*Revision del rendimiento de la clasificacion de imágenes de retinopatia diabetica empleando modelos de redes neuronales convolucionales*

Abstract — Este estudio examina la importancia de la detección temprana de la retinopatía diabética (RD), una condición común en las personas con diabetes que puede conducir a la pérdida de la visión, y cómo las redes neuronales convolucionales (CNN) pueden ayudar a detectar las características principales en las imágenes de fondo de ojo para detectar señales tempranas de la RD. Las CNN es un método poderoso para procesar datos no estructurados, especialmente imágenes, audio y texto, que puede analizar de manera efectiva datos complejos y usarlos para detectar RD. Este artículo explica la arquitectura de CNN; que incluye una capa de entrada, una capa de convolución, una capa de agrupación y una capa completamente conectada, y analiza dos arquitecturas populares de CNN: Inceptionv3[1] [2] y ResNet34 [3][4][5]. Inceptionv3 es conocido por su capacidad para ejecutar filtros convolucionales de diferentes tamaños en una sola capa y utilizar módulos de transformación convolucional. Por otro lado, ResNet34 usa conexiones residuales o salta conexiones para permitir que la red aprenda las características de la imagen en diferentes niveles de abstracción. El documento destaca la importancia de conjuntos de datos grandes y diversos para usar modelos CNN para analizar RD.

Keywords - Retinopatía Diabética, Clasificación de imágenes, Captura de imágenes, Redes Neuronales Convolucionales, Nvidia Jetson Nano, classification algorithm.

1. Introducción

La detección temprana de la Retinopatía Diabética (RD) es crucial debido a su prevalencia entre pacientes con diabetes y puede causar pérdida de visión. En este estudio, se explora cómo las Redes Neuronales Convolucionales (CNN) pueden desempeñar un papel importante en la identificación de características predictivas en imágenes de fondo de ojo para detectar signos tempranos de RD.

Las CNN son herramientas poderosas para procesar datos no estructurados, especialmente en los dominios de imagen, audio y texto. Su capacidad para analizar datos complejos de manera eficiente ha demostrado ser útil en varios campos, incluida la detección de enfermedades como la RD. Este artículo presenta la arquitectura CNN, que comprende la capa de entrada, la capa de convolución, la capa de agrupación y la capa totalmente conectada.

La elección de una arquitectura de CNN adecuada depende de varios factores, como la eficiencia de los recursos, la precisión de la detección y la complejidad del modelo. Este estudio compara cuatro arquitecturas de uso común: DenseNet121, MobileNetV2, InceptionV3 y ResNet34. Están clasificados por su eficiencia, precisión y complejidad.

Además, se enfatiza la importancia de conjuntos de datos grandes y diversos para entrenar y aplicar modelos CNN para la detección de RD. El estudio también examina el proceso de entrenamiento y validación de los modelos, así como los resultados obtenidos.

Es cierto que, aunque las CNN son una herramienta invaluable para la detección de RD, existen limitaciones en términos de calidad de los datos de entrenamiento, variabilidad en la presentación clínica de la enfermedad y el potencial de signos ambiguos. Por lo tanto, la información obtenida debe ser complementada con otras pruebas y evaluaciones por parte de un médico calificado.

La elección adecuada de la arquitectura CNN y la consideración de las restricciones son aspectos clave para garantizar resultados de detección de RD precisos y válidos.

1. Antecedentes
   1. Retinopatía Diabética

La RD es una enfermedad muy común en pacientes con diabetes y es la principal causa de pérdida de visión [6] - [8]. Según un estudio publicado por la Academia Estadounidense de Oftalmología, que recaudo 59 estudios poblacionales publicados hasta marzo de 2020 de fuentes como PubMed, Medline, Web of Science y Scopus, se encontró que la incidencia de RD en adultos en todo el mundo era una medida de prevalencia es 6,17% [9][10]. La RD puede ser asintomática en sus primeras etapas, lo que significa que la detección temprana es fundamental para prevenir la pérdida de visión. La edad, el tabaquismo y la hipertensión son factores que pueden exacerbar las complicaciones y acelerar la pérdida de la visión [7][11][12]. Con los avances en los algoritmos de redes neuronales convolucionales, las características predictivas de RD pueden identificarse en imágenes de fondo de ojo obtenidas de pacientes diabéticos diagnosticados con RD para identificar señales en otros pacientes [13], [14].

* 1. Arquitectura de redes neuronales convolucionales

Las redes neuronales convolucionales (CNN) son un método poderoso para procesar datos no estructurados, especialmente imágenes, audio y texto. Es importante comprender que las CNN están diseñadas para procesar matrices de datos y utilizar capas convolucionales para extraer características de los datos de entrada. Esta arquitectura permite que CNN identifique de manera eficiente patrones y características relevantes en los datos de entrada, lo que la convierte en una herramienta importante para la clasificación y el análisis de datos complejos. Cuando se desarrollan sistemas de software que involucran procesamiento de datos no estructurados, es importante considerar el uso de redes neuronales convolucionales para proporcionar resultados precisos y eficientes [1].

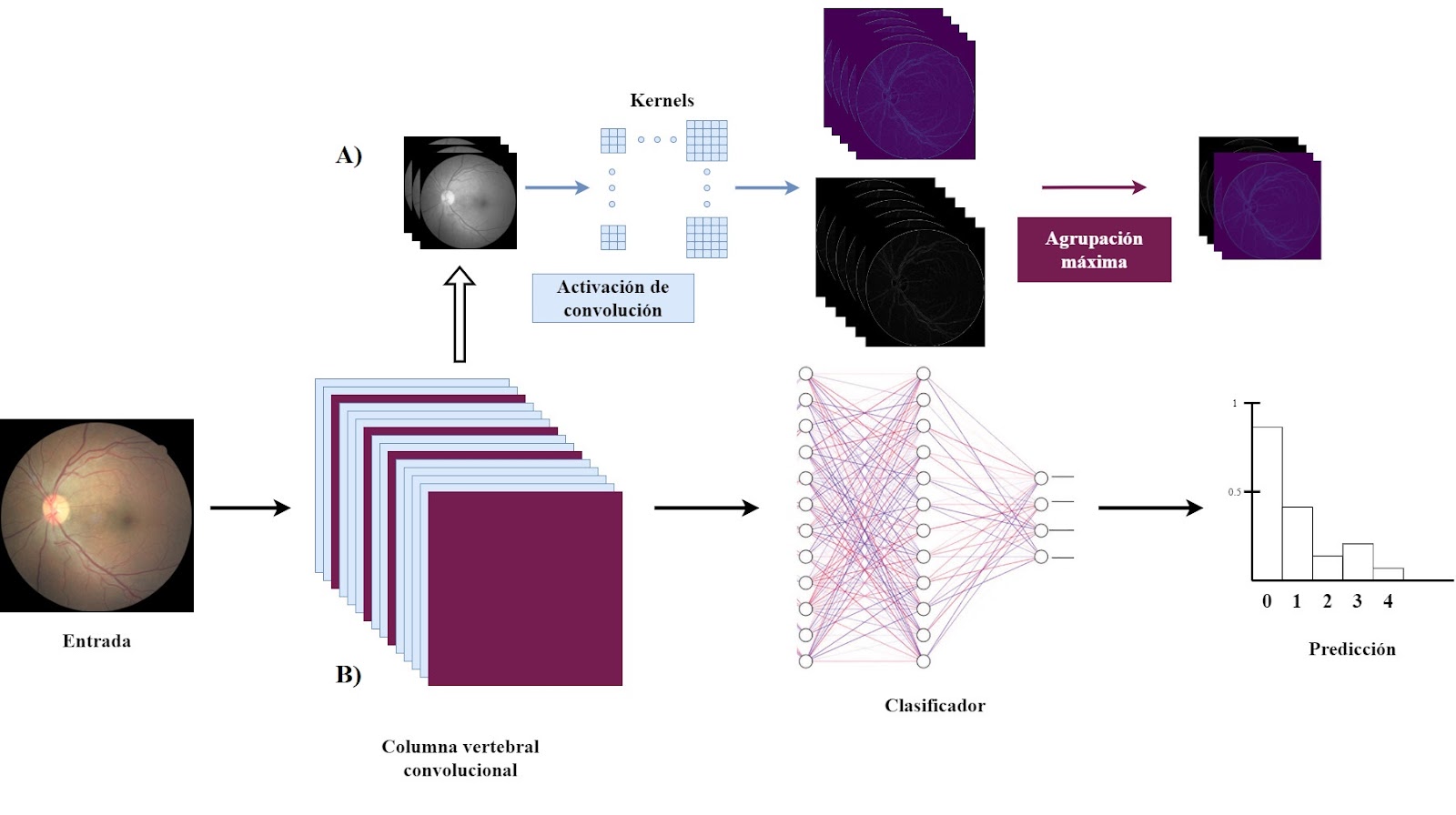


Figura 1. La arquitectura de una red neuronal convolucional (CNN) para el reconocimiento de imágenes utiliza operaciones de convolución y activación ReLU (Unidad Lineal Rectificada), junto con operaciones de agrupación de máximos adyacentes, en una columna vertebral convolucional. Estas operaciones se aplican repetidamente para generar mapas de características y profundizar en ellos hasta llegar a un clasificador, que en este caso es otra CNN, que emite una predicción.

La arquitectura CNN consta de varias capas, incluida una capa de entrada, capas convolucionales para extraer características de los datos de entrada, capas de agrupación para reducir el tamaño de la salida y capas totalmente conectadas para generar la salida final. La arquitectura de una CNN variará según los datos de entrada y la tarea en cuestión. Además, los parámetros deben ajustarse para lograr el mejor rendimiento en la tarea dada [15][16].

La elección de la arquitectura CNN (Red Neural Convolucional) en visión artificial es muy importante para lograr modelos de aprendizaje profundo de alta precisión y alto rendimiento. DenseNet121, MobileNetV2, InceptionV3 y ResNet34 son las arquitecturas de CNN más utilizadas debido a su precisión en varias tareas de visión artificial. Para determinar qué arquitectura CNN es mejor para un problema de visión por computadora en particular, comparamos estas arquitecturas entre sí. Aunque hay muchas otras arquitecturas de CNN disponibles, se eligieron estas cuatro por su complejidad y precisión, así como por la eficiencia de los recursos.

Se comparan cuatro arquitecturas de redes neuronales convolucionales (CNN) que se usan comúnmente en aplicaciones de visión por computadora: DenseNet121, MobileNetV2, InceptionV3 y ResNet34. Se evalúan estas arquitecturas según su eficiencia, precisión y complejidad.

1. Comparación de las arquitecturas CNN. Se comparan cuatro arquitecturas de redes neuronales convolucionales (CNN) en visión por computadora: DenseNet121, MobileNetV2, InceptionV3 y ResNet34. Se evalúan según eficiencia, exactitud y complejidad. MobileNetV2 es la más eficiente, DenseNet121 e InceptionV3 son las más precisas y DenseNet121 e InceptionV3 son más complejas.

|  |  |  |  |
| --- | --- | --- | --- |
| **CNN** | **Año** | **Profundidad** | **Top-1 err en ImageNet** |
| DenseNet121[17] | 2017 | 121 | 25.02% |
| MobileNetV2[18] | 2018 | 53 | 28% |
| InceptionV3[2] | 2015 | 48 | 18.77% |
| ResNet34[3][5] | 2015 | 34 | 25.03% |

En términos de eficiencia, MobileNetV2 es la arquitectura más eficiente entre las cuatro porque está diseñada específicamente para dispositivos móviles e integrados, lo que reduce el consumo de recursos como memoria y energía. Por el contrario, DenseNet121 e InceptionV3 tienen requisitos de recursos medios.

En términos de precisión, DenseNet121 e InceptionV3 son las arquitecturas más precisas y, por lo general, superan a MobileNetV2 y ResNet34. Sin embargo, la precisión de estas arquitecturas puede variar mucho según el conjunto de datos y el problema que se está resolviendo.

Complejidad, DenseNet121 e InceptionV3 tienen arquitecturas de redes neuronales convolucionales más complejas que MobileNetV2 y ResNet34. Esta complejidad está relacionada con el número de capas y la naturaleza de las unidades de procesamiento funcional utilizadas en ellas. Aunque una mayor complejidad mejora la precisión de la red, también aumenta la carga computacional y los requisitos de memoria, lo que puede ser un factor limitante en algunos casos.

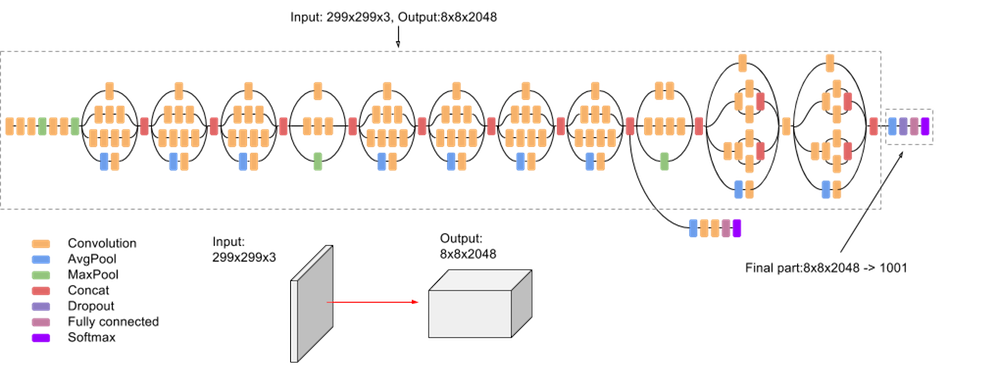
La elección de la CNN más adecuada dependerá de las necesidades específicas del problema a resolver. Si la eficiencia de los recursos es crítica, MobileNetV2 es la mejor opción. Sin embargo, si busca una alta precisión en la mayoría de los conjuntos de datos, DenseNet121 e InceptionV3 son los más precisos, aunque su complejidad puede limitar su uso en algunos casos. Por lo tanto, para problemas que requieren un equilibrio entre precisión y eficiencia moderada, se consideraron ResNet34 e InceptionV3.

Figura 2. Arquitectura de InceptionV3. Este emplea una estrategia conocida como "módulos de Inception", los cuales fusionan de manera simultánea filtros convolucionales de distintos tamaños para capturar características a diferentes escalas. Asimismo, esta arquitectura incorpora operaciones como convoluciones 1x1 y reducción de dimensionalidad, con el fin de mejorar la eficiencia computacional y disminuir la cantidad de parámetros que necesitan ser entrenados [19].

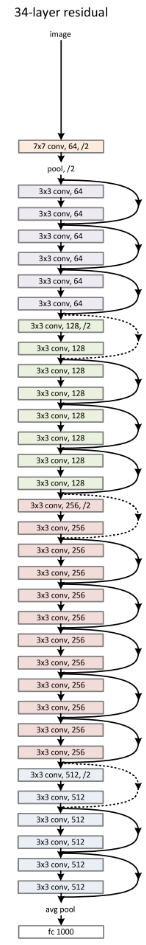
* 1. Inceptionv3

Inceptionv3 es una arquitectura CNN (Convolutional Neural Network) utilizada para la clasificación de imágenes desarrollada por Google en 2015 como una versión mejorada de la red Inception original, también conocida como GoogLeNet [1].

La arquitectura Inceptionv3 se destaca por su capacidad para ejecutar filtros convolucionales de diferentes tamaños en una sola capa y utilizar módulos de transformación convolucional. Estos módulos son bloques de construcción que realizan convoluciones en paralelo en diferentes tamaños de kernel, lo que les permite aprender características a diferentes escalas. Además, Inceptionv3 utiliza la normalización de paquetes y las funciones de activación de ReLU para mejorar la estabilidad y la no linealidad de la red[16].

* 1. ResNet34

ResNet34 es una arquitectura de red neuronal convolucional (CNN) para la clasificación de imágenes desarrollada por Microsoft en 2015. Es una versión mejorada de la arquitectura ResNet original, también conocida como ResNet18 [3] [4].

Se sabe que la arquitectura ResNet34 usa conexiones residuales (también conocidas como saltos de conexión) que permiten que la red aprenda características en imágenes en diferentes niveles de abstracción. Estas conexiones residuales evitan el problema de desvanecimiento de gradiente que ocurre cuando se usan redes profundas. En ResNet34, las capas convolucionales se agrupan en bloques, cada bloque tiene varias capas y las conexiones restantes se utilizan para omitir uno o más bloques[3][4][16].

1. Caso de estudio

En este caso de estudio, se analizarán conjuntos de imágenes de RD con el objetivo de evaluar el rendimiento de Inceptionv3 y ResNet34.

* 1. Conjunto de datos

El uso de modelos de redes neuronales convolucionales (CNN) para analizar la retinopatía diabética requiere un conjunto de datos amplio y diverso que contenga imágenes de la retina en diferentes etapas de la enfermedad. El conjunto de datos de retinopatía diabética proliferativa y no proliferativa en estadios leves, moderados y graves es un recurso importante para los investigadores que desean entrenar modelos de redes neuronales convolucionales (CNN) para analizar la enfermedad.

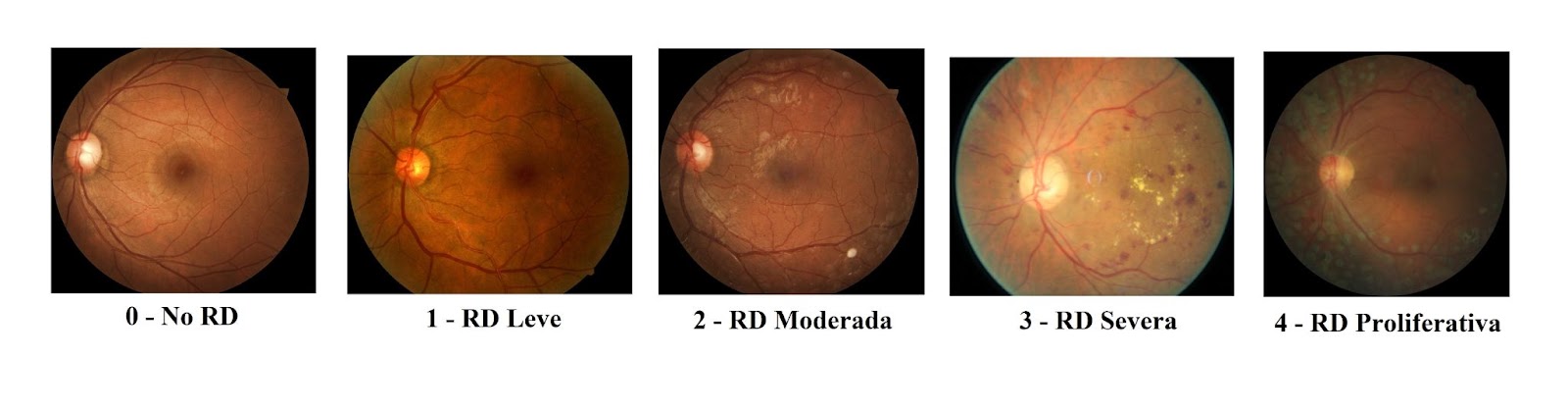
Conjuntos de datos de retinopatía diabética disponibles en la web, como Kaggle [18] y otras plataformas. Es una fuente invaluable de recursos para la investigación oftalmológica porque contiene imágenes de la retina de pacientes diabéticos en varias etapas de la enfermedad. Algunos de estos conjuntos de datos "Detección de retinopatía diabética" y "Detección de ceguera APTOS 2019". Ambos conjuntos de datos contienen imágenes retinales de pacientes diabéticos y etiquetas que indican la gravedad de la retinopatía diabética.

Figura 3. Imágenes de retinopatía diabética en sus diferentes etapas recuperadas del conjunto de imágenes ‘Diabetic Retinopathy Detection’ de Kaggle[18].

El acceso a conjuntos de datos grandes y variados es fundamental para el desarrollo de modelos de CNN precisos y eficientes. Los conjuntos de datos de retinopatía diabética descargados desde Kaggle proporcionan una fuente importante de imágenes y datos clínicos para el análisis de la enfermedad.

1. Información sobre como la estructura de los conjuntos de datos de retinopatia diabetrica recuperdas de Kaggle.

Figura 4. Arquitectura de ResNet34. Está conformada por 34 capas, que incluyen capas convolucionales, capas de agrupación y capas completamente conectadas. La arquitectura sigue un patrón de bloques residuales, donde cada bloque está compuesto por capas de convolución y normalización, seguidas de una conexión residual que combina la entrada original con la salida del bloque. En cuanto a los filtros convolucionales utilizados en ResNet34, mayoritariamente tienen un tamaño de 3x3, lo cual permite capturar características locales. También se emplean capas de agrupación máxima (max pooling) para reducir la dimensión espacial de las características[3].

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre** | **Imágenes** | **Dimensión** | **Descripción** |
| Diabetic Retinopathy Detection | 34108 | 1024 x 1024 | Conjunto de imágenes de retina de alta resolución estandarizado. |
| APTOS 2019 Blindness Detection | 3662 | 1024 x 1024  3216 x 2136 | Imágenes recopiladas de diferentes conjuntos de imágenes, cuentan con tamaño y resoluciones diferentes. |

En la Tabla 2. Se muestran los conjuntos de datos que se emplearán para entrenar los modelos. Estos incluyen imágenes sobre las diferentes etapas del padecimiento, tales como retinopatía diabética no proliferativa en sus etapas leve, moderada y severa, así como la retinopatía diabética proliferativa. También, contiene imágenes de retina sin retinopatía diabética.

* + 1. Distribución del conjunto de datos

La composición del conjunto de datos de retinopatía diabética contiene las imágenes sobre el grado de daño en los vasos sanguíneos, así como otros factores relevantes según la clase correspondiente de retinopatía diabética.

* + - 1. Diabetic Retinopathy Detection

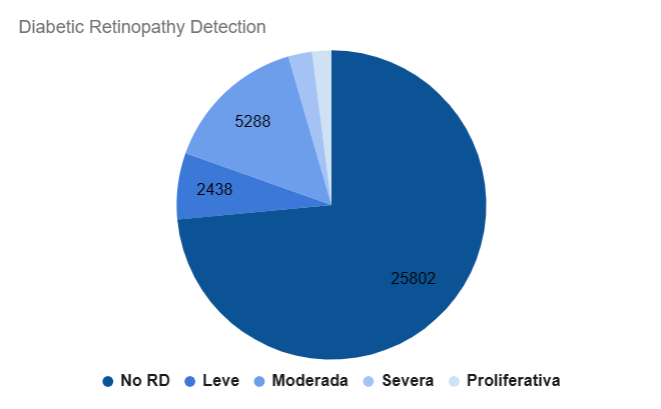
El conjunto de datos Diabetic Retinopathy Detection de Kaggle, está compuesto por las  diferentes clases del padecimiento y su distribución es la siguiente [19].

Figura 5. Distribución de las diferentes clases que componen el conjunto de datos Diabetic Retinopathy Detection[19]. 25802 de No RD, 2438 son Leve, 5288 son Moderada, 872 son Severa, y 708 son Proliferativa

* + - 1. APTOS 2019 Blindness Detection

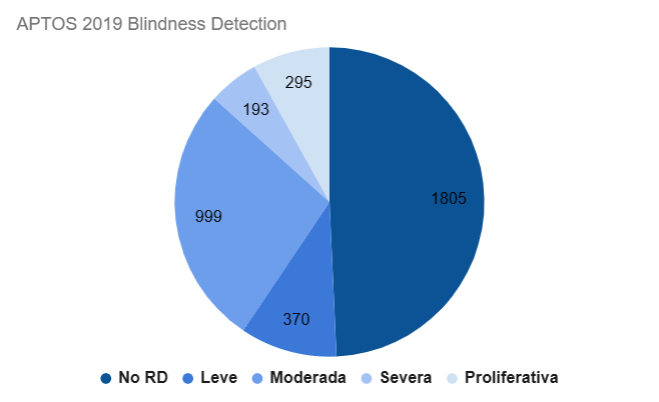
El conjunto de datos APTOS 2019 Blindness Detection de Kaggle, está compuesto por las  diferentes clases del padecimiento y su distribución es la siguiente [20].

Figura 6. Distribución de las diferentes clases que componen el conjunto de datos APTOS 2019 Blindness Detection[20]. 1805 de No RD, 370 son Leve, 999 son Moderada, 193 son Severa, y 295 son Proliferativa.

Los conjuntos de datos contienen imágenes de las diferentes clases o etapas de retinopatía diabética desde, no retinopatía diabética hasta proliferativa. Las imágenes se encuentran en formatos JPEG y PNG.

1. Redes Neuronales Convolucionales

La retinopatía diabética es una enfermedad ocular común en personas con diabetes que puede conducir a la pérdida de la visión si no se diagnostica y trata a tiempo. Para detectar la enfermedad de forma temprana mediante un modelo de aprendizaje profundo, se analizan las imágenes de la retina de personas con RD para predecir la presencia o indicios de RD en otras personas.

Se realizó una división aleatoria del conjunto de datos, asignando el 85% de las imágenes para el entrenamiento y el 15% restante para realizar una validación inicial. Luego, se utilizará todo el conjunto de datos para realizar una validación final de los modelos entrenados y verificar su desempeño general.

* 1. Entrenamiento

Un modelo de aprendizaje profundo puede analizar grandes cantidades de datos de imágenes de la retina y descubrir patrones que pueden indicar la presencia de una enfermedad. Además, estos modelos son de autoaprendizaje, lo que significa que pueden mejorar la precisión y la eficiencia a medida que se les suministran más datos. En última instancia, mediante el uso de modelos de aprendizaje profundo para detectar la retinopatía diabética de manera temprana, los médicos pueden diagnosticar y tratar la enfermedad de manera más efectiva, lo que puede mejorar la calidad de vida de los pacientes y reducir los costos asociados con la atención a largo plazo.

1. División de conjuntos. Destina un 85% para entrenamiento y un 15% para realizar una primera validación

|  |  |  |
| --- | --- | --- |
| **Conjunto de datos** | **Entrenamiento** | **Validación** |
| Diabetic Retinopathy Detection | 29841 | 5267 |
| APTOS 2019 Blindness Detection | 3112 | 550 |

Se crearon dos instancias de entrenamiento para cada modelo, de esta forma se puede realizar una comparación en el rendimiento que presenta cada modelo acorde al conjunto con el que fue entrenado.

Para realizar el entrenamiento se usó Google Colab y RunPod, estas son herramientas disponibles en la nube para entrenar modelos de aprendizaje profundo, ambas proporcionan recursos como GPU y TPU, que pueden acelerar el proceso de entrenamiento. A pesar de sus similitudes, hay algunas diferencias importantes entre ellas que pueden afectar la elección de la herramienta adecuada para un proyecto específico. Para el entrenamiento se establecio una configuración de 50 épocas, Python 3.9, Pytorch 1.13.1 cu116 y una GPU, según la herramienta utilizada el tiempo total de entrenamiento varía desde 8 horas en RunPod a 20 horas en Google Colab.

* 1. Validación

Como resultado del entrenamiento, se cuenta con 4 modelos que pueden ser utilizados para detectar retinopatía diabética. Para determinar qué modelo presenta un mejor desempeño, se realiza una comparación mediante el cálculo de imágenes clasificadas correctamente

1. Resultados de la validación de  los modelos. InceptionV3(1) y ResNet(1) - Modelos entrenados con el conjunto de datos Diabetic Retinopathy, InceptionV3(2) y ResNet(2) - Modelo entrenado con el conjunto de datos APTOS 2019 Blindness. Imágenes: Numero de imágenes que se Utilizaron para realizar la evaluación del modelo. Exactitud: Porcentaje del número de imágenes clasificadas correctamente. Error: Porcentaje de imágenes que se clasificaron de manera incorrecta.

|  |  |  |  |
| --- | --- | --- | --- |
| **Modelo** | **Imágenes** | **Exactitud** | **Error** |
| InceptionV3(1)  InceptionV3(2) | 5267  550 | 74.407%  78.181% | 25.593%  21.819% |
| ResNet34(1)  ResNet34(2) | 5267  550 | 74.197%  73.454% | 25.803%  26.546% |

Los resultados del uso de modelos InceptionV3 y ResNet34 para detectar retinopatía diabética indica que estos modelos pueden ser efectivos en la detección de enfermedades oculares en pacientes diabéticos. De cada 10 pacientes que puedan presentar una retinopatía diabética en sus diferentes etapas, al menos 7 de ellas serán detectadas correctamente.

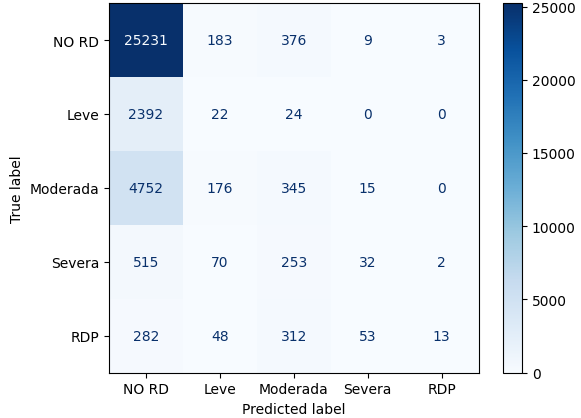


Figura 7. Matriz de confusión de los resultados de la clasificación del conjunto de datos 1(Diabetic Retinopathy) utilizando ResNet34.

El desempeño presentado en los modelos InceptionV3 frente a ResNet34 indicaron que presentan una mayor exactitud. Por otro lado, los modelos ResNet34 mostraron una facilidad al concluir su entrenamiento en menor tiempo. Además, la brecha de exactitud entre los modelos no es muy amplia.

1. Resultados

Los cuatro modelos resultantes validados y que presentan una exactitud mayor al 70%, son empleados para procesar en su totalidad los conjuntos de datos de imágenes de retinopatía diabética. Una vez terminada la clasificación, se calcularon los resultados para determinar el número de imágenes que fueron clasificadas correctamente.

1. Resultados del desempeño de los modelos InceptionV3 y ResNet34. Modelo: InceptionV3(1) y ResNet(1) - Modelo entrenado con el conjunto de datos Diabetic Retinopathy, InceptionV3(2) y ResNet(2) - Modelo entrenado con el conjunto de datos APTOS 2019 Blindness. Evaluar: Nombre del conjunto de datos que se empleó para validar el modelo. Correctos(Corr): Número de imágenes que fueron clasificadas correctamente del conjunto de datos. Error(Err): Número de imágenes que fueron clasificadas incorrectamente del conjunto de datos. Exactitud: Porcentaje obtenido de las imágenes clasificadas correctamente.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Modelo** | **Evaluar** | **Corr** | **Err** | **Exactitud** |
| InceptionV3(1)  InceptionV3(2)  InceptionV3(1)  InceptionV3(2) | Diabetic Retinopathy  APTOS 2019 Blindness | 26134  25192  1969  2842 | 8974  9916  1693  820 | 74.44%  71.76%  53.77%  77.61% |
| ResNet34(1)  ResNet34(2)  ResNet34(1)  ResNet34(2) | Diabetic Retinopathy  APTOS 2019 Blindness | 33680  25643  2336  2760 | 1428  9465  1326  902 | 95.93%  73.04%  63.79%  75.37% |

Al analizar los resultados de los modelos de clasificación, se observa que el modelo ResNet34 entrenado con el conjunto de datos 2 presenta una exactitud más estable frente a los otros modelos. En particular, este modelo logró clasificar correctamente la mayoría de los datos tanto en el conjunto de datos 1 como en el conjunto de datos 2, obteniendo una tasa de error menor en comparación con los otros modelos. Esto sugiere que el modelo ResNet34 entrenado con el conjunto de datos 2 es más robusto y generalizable en comparación con los otros modelos. En conclusión, el modelo ResNet34 entrenado con el conjunto de datos 2 es el más adecuado para la tarea de clasificación en cuestión, ya que presenta un mejor rendimiento y una mayor estabilidad en la precisión de las predicciones.

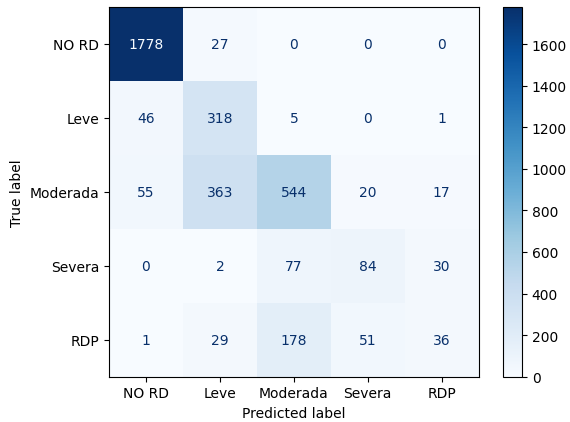
La matriz de confusión de la Figura 5 y 6. Proporciona una visualización detallada de los resultados de la clasificación del conjunto de datos 1 y 2 utilizando el modelo ResNet34. En la diagonal principal de la matriz, se encuentran los valores que representan los aciertos, mientras que fuera de la diagonal se encuentran los errores de clasificación. Al analizar la matriz de confusión, se puede observar que el modelo ResNet34 logró clasificar correctamente la mayoría de las imágenes de ambos conjuntos de datos. Sin embargo, también se observa que hubo algunos casos de falsos positivos y falsos negativos, lo que indica que aún hay margen de mejora para este modelo.

Figura 8. Matriz de confusión de los resultados de la clasificación del conjunto de datos 2(APTOS 2019 Blindness) utilizando ResNet34

1. Detalles de la clasificación

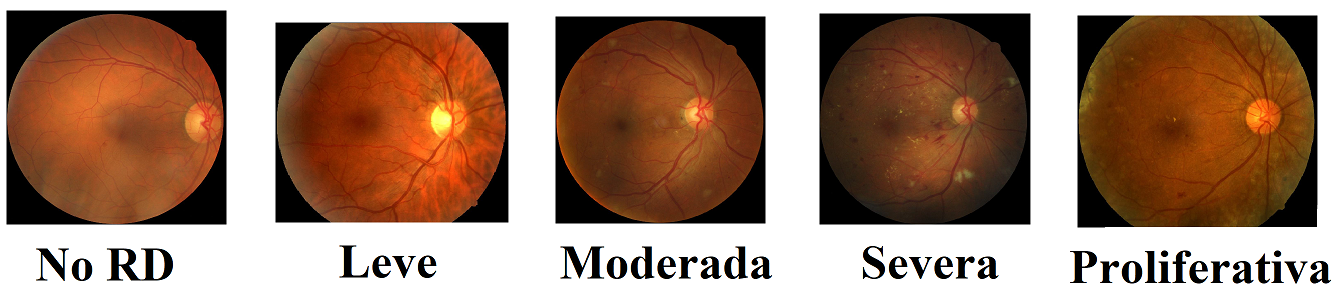
Vale la pena señalar que CNN es un modelo de aprendizaje automático que utiliza datos de entrada para identificar patrones y características que ayudan a clasificar la información. En el caso de la retinopatía diabética, las CNN pueden analizar imágenes de la retina y buscar patrones específicos para cada etapa de la enfermedad. Por ejemplo, La RD en etapa leve puede tener pequeñas áreas de hemorragia o exudado, mientras que en caso moderado puede tener múltiples áreas de hemorragia, exudado y abultamiento. En etapa severa, las anomalias son graves y notables.

Figura 9. Imágenes de retinopatía diabética clasificadas correctamente por el modelo ResNet34. Imágenes recuperadas del conjunto ‘Diabetic Retinopathy Detection’ de Kaggle.

También es posible que algunas imágenes presenten características ambiguas o que no sean claramente atribuibles a una etapa específica de la enfermedad, lo que puede llevar a una clasificación errónea. Además, otros factores como la calidad de la imagen o la variabilidad en la presentación clínica de la enfermedad pueden afectar la precisión de la clasificación. Asimismo, si la cantidad o calidad de los datos de entrenamiento no es suficiente o adecuada, la CNN puede no estar completamente equipada para identificar todas las características relevantes.

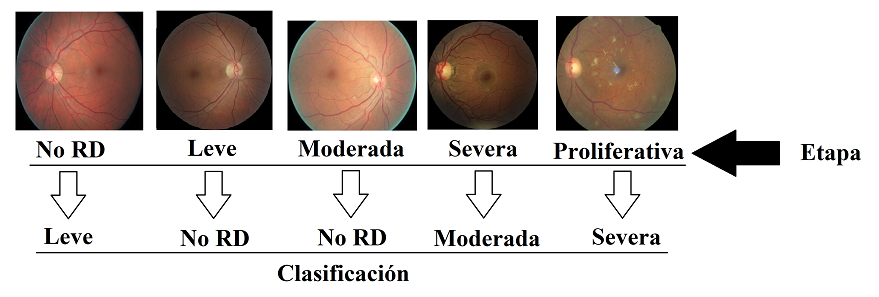
Es importante tener en cuenta que aunque una CNN puede ser útil para la detección de la retinopatía diabética, existen limitaciones en su precisión. Por ejemplo, una CNN podría confundir una retinopatía diabética leve con una persona que no tiene retinopatía diabética, lo cual puede generar falsos negativos o falsos positivos de ser el caso contrario. De igual manera, una CNN podría confundir una retinopatía diabética moderada con una persona que no tiene retinopatía diabética, o incluso confundirla con una retinopatía diabética severa. Por lo tanto, es importante tener en cuenta las limitaciones de la tecnología y utilizarla en conjunto con la evaluación clínica de un especialista para un diagnóstico preciso.

Figura 10. Imágenes de retinopatía diabética clasificadas incorrectamente por el modelo ResNet34. Se indica una imagen por etapa y la clasificación realizada por el modelo. Imágenes recuperadas del conjunto ‘Diabetic Retinopathy Detection’ de Kaggle.

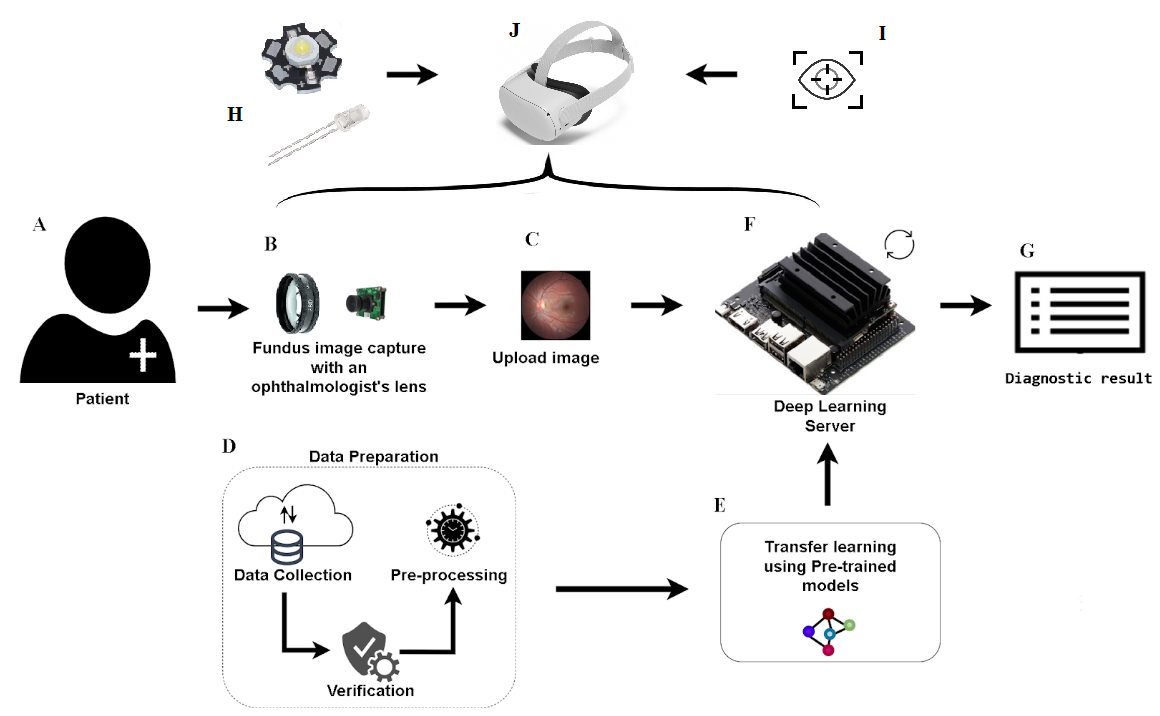
1. Limitantes

Una CNN es una herramienta útil para la identificación de las diferentes etapas de la retinopatía diabética a partir de imágenes de la retina, pero su precisión está limitada por la calidad de los datos de entrenamiento, la variabilidad de la presentación clínica de la enfermedad y la presencia de características ambiguas o no específicas. Por lo tanto, es importante tener en cuenta estas limitaciones al utilizar una CNN para el diagnóstico de la retinopatía diabética y siempre se debe complementar la información obtenida con otras pruebas y la evaluación de un profesional médico capacitado

1. Conclusiones

La retinopatía diabética es una afección común en las personas con diabetes y es la principal causa de pérdida de la visión. La detección temprana es fundamental para prevenir la pérdida de visión. Gracias al desarrollo de algoritmos de redes neuronales convolucionales, se pueden identificar características predictivas de RD en imágenes de fondo de ojo obtenidas de pacientes diabéticos para identificar señales de RD en pacientes jóvenes. Las redes neuronales convolucionales son una herramienta importante para analizar y clasificar datos complejos, especialmente imágenes. Inceptionv3 y ResNet34 son arquitecturas de redes neuronales convolucionales utilizadas para la clasificación de imágenes y son muy eficientes para esta tarea. El uso de modelos de redes neuronales convolucionales para analizar la retinopatía diabética requiere un conjunto de datos grande y diverso que consta de imágenes de la retina en diferentes etapas de la enfermedad.

1. Trabajo a futuro

Se está desarrollando un sistema para la clasificación de imágenes de retinopatía diabética utilizando ResNet34. Este modelo se entrenó para identificar las etapas de la enfermedad: Leve, Moderada, Severa y Proliferativa. Además, se implementa un sistema de captura de imagen que aprovecha un lente de 20 dioptrías para obtener imágenes del fondo de ojo de los pacientes. Un código específico detecta la retina del ojo izquierdo y rastrea su movimiento. Cuando el paciente dirige su mirada hacia adelante, se captura la imagen del ojo derecho. Para mejorar la detección de relieve y contornos oculares, se utilizan LEDs en el lado izquierdo. Estos LEDs se controlan mediante una Nvidia Jetson Nano, al igual que la cámara y los códigos desarrollados. Finalmente, todo el sistema, incluyendo la cámara, los LEDs y el lente de 20 dioptrías, se monta en la carcasa de unos lentes de realidad virtual (VR). Además, se implementa un control de luz en el lado derecho, donde se enciende un LED simulando el efecto de un flash, proporcionando una ráfaga de luz hacia el lente de 20 dioptrías para capturar la imagen del fondo de ojo.

* 1. Paciente

Se propone utilizar el sistema para diagnosticar a pacientes que presentan problemas o malestar en los ojos, como visión borrosa o presencia de manchas o hebras oscuras en la vista. También se utiliza como medida preventiva para detectar retinopatía diabética.

* 1. Captura de fondo de ojo

La captura de imágenes del fondo de ojo se realiza con la ayuda de un lente de 20 dioptrías, lo que permite visualizar los vasos sanguíneos y el tejido ocular. Esta imagen se utiliza para detectar posibles daños o pérdida de visión.

* 1. Carga de Imagen

Se cargan las imágenes capturadas en una jetson nano developer kit de NVIDIA, ya que es adecuada para algoritmos de inteligencia artificial y modelos de aprendizaje profundo.

* 1. Preparación de datos

Como se indica en este estudio, se utiliza el modelo ResNet 34, este modelo ya está entrenado y validado con conjuntos de imágenes de retina disponibles en Kaggle.

* 1. Transferencia de aprendizaje

Se transfiere el modelo Resnet34 pre-entrenado para su uso en la jetson nano, con la capacidad de detectar la Retinopatía Diabética en sus etapas leve, moderada, grave y proliferativa, así como a los que no presenten este padecimiento.

Figura 11. Arquitectura del sistema propuesto.

* 1. Servidor de Aprendizaje profundo

Una vez capturadas las imágenes del fondo de ojo, se utiliza el modelo ResNet34 pre-entrenado para determinar si la imagen es de un paciente que padece o no de Retinopatía Diabética, y en caso de padecerla, indica en qué etapa se encuentra.

* 1. Diagnostico de resultado

Con el resultado del diagnóstico obtenido, se informa al paciente sobre su condición y se permite que el público en general, sin experiencia médica, se someta al examen mediante el sistema propuesto. En caso necesario, se puede derivar al paciente a un especialista para un tratamiento especializado.

* 1. LEDs

Para el control de la luminosidad en el ojo izquierdo, se utilizan leds de 5mm blancos para iluminar los contornos y detectarlos. Además, para el ojo derecho, se emplea un led de 3.5 mm que funciona como un flash.

* 1. Ubicacion de la pupila

Con una cámara y mediante el software desarrollado, se reconoce la pupila izquierda del usuario y cuando la pupila se encuentra alineada en el centro, se realiza una captura con la cámara que está en el ojo derecho, donde se encuentra el lente de 20 dioptrías, obteniendo así una imagen del fondo de ojo del usuario.

* 1. Gafas de realidad virtual(VR)

El sistema de captura de imagen, que comprende las dos cámaras, los leds y la jetson nano de Nvidia con el modelo pre-entrenado y el código que reconoce la pupila y el control de leds, se encuentra montado dentro de unas gafas de realidad virtual (VR).

Este sistema ofrece un enfoque prometedor para la detección temprana y el monitoreo de esta enfermedad ocular. Además, el sistema permite que tanto pacientes con problemas o malestar en los ojos como personas sin experiencia médica se sometan al examen, proporcionando un diagnóstico rápido. En caso necesario, se deriva a los pacientes a especialistas para un tratamiento especializado.

Reconocimiento

Este trabajo es apoyado por un estímulo de estudiante de posgrado otorgado a Isaul Ibarra Belmonte por el Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT).

References

[1] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, et al., "Backpropagation applied to handwritten zip code recognition", *Neural Comput.*, vol. 1, no. 4, pp. 541-551, 1989.  
[2] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2818-2826, Jun. 2016.  
[3] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 770-778, 2016.  
[4] M. Shafiq y Z. Gu, "Deep Residual Learning for Image Recognition: A Survey", Appl. Sci., vol. 12, n.º 18, p. 8972, septiembre de 2022. Accedido el 14 de junio de 2023. [En línea]. Disponible: <https://doi.org/10.3390/app12188972>.  
[5] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions", SN Comput. Sci., vol. 2, n.º 3, marzo de 2021. Accedido el 14 de junio de 2023. [En línea]. Disponible: <https://doi.org/10.1007/s42979-021-00592-x>.  
[6] W. Chan y L. Sobrin, "Epidemiology and Genetics of Diabetic Retinopathy", en Albert and Jakobiec's Principles and Practice of Ophthalmology. Cham: Springer International Publishing, 2022, pp. 2935–2953. Accedido el 14 de junio de 2023. [En línea]. Disponible: <https://doi.org/10.1007/978-3-030-42634-7_35>.

[7] "Causes and Prevalence of Visual Impairment Among Adults in the UnitedStates", Arch. Ophthalmol., vol. 122, n.º 4, p. 477, abril de 2004. Accedido el 14 de junio de 2023. [En línea]. Disponible: <https://doi.org/10.1001/archopht.122.4.477>.

[8] N. G. Forouhi y N. J. Wareham, "Epidemiology of diabetes", Medicine, vol. 47, n.º 1, pp. 22–27, enero de 2019. Accedido el 14 de junio de 2023. [En línea]. Disponible: <https://doi.org/10.1016/j.mpmed.2018.10.004>.

[9] P. Saeedi et al., "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition", Diabetes Res. Clin. Pract., vol. 157, p. 107843, noviembre de 2019. Accedido el 14 de junio de 2023. [En línea]. Disponible: <https://doi.org/10.1016/j.diabres.2019.107843>.

[10] Z. L. Teo et al., "Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045", Ophthalmology, mayo de 2021. Accedido el 14 de junio de 2023. [En línea]. Disponible: <https://doi.org/10.1016/j.ophtha.2021.04.027>.

[11] W. L. Wong et al., "Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis", Lancet Global Health, vol. 2, n.º 2, febrero de 2014, art. n.º e106-e116. Accedido el 14 de junio de 2023. [En línea]. Disponible: <https://doi.org/10.1016/s2214-109x(13)70145-1>

[12] D. L. Demmin y S. M. Silverstein, "Visual Impairment and Mental Health: Unmet Needs and Treatment Options", Clin. Ophthalmol., Volume 14, pp. 4229–4251, diciembre de 2020. Accedido el 14 de junio de 2023. [En línea]. Disponible: <https://doi.org/10.2147/opth.s258783>.

[13] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs", *J. Amer. Med. Assoc.*, vol. 316, no. 22, pp. 2402-2410, 2016.

[14] R. Gargeya and T. Leng, "Automated identification of diabetic retinopathy using deep learning", *Ophthalmology*, vol. 124, no. 7, pp. 962-969, 2017.

[15] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift" in arXiv:1502.03167, 2015, [online] Available: http://arxiv.org/abs/1502.03167.

[16] C. Szegedy et al., "Going deeper with convolutions", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1-9, Jun. 2015.

[17] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely connected convolutional networks", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2261-2269, Jul. 2017.

[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks", *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, pp. 4510-4520, 2018.

[19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2818-2826, Jun. 2016.

[18] "Kaggle: Your Machine Learning and Data Science Community". Kaggle: Your Machine Learning and Data Science Community. <https://www.kaggle.com/> (accedido el 4 de mayo de 2023).

[19] Fanbyprinciple. (2021). pytorch diabetic retinopathy. *Kaggle*. <https://www.kaggle.com/code/fanbyprinciple/pytorch-diabetic-retinopathy/data>.

[20] Balajiai. (2021). Diabetic Retinopathy Detection using PyTorch. *Kaggle*.