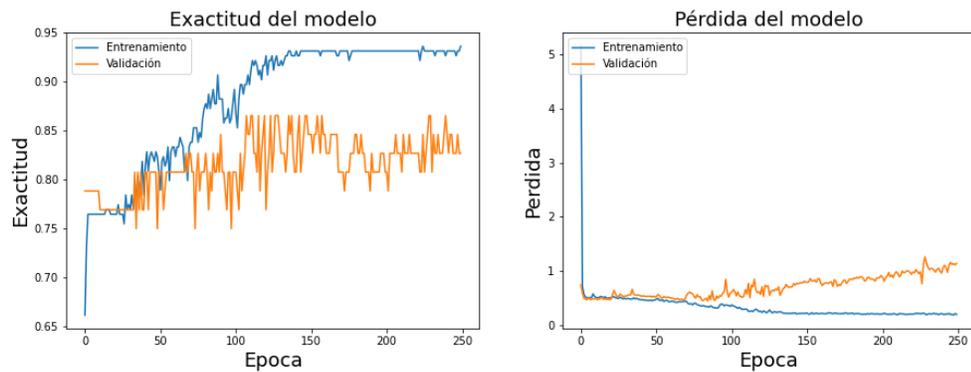


Figura 29: Gráficas de entrenamiento del modelo CNN Paralela, exactitud y pérdida respectivamente.

Métrica	Porcentaje AlexNet	Porcentaje HyperResNet	Porcentaje CNN Paralela
Exactitud	86.70 %	78 %	79 %
Precisión	86 %	54 %	68 %
Sensibilidad	86 %	52 %	65 %
F1-Score	86 %	51 %	62 %

Tabla 7: Comparativa de métricas de los modelos implementados.

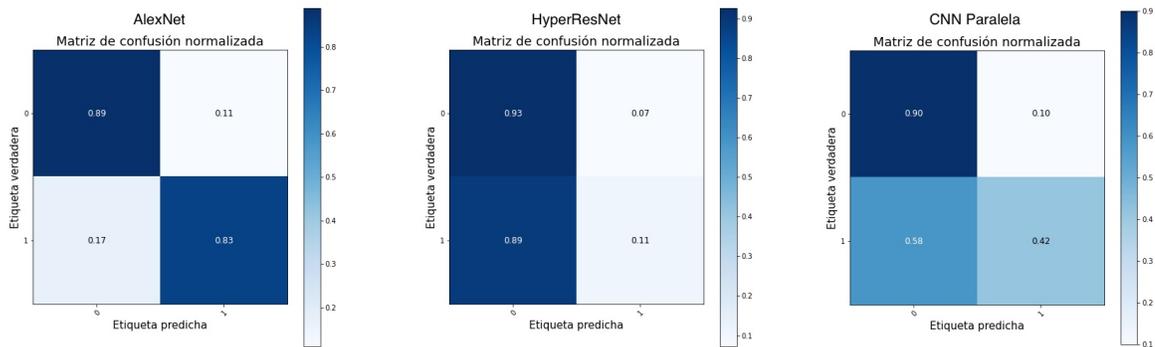


Figura 30: Comparativa de matrices de confusión de los modelos implementados.

tos. En la matriz de confusión se pueden observar algunos falsos positivos y falsos negativos lo cual indica que la red puede llegar a confundir algunos patrones en las imágenes de escalograma.

En cuanto a la HyperResNet se encuentran un alto número de falsos negativos lo cual indica que la CNN muestra un sobre entrenamiento clasificando imágenes con RD como NoRD, confunde bastante los patrones de retinopatía diabética y parece no ser tan capaz de identificarlos correctamente, esta CNN resultó ser una red bastante profunda. Al ser pocos los datos que se utilizan para el entrenamiento y a pesar de utilizar técnicas de balanceo, son pocas las muestras utilizadas en cada clase lo que afecta por mucho el proceso del entrenamiento.

La CNN paralela parece estar en un punto intermedio entre la HyperResNet y la AlexNet, mantiene también un número alto de falsos negativos pero parece reconocer los patrones de RD mejor que la HyperResNet, sin embargo no logra alcanzar la exactitud de la AlexNet, le cuesta identificar los patrones de RD confundiéndolos con NoRD. Al igual que en el caso anterior el número de muestras es pequeño por cada clase pero parece responder un poco mejor a la técnica de balanceo de datos por *Stratified K-fold*.

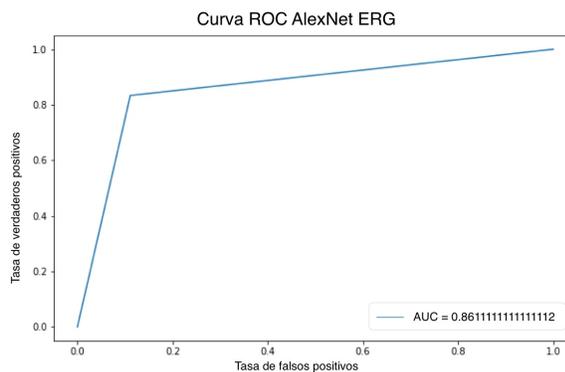


Figura 31: Curvas de análisis ROC modelo Alexnet.

En las Figuras 31 a 33 se muestran las curvas ROC de cada uno de los modelos al ser evaluados con el conjunto de pruebas. Nuevamente se observa como la AlexNet es el modelo que mejor se desempeñó en la tarea de clasificación obteniendo un 86 % de área bajo la curva considerado como una prueba buena.

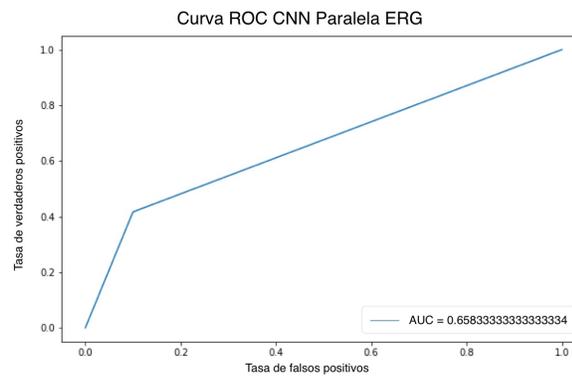


Figura 32: Curvas de análisis ROC CNN Paralela.

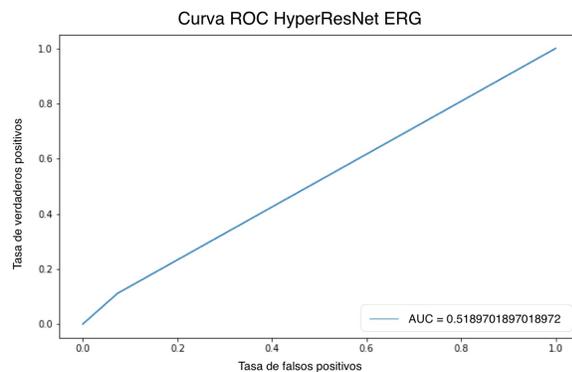


Figura 33: Curvas de análisis ROC HyperResNet

El modelo HyperResNet muestra una evaluación bastante mala alcanzando solamente un 51.8 % lo cual se considera como una prueba sin capacidad discriminativa, esto se interpreta como si el modelo decidiera al azar la clasificación tal cual como lanzar una moneda.

El último caso con la CNN Paralela se considera un test regular alcanzando solamente un 65 % en el área bajo la curva, no ofrece resultados muy relevantes con las técnicas aplicadas, tras ver estos resultados es lógico que la CNN llega a tener problemas para el reconocimiento de ciertos patrones y esto puede ser causado por los canales cuya entrada es completamente imágenes en negro.



## 13. Conclusiones

En resumen, esta investigación se centra en la comparativa de tres modelos de CNN. Modificación a CNN Alexnet, ResNet con AutoML (HyperResNet) y una CNN con capas paralelas. Cada una de ellas se desempeñó de manera diferentes con el mismo conjunto de datos. La red Alexnet mostró el mejor desempeño obteniendo no solo la exactitud más alta con 86.7% si no que también demuestra tener una mejor sensibilidad para detectar las características de los escalogramas. La CNN Paralela parece encontrarse en un punto intermedio logrando una exactitud de 79% con una sensibilidad baja de 65% esto ofrece una predicción regular en el análisis de la curva ROC. La HyperResNet fue modelo con el peor desempeño, demuestra que no regulariza correctamente al momento del entrenamiento ocasionando una confusión muy alta identificando los patrones de RD, su predicción es completamente aleatoria obteniendo un área bajo la curva ROC de 0.5.

Cada CNN dependiendo de sus hiperparámetros se desempeñará diferente para cada conjunto de datos. Son múltiples las combinaciones de técnicas de preprocesamiento de datos con diferentes topologías, la combinación adecuada de cantidad de datos, calidad de datos, topología, hiperparámetros o incluso el tipo de red neuronal, pueden llegar a lograr distinguir las características relevantes de los escalogramas. Todo lo anterior demuestra que indagar y probar con una gran diversidad de modelos y técnicas de preprocesamiento de señales es un camino para encontrar el mejor compromiso entre el desempeño de la predicción, maximización de los datos y recursos computacionales.

Se puede observar que los escalogramas de señales ffERG son útiles para mostrar las características relevantes para la detección de enfermedades como la RD y que modelos de datos basados en DL o ML pueden identificar dichas características para clasificarlas y así entregar un veredicto que pueden ser de utilidad al momento de diagnosticar a un paciente.

Uno de los factores más importantes es el preprocesamiento de los datos, ya que es fácil alterar los datos y caer en situaciones que serían fatales para los modelos. Algunas de ellas son la limpieza de los datos de las señales y la conversión a imágenes de escalograma. La utilización de familias *Wavelet* y escalas debe ser precisa ya que una mala elección podría alterar la información de la señal y perder patrones que serían de gran utilidad al momento del entrenamiento.

Uno de los objetivos principales de esta investigación es identificar los patrones de ffERG desde las etapas más tempranas de la RD, sin embargo es posible se requiera un número mucho más alto de registros en la base de datos. Para lograr identificar a detalle las características incluso más pequeñas. Y que los modelos sean capaces de diferenciar la RDNP mínima, leve, moderada, severa y muy severa, para el caso de RDP moderada, alto riesgo y avanzada. Con los modelos aquí propuestos se necesitarán al menos unos cientos de muestras por cada categoría para lograr llegar a una clasificación confiable.

Como se puede ver las técnicas de balanceo de datos ayudan a la exactitud de los modelos, sin embargo, por otro lado parece afectar la sensibilidad, en cualquier caso el aumento de datos funciona pero en porcentajes bajos, además esto no asegura del todo la correcta identificación de los patrones necesarios en las señales, siempre es mucho mejor tener muestras reales.

Finalmente, los modelos de CNN utilizando imágenes de escalograma pueden ser una excelente opción para el análisis de señales de ERG en ambientes donde se encuentre una gran cantidad de datos. Esto por la razón de que será mucho más fácil y rápido analizar una imagen de escalograma de un tamaño de 300x300 sin perder los patrones importantes que la caracterizan, que evaluar una señal de bioeléctrica conformada por millones de datos.



### **13.0.1. Trabajo futuro**

Con este trabajo de investigación se realizó un análisis y estudio de las señales ERG espontáneas con técnicas modernas de IA. La base de datos de ERG sigue en constante crecimiento, pues la intención es que llegue a tener un tamaño adecuado para que los modelos sean capaces de distinguir no solo la RD de otras enfermedades si no entre diferentes grados de la enfermedad, prestando especial atención en las etapas tempranas.

Se buscará realizar algunas mejoras a los modelos y la extracción de las imágenes de escalograma para incrementar los resultados y encontrar la mejor relación entre exactitud, aprovechamiento de los datos y los recursos computacionales. Aplicar diferentes técnicas a los escalogramas y mejorar su calidad podría ayudar mucho a lograr este objetivo, así mismo intentar un aumento de datos reduciendo el tiempo de las señales y tomar secciones de las señales más grandes para lograr un balanceo más adecuado podría ser una técnica que ayude al rendimiento de los modelos.

El presente trabajo de investigación nos ayudará a sembrar las bases para el desarrollo de un sistema inteligente con equipo portátil, confiable y de bajo costo que pueda estar al alcance de cualquier sector poblacional para reducir la brecha en el tratamiento y diagnóstico fácil, rápido y seguro de la RD.

## Referencias

- [1] D. N. M. García, V. M. D. Flores, J. L. H. López, E. I. A. Jiménez, and E. F. V. Acurio, “Avances de la inteligencia artificial en salud,” *Dominio de las Ciencias*, vol. 5, no. 3, pp. 603–613, 2019.
- [2] D. F. Edreira, J. L. Blanco, B. C. Galdo, V. M. M. García, and A. Pazos, “La inteligencia artificial como herramienta para la gestión y explotación de datos, informaciones y conocimientos biomédicos en entornos “big data” en la nube,” *I+ S: Revista de la Sociedad Española de Informática y Salud*, no. 143, pp. 30–39, 2021.
- [3] R. Arencibia Jorge, L. Perezleo Solorzano, G. Achón Veloz, and J. A. Araújo Ruiz, “La informática biomédica desde una perspectiva bibliométrica,” *Acimed*, vol. 9, no. 3, pp. 201–208, 2001.
- [4] S. Thébault, “Pérdida de visión por diabetes: detección y tratamiento,” *Revista Digital Universitaria*, vol. 20, no. 3, pp. 1–13, 2019.
- [5] V. Flores-Peredo, V. Robredo-Torres, and F. Graue-Wiechers, “Evaluación del campo visual en pacientes diabéticos tratados con fotocoagulación panretiniana,” *Revista Mexicana de Oftalmología*, vol. 80, no. 3, pp. 116–125, 2006.
- [6] M. D. Abràmoff, P. T. Lavin, M. Birch, N. Shah, and J. C. Folk, “Pivotal trial of an autonomous ai-based diagnostic system for detection of diabetic retinopathy in primary care offices,” *NPJ digital medicine*, vol. 1, no. 1, pp. 1–8, 2018.
- [7] C. Valverde, M. Garcia, R. Hornero, and M. I. Lopez-Galvez, “Automated detection of diabetic retinopathy in retinal images,” *Indian journal of ophthalmology*, vol. 64, no. 1, p. 26, 2016.
- [8] R. Ortiz-Feregrino, S. Tovar-Arriag, J. Ramos-Arreguin, and E. Gorrostieta, “Classification of proliferative diabetic retinopathy using deep learning,” in *2019*

- IEEE Colombian Conference on Applications in Computational Intelligence (ColCACI)*. IEEE, 2019, pp. 1–6.
- [9] F. Cisneros-Guzmán, S. Tovar-Arriaga, C. Pedraza, and A. González-Gutierrez, “Classification of diabetic retinopathy based on hard exudates patterns, using images processing and svm,” in *2019 IEEE Colombian Conference on Applications in Computational Intelligence (ColCACI)*. IEEE, 2019, pp. 1–5.
- [10] G. A. Francia, C. Pedraza, M. Aceves, and S. Tovar-Arriaga, “Chaining a u-net with a residual u-net for retinal blood vessels segmentation,” *IEEE Access*, vol. 8, pp. 38 493–38 500, 2020.
- [11] F. C. Martinez and D. Galaz, “Dart inteligencia artificial para la prevención de la ceguera, entrevista a jose tomás arenas y dra. maría eugenia chadwick,” *Laboratorio de gobierno*, 2017.
- [12] M. Bhaskaranand, C. Ramachandra, S. Bhat, J. Cuadros, M. G. Nittala, S. R. Sadda, and K. Solanki, “The value of automated diabetic retinopathy screening with the eyeart system: a study of more than 100,000 consecutive encounters from people with diabetes,” *Diabetes technology & therapeutics*, vol. 21, no. 11, pp. 635–643, 2019.
- [13] E. V. Carrera, A. González, and R. Carrera, “Automated detection of diabetic retinopathy using svm,” in *2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*. IEEE, 2017, pp. 1–4.
- [14] I. Perlman, “The electroretinogram: Erg by ido perlman,” *Webvision: The Organization of the Retina*, 2015.
- [15] L. de Santiago, M. Ortiz del Castillo, E. Garcia-Martin, M. J. Rodrigo, E. M. Sánchez Morla, C. Cavaliere, B. Córdón, J. M. Miguel, A. López, and L. Boquete,

- “Empirical mode decomposition-based filter applied to multifocal electroretinograms in multiple sclerosis diagnosis,” *Sensors*, vol. 20, no. 1, p. 7, 2020.
- [16] D. Jakhar and I. Kaur, “Artificial intelligence, machine learning and deep learning: definitions and differences,” *Clinical and experimental dermatology*, vol. 45, no. 1, pp. 131–132, 2020.
- [17] M. Sewak, M. R. Karim, and P. Pujari, *Practical convolutional neural networks: implement advanced deep learning models using Python*. Packt Publishing Ltd, 2018.
- [18] J. JALOMO-CUEVAS, E. PRECIADO-MARTÍNEZ, and J. GUDIÑO-LAU, “Diseño de una red neuronal convolucional para la clasificación de señales biomédicas design of a convolutional neural network for classification of biomedical signals.”
- [19] J. D. Kelleher, *Deep learning*. MIT press, 2019.
- [20] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, “Deep learning for computer vision: A brief review,” *Computational intelligence and neuroscience*, vol. 2018, 2018.
- [21] J. Bagnato, “Convolutional neural networks: La teoría explicada en español,” *Aprende Machine Learning. Retrieved November*, vol. 19, p. 2021, 2018.
- [22] DeepAI, “Stride (machine learning),” *DeepAI*, 2019. [Online]. Available: <https://deepai.org/machine-learning-glossary-and-terms/stride>
- [23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, 2012.
- [24] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Esesn, A. A. S. Awwal, and V. K. Asari, “The history began from alex-

- net: A comprehensive survey on deep learning approaches,” *arXiv preprint arXiv:1803.01164*, 2018.
- [25] S. Targ, D. Almeida, and K. Lyman, “Resnet in resnet: Generalizing residual architectures,” *arXiv preprint arXiv:1603.08029*, 2016.
- [26] F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to forget: Continual prediction with lstm,” *Neural computation*, vol. 12, no. 10, pp. 2451–2471, 2000.
- [27] K. the Python deep learning API, “Keras documentation: Hyperresnet.” [Online]. Available: [https://keras.io/api/keras\\_tuner/hypermodels/hyper\\_resnet/](https://keras.io/api/keras_tuner/hypermodels/hyper_resnet/)
- [28] S. Lee, D. Jha, A. Agrawal, A. Choudhary, and W.-k. Liao, “Parallel deep convolutional neural network training by exploiting the overlapping of computation and communication,” in *2017 IEEE 24th International Conference on High Performance Computing (HiPC)*. IEEE, 2017, pp. 183–192.
- [29] T. Lidy, A. Schindler *et al.*, “Parallel convolutional neural networks for music genre and mood classification,” *MIREX2016*, vol. 3, 2016.
- [30] A. Olugboja, Z. Wang, and Y. Sun, “Parallel convolutional neural networks for object detection,” *Journal of Advances in Information Technology Vol*, vol. 12, no. 4, 2021.
- [31] W. Q. Zhang, T. H. Chan, and S. A. Vahid, “Serial and parallel convolutional neural network schemes for nfdm signals,” *Scientific reports*, vol. 12, no. 1, pp. 1–12, 2022.
- [32] S. Sakhavi, C. Guan, and S. Yan, “Parallel convolutional-linear neural network for motor imagery classification,” in *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 2736–2740.
- [33] Z. H. Hoo, J. Candlish, and D. Teare, “What is an roc curve?” pp. 357–359, 2017.

- 
- [34] I. e. I. Wikipedia, "Curva roc." [Online]. Available: [https://es.wikipedia.org/wiki/Curva\\_ROC](https://es.wikipedia.org/wiki/Curva_ROC)
- [35] L. D. Naya and E. D. Álvarez, "Diabetes mellitus. criterios diagnósticos y clasificación. epidemiología. etiopatogenia. evaluación inicial del paciente con diabetes," *Medicine-Programa de Formación Médica Continuada Acreditado*, vol. 12, no. 17, pp. 935–946, 2016.
- [36] A. D. Association *et al.*, "Diagnosis and classification of diabetes mellitus," *Diabetes care*, vol. 33, no. Supplement 1, pp. S62–S69, 2010.
- [37] M. Hernández-Ávila, J. P. Gutiérrez, and N. Reynoso-Noverón, "Diabetes mellitus en México: El estado de la epidemia," *Salud pública de México*, vol. 55, pp. s129–s136, 2013.
- [38] S. Yang, C. Nien, C. Lee, H. Chen, S. Chao, H. Hsu, S. Tzeng, J. Huang, H. Miao, and H. Lin, "The elevated risk of sight-threatening cataract in diabetes with retinopathy: a cohort study," 2020.
- [39] O. A. B. Salih and E. Hamed, "Open-angle glaucoma in diabetic patients' prevalence and risk factors: a cross-sectional study in Sudan," *International Journal of Research in Medical Sciences*, vol. 8, no. 6, p. 2205, 2020.
- [40] D. S. Fong, L. Aiello, T. W. Gardner, G. L. King, G. Blankenship, J. D. Cavallerano, F. L. Ferris, and R. Klein, "Retinopathy in diabetes," *Diabetes care*, vol. 27, no. suppl 1, pp. s84–s87, 2004.
- [41] M. Gandhi and R. Dhanasekaran, "Diagnosis of diabetic retinopathy using morphological process and svm classifier," in *2013 International Conference on Communication and Signal Processing*. IEEE, 2013, pp. 873–877.
- [42] Y. Liu, J. Li, J. Ma, and N. Tong, "The threshold of the severity of diabetic retinopathy below which intensive glycemic control is beneficial in diabetic patients:

- Estimation using data from large randomized clinical trials,” *Journal of diabetes research*, vol. 2020, 2020.
- [43] N. M. Tehrani, H. Riazi-Esfahani, E. Jafarzadehpur, A. Mirzajani, H. Talebi, A. Amini, M. Mazloumi, R. Roohipour, and M. Riazi-Esfahani, “Multifocal electroretinogram in diabetic macular edema; correlation with visual acuity and optical coherence tomography,” *Journal of ophthalmic & vision research*, vol. 10, no. 2, p. 165, 2015.
- [44] J. K. Sun, L. P. Aiello, M. D. Abràmoff, D. A. Antonetti, S. Dutta, M. Pragnell, S. R. Levine, and T. W. Gardner, “Updating the staging system for diabetic retinal disease,” *Ophthalmology*, 2020.
- [45] P. Rivas-Aguino, R. García-Amaris, M. Berrocal, J. Sánchez, A. Rivas, and J. F. Arevalo, “Vitrecomía pars plana, facoemulsificación e implante de lente intraocular para el manejo de catarata y retinopatía diabética proliferativa: comparación de técnica quirúrgica combinada versus en dos tiempos,” *Archivos de la sociedad española de oftalmología*, vol. 84, no. 1, pp. 31–37, 2009.
- [46] D. Toslak, T. Son, M. K. Erol, H. Kim, T.-H. Kim, R. P. Chan, and X. Yao, “Portable ultra-widefield fundus camera for multispectral imaging of the retina and choroid,” *Biomedical Optics Express*, vol. 11, no. 11, pp. 6281–6292, 2020.
- [47] S. Ravindran, “How artificial intelligence is helping to prevent blindness.” *Nature*, 2019.
- [48] L. Balyen and T. Peto, “Promising artificial intelligence-machine learning-deep learning algorithms in ophthalmology,” *The Asia-Pacific Journal of Ophthalmology*, vol. 8, no. 3, pp. 264–272, 2019.
- [49] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan *et al.*, “Identifying medical diagnoses and

- treatable diseases by image-based deep learning,” *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.
- [50] J. C. Armington, *The electroretinogram*. Academic Press, 1974.
- [51] G. H. Bresnick, K. Korth, A. Groo, and M. Palta, “Electroretinographic oscillatory potentials predict progression of diabetic retinopathy: preliminary report,” *Archives of ophthalmology*, vol. 102, no. 9, pp. 1307–1311, 1984.
- [52] R. Tzekov and G. Arden, “The electroretinogram in diabetic retinopathy,” *Survey of ophthalmology*, vol. 44, no. 1, pp. 53–60, 1999.
- [53] R. Rajagopal, G. W. Bligard, S. Zhang, L. Yin, P. Lukasiewicz, and C. F. Semenkovich, “Functional deficits precede structural lesions in mice with high-fat diet-induced diabetic retinopathy,” *Diabetes*, vol. 65, no. 4, pp. 1072–1084, 2016.
- [54] B. Fortune, M. E. Schneck, and A. J. Adams, “Multifocal electroretinogram delays reveal local retinal dysfunction in early diabetic retinopathy,” *Investigative ophthalmology & visual science*, vol. 40, no. 11, pp. 2638–2651, 1999.
- [55] M. Iragui-Madoz, “Tests electrofisiológicos en el estudio de la patología visual [electrophysiological test in the assessment of visual disorders],” *An. Sist. Sanit. Navar*, vol. 32, no. Suplemento 3, p. 93, 2009.
- [56] J. Catalá-Mora, M. Castany-Aregall, J. Berniell-Trota, L. Arias-Barquet, G. Rocalinares, and I. Jürgens-Mestre, “Electrorretinograma multifocal y degeneración macular asociada a la edad,” *Archivos de la Sociedad Española de Oftalmología*, vol. 80, no. 7, pp. 395–404, 2005.
- [57] K.-i. Murai, Y. Tazawa, M. Kobayashi, and A. Hayasaka, “Amplitude of the s-wave of multifocal electroretinograms can indicate local retinal sensitivity in glaucomatous eyes,” *Japanese journal of ophthalmology*, vol. 48, no. 3, pp. 215–221, 2004.

- [58] R. C. Ferras, Y. A. Escobar, M. M. Pastrana, I. M. Forzate, and R. L. Rodríguez, “Estudio normativo del electroretinograma a patrón,” *Revista Cubana de Tecnología de la Salud*, vol. 6, no. 1, pp. 13–23, 2015.
- [59] S. Kara, A. Güven, and A. Ö. Öner, “Utilization of artificial neural networks in the diagnosis of optic nerve diseases,” *Computers in Biology and Medicine*, vol. 36, no. 4, pp. 428–437, 2006.
- [60] E. P. García, O. H. Echevarría, Y. R. Martín, and J. G. González, “Obtención, interpretación y usos del electroretinograma multifocal,” *Revista Cubana de Oftalmología*, vol. 29, no. 1, pp. 105–123, 2016.
- [61] M. Bach, M. Hawlina, G. E. Holder, M. F. Marmor, T. Meigen, Y. Miyake *et al.*, “Standard for pattern electroretinography,” *Documenta ophthalmologica*, vol. 101, no. 1, pp. 11–18, 2000.
- [62] K. Penkala, T. Rogala, and A. Brykalski, “Selected methods of the pattern electroretinogram signal analysis,” *Pomiary Automatyka Kontrola*, vol. 53, pp. 22–25, 2007.
- [63] S. Ryan, *Elsevier Saunders*, 2013.
- [64] V. J. Bolós and R. Benítez, “The wavelet scalogram in the study of time series,” in *Advances in Differential Equations and Applications*. Springer, 2014, pp. 147–154.
- [65] R. Benítez, V. J. Bolós, and M. Ramírez, “A wavelet-based tool for studying non-periodicity,” *Computers & Mathematics with Applications*, vol. 60, no. 3, pp. 634–641, 2010.
- [66] G. Kaiser and L. H. Hudgins, *A friendly guide to wavelets*. Springer, 1994, vol. 300.

- [67] A. Alaql, *Analysis and processing of human electroretinogram*. University of South Florida, 2016.
- [68] M. X. Cohen, "A better way to define and describe morlet wavelets for time-frequency analysis," *NeuroImage*, vol. 199, pp. 81–86, 2019.
- [69] A. Güven and S. Kara, "Diagnosis of the macular diseases from pattern electroretinography signals using artificial neural networks," *Expert Systems with Applications*, vol. 30, no. 2, pp. 361–366, 2006.
- [70] S. Kara and A. Güven, "Training a learning vector quantization network using the pattern electroretinography signals," *Computers in Biology and Medicine*, vol. 37, no. 1, pp. 77–82, 2007.
- [71] L. Boquete, J. M. Miguel-Jiménez, S. Ortega, J. Rodríguez-Ascariz, C. Pérez-Rico, and R. Blanco, "Multifocal electroretinogram diagnosis of glaucoma applying neural networks and structural pattern analysis," *Expert Systems with Applications*, vol. 39, no. 1, pp. 234–238, 2012.
- [72] B. Fortune, M. A. Bearse, G. A. Cioffi, and C. A. Johnson, "Selective loss of an oscillatory component from temporal retinal multifocal erg responses in glaucoma," *Investigative ophthalmology & visual science*, vol. 43, no. 8, pp. 2638–2647, 2002.
- [73] P. Thienprasiddhi, V. C. Greenstein, C. S. Chen, J. M. Liebmann, R. Ritch, and D. C. Hood, "Multifocal visual evoked potential responses in glaucoma patients with unilateral hemifield defects," *American journal of ophthalmology*, vol. 136, no. 1, pp. 34–40, 2003.
- [74] S. Stiefelmeyer, A. S. Neubauer, T. Berninger, G. B. Arden, and G. Rudolph, "The multifocal pattern electroretinogram in glaucoma," *Vision research*, vol. 44, no. 1, pp. 103–112, 2004.

- [75] P. H. Chu, H. H. Chan, and B. Brown, "Glaucoma detection is facilitated by luminance modulation of the global flash multifocal electroretinogram," *Investigative ophthalmology & visual science*, vol. 47, no. 3, pp. 929–937, 2006.
- [76] J. M. Miguel-Jiménez, R. Blanco, L. De-Santiago, A. Fernández, J. M. Rodríguez-Ascariz, R. Barea, J. L. Martín-Sánchez, C. Amo, E. Sánchez-Morla, and L. Boquete, "Continuous-wavelet-transform analysis of the multifocal erg waveform in glaucoma diagnosis," *Medical & biological engineering & computing*, vol. 53, no. 9, pp. 771–780, 2015.
- [77] J. S. Ng, M. A. Bearse, M. E. Schneck, S. Barez, and A. J. Adams, "Local diabetic retinopathy prediction by multifocal erg delays over 3 years," *Investigative ophthalmology & visual science*, vol. 49, no. 4, pp. 1622–1628, 2008.
- [78] W. W. Harrison, M. A. Bearse, J. S. Ng, N. P. Jewell, S. Barez, D. Burger, M. E. Schneck, and A. J. Adams, "Multifocal electroretinograms predict onset of diabetic retinopathy in adult patients with diabetes," *Investigative ophthalmology & visual science*, vol. 52, no. 2, pp. 772–777, 2011.
- [79] I. S. Yapici, O. ErKaymaz, and R. U. Arslan, "A hybrid intelligent classifier to estimate obesity levels based on erg signals," *Physics Letters A*, vol. 399, p. 127281, 2021.
- [80] S. Gopal Krishna Patro, P. Parimita Sahoo, I. Panda, and K. K. Sahu, "Technical analysis on financial forecasting," *arXiv e-prints*, pp. arXiv–1503, 2015.
- [81] L. Al Shalabi, Z. Shaaban, and B. Kasasbeh, "Data mining: A preprocessing engine," *Journal of Computer Science*, vol. 2, no. 9, pp. 735–739, 2006.

## 14. Anexos

### 14.1. Anexo 1 Carta de autorización para uso de base de datos



A quien corresponda:

Yo, DRA. STÉPHANIE COLETTE THÉBAULT, en calidad de codirectora del trabajo de tesis “Modelo de clasificación de señales de electroretinograma para diagnóstico de retinopatía diabética temprana” autorizo a ING. CARLOS OLIVER HERNÁNDEZ MONTEJANO, estudiante la Maestría en Ciencias en Inteligencia Artificial de la Universidad Autónoma de Querétaro, para que haga uso de la base de datos de señales de electroretinograma, conformada por información de señales bioeléctricas de la retina de diferentes pacientes anónimos, tanto sanos como con algún tipo de diabetes para fines únicamente de investigación, garantizando la más estricta confidencialidad, uso responsable e integridad. La publicación de resultados de investigación se realizará únicamente bajo mi previa autorización.

Sin más que agregar, quedo a sus órdenes.

Atentamente,  
“POR MI RAZA HABLARA EL ESPIRITU”  
Juriquilla, Qro. 18 de enero de 2021

Dra. Stéphanie Colette Thébault

## 14.2. Anexo 2 Registro de pruebas de modelos

ID	Tipo de señales	Cantidad	Distribución	Técnica	Métricas	Tiempo de entrenamiento	Comentarios
1	Condiciones de luz	V1 Entrenamiento: 35 señales Pruebas: 15 señales Total: 50 señales	0 – No RD – 36 señales 1 – RD – 14 señales	Perceptrón multicapa Espectrograma NFFT de 300x300x1 Frecuencia muestral: 130 Hz Ventana: Hamming	Accuracy: 27%	10 horas+	
2	Condiciones de luz	V1 Entrenamiento: 35 señales Pruebas: 15 señales Total: 50 señales	0 – No RD – 36 señales 1 – RD – 14 señales	CNN Alexnet Espectrograma NFFT de 300x300x3 Frecuencia muestral: 130 Hz Ventana: Hamming	Accuracy: 65%	10 a 15 minutos	
3	Condiciones de oscuridad	V1 Entrenamiento: 36 señales Pruebas: 16 señales Total: 52 señales	0 – No RD – 38 señales 1 – RD – 14 señales	CNN Alexnet Espectrograma NFFT de 300x300x3 Frecuencia muestral: 130 Hz Ventana: Hamming	Accuracy: 72%	10 a 15 minutos	
4	Condiciones de luz	V1 Entrenamiento: 35 señales Pruebas: 15 señales Total: 50 señales	0 – No RD – 36 señales 1 – RD – 14 señales	CNN Alexnet Espectrograma NFFT de 300x300x3 Frecuencia muestral: 130 Hz Ventana: Hamming Descomposición por DWT DB10 Umbral SQTWLOG	Accuracy: 80%	10 a 15 minutos	Accuracy alto por desbalanceo de clases. La predicción decía que todos eran RD.
5	Condiciones de oscuridad	V1 Entrenamiento: 36 señales Pruebas: 16 señales Total: 52 señales	0 – No RD – 38 señales 1 – RD – 14 señales	CNN Alexnet Espectrograma NFFT de 300x300x3 Frecuencia muestral: 130 Hz Ventana: Hamming Descomposición por DWT DB10 Umbral SQTWLOG	Accuracy: 81%	10 a 15 minutos	Accuracy alto por desbalanceo de clases. La predicción decía que todos eran RD.
6	Condiciones de luz	V1 <b>Datos aumentados</b> Entrenamiento: 517 Pruebas: 129 Total: 646	0 – No RD – 320 señales 1 – RD – 319 señales	SVM Espectrograma NFFT de 300x300x3 Frecuencia muestral: 130 Hz Ventana: Hamming Descomposición por DWT DB 3-10 y Sym 2-12 Umbral SQTWLOG	Accuracy: 62%	20 a 25 minutos	
7	Condiciones de luz	V1 <b>Balaceo de muestras en cada clase</b> Entrenamiento: 57 Pruebas: 15 Total: 72	0 – No RD – 36 señales 1 – RD – 36 señales	AlexNet Modificada Espectrograma Librosa de 300x300x3 Sample rate: 1004 Espectrograma de frecuencia de Mel  Balanceo de datos Descomposición por DWT DB Nivel 2 y 3	Accuracy: 86.7 % Precision: 0.91 Recall: 0.83 F1: 0.85	30 a 45 minutos	
	Datos adicionales de validación	V1 Pruebas: 36 señales	0 – No RD – 26 señales 1 – RD – 10 señales	2	Accuracy: 55%		
	Datos adicionales de validación	V1 Pruebas: 36 señales	0 – No RD – 26 señales 1 – RD – 10 señales	4	Accuracy: 27%		
	Datos adicionales de validación	V1 Pruebas: 36 señales	0 – No RD – 26 señales 1 – RD – 10 señales	7	Accuracy: 52.8 % Precision: 0.39 Recall: 0.4 F1: 0.39		
	Datos originales aumentados	V1 Pruebas: 50 señales	0 – No RD – 36 señales 1 – RD – 14 señales	7 Aumento de datos Descomposición por DWT DB Nivel 1	Accuracy: 80.0 % Precision: 0.76 Recall: 0.71 F1: 0.73		
8	Condiciones de luz TRAIN 1 Espontaneos TRAIN 1 Espontaneos INDEREB TRAIN2 Modelo 1ra Generación	V2 Entrenamiento: 157 Validación: 40 Pruebas: 50 Total: 247	0 – No RD – 186 señales 1 – RD – 61 señales	AlexNet Modificada Espectrograma Librosa de 300x300x3 Sample rate: n / (n * 0.001) Espectrograma de frecuencia de Mel	Accuracy: 70.0 % Precision: 0.74 Recall: 0.68 F1: 0.7	25 a 30 min	Modificaciones pequeñas al modelo kernel_size strides loss_function: binary_crossentropy
9	Condiciones de luz TRAIN 1 Espontaneos TRAIN 1 Espontaneos INDEREB TRAIN2 Modelo 1ra Generación	V2 Entrenamiento: 157 Validación: 40 Pruebas: 50 Total: 247	0 – No RD – 186 señales 1 – RD – 61 señales	AlexNet Modificada Espectrograma Librosa de 300x300x3 Sample rate: n / (n * 0.001) Espectrograma de frecuencia de Mel	Accuracy: 74.0 % Precision: 0.72 Recall: 0.7 F1: 0.74	25 a 30 min	Modificaciones pequeñas al modelo kernel_size strides loss_function: binary_crossentropy
10	Condiciones de luz TRAIN 1 Espontaneos TRAIN 1 Espontaneos INDEREB TRAIN2 Modelo 1ra Generación	V2 Entrenamiento: 157 Validación: 40 Pruebas: 50 Total: 247	0 – No RD – 186 señales 1 – RD – 61 señales	AlexNet Modificada Espectrograma Librosa de 300x300x3 Sample rate: n / (n * 0.001) Espectrograma de frecuencia de Mel	Accuracy: 78.0 % Precision: 0.75 Recall: 0.69 F1: 0.75	25 a 30 min	Modificaciones pequeñas al modelo kernel_size strides loss_function: binary_crossentropy
11	Condiciones de luz TRAIN 1 Espontaneos TRAIN 1 Espontaneos INDEREB TRAIN2 Modelo 1ra Generación	V2 Entrenamiento: 157 Validación: 40 Pruebas: 50 Total: 247	0 – No RD – 186 señales 1 – RD – 61 señales	AlexNet Modificada Espectrograma Librosa de 300x300x3 Sample rate: n / (n * 0.001) Espectrograma de frecuencia de Mel	Accuracy: 86.0 % Precision: 0.8 Recall: 0.72 F1: 0.75	25 a 30 min	Modificaciones pequeñas al modelo kernel_size strides loss_function: binary_crossentropy
12	Condiciones de luz TRAIN 1 Espontaneos TRAIN 1 Espontaneos INDEREB TRAIN2 Modelo 1ra Generación	V2 Entrenamiento: 157 Validación: 40 Pruebas: 50 Total: 247	0 – No RD – 186 señales 1 – RD – 61 señales	AlexNet Modificada Espectrograma Librosa de 300x300x3 Sample rate: n / (n * 0.001) Frecuencia Máxima 15 Hz Espectrograma de frecuencia de Mel Filtrado por descomposición DWT DB Nivel 8	Accuracy: 94.0 % Precision: 0.97 Recall: 0.81 F1: 0.87	25 a 30 min	Modificaciones pequeñas al modelo kernel_size strides loss_function: binary_crossentropy Aplicación de filtrado por DWT DB con nivel 8 de descomposición La frecuencia Máxima a 15 Hz realiza un anecamiento resultando la sección del espectrograma

### 14.3. Anexo 3 Dictamen de aprobación de artículo en revista Mecamex

## La Mecatrónica en México



### DICTÁMEN DE EVALUACIÓN

6 de mayo de 2022

**OFICIO LMEM/20220506-1**

Por este medio, el comité editorial de la revista LMEM se les informa a los autores que el trabajo titulado

#### **Clasificación de Señales de Audio FSDD Utilizando Redes Neuronales Convolutivas**

Con número de registro **LMEM20211130-1**, cuyos autores son

**Hernández Montejano Carlos Oliver, González Huerta Rodrigo, Tovar Arriaga Saúl, Ramos Arreguin Juan Manuel, Takacs Andras**

Ha tenido como dictamen **APROBADO**.

Editorial

En la versión aprobada hay cambios de autor, además de haber eliminado al autor Juan Manuel Ramos Arreguín. Favor de indicar en el borrador la lista definitiva de los autores.

**ATENTAMENTE,**  
Comité Evaluador de los Artículos de la revista  
"La Mecatrónica en México"

