

Estudio del uso de GAN para el balanceo de datos en el conjunto BreakHis y los efectos en la clasificación de tumores de mama

Alfredo Gutiérrez-Alfaro¹, Angel E. Rosales-Morales²,
Andrés Espinal¹, Manuel Ornelas-Rodríguez²,
Marco Sotelo-Figueroa¹, Horacio Rostro-Gonzalez³

¹ Universidad de Guanajuato,
Departamento de Estudios Organizacionales,
División de Ciencias Económico Administrativas,
México

² Instituto Tecnológico de León,
Tecnológico Nacional de México,
México

³ Universidad de Guanajuato,
Departamento de Electrónica,
División de Ingenierías,
México

{a.gutierrezalfaro, aespinal}@ugto.mx,
m22240870@leon.tecnm.mx

Resumen. El conjunto BreakHis es una base de datos de imágenes histopatológicas de tejido de mamario, cuyo propósito es permitir el desarrollo de modelos para la clasificación automática de tumores (malignos y benignos). Las clases principales del conjunto están distribuidas en 5, 429 y 2, 480 imágenes para las clases maligno y benigno, respectivamente. El desbalance de datos en las clases, un problema común en el análisis de imágenes médicas, puede impactar negativamente en el desempeño de modelos de aprendizaje máquina. En este trabajo, se hace uso un modelo generativo (GAN) para el balanceo de imágenes. Posteriormente, se comprueba el comportamiento y desempeño de clasificadores basados en aprendizaje profundo usando los conjuntos no balanceado y balanceado. Los resultados obtenidos son comparados usando cuatro diferentes métricas.

Palabras clave: Conjunto BreakHis, balanceo de datos, clasificación, aumento de datos, GAN.

Study of GAN Usage for Data Balancing in BreakHis Dataset and the Effects in Breast Tumor Classification

Abstract. The BreakHis set is a histopathologic images database of mammary tissue, which purpose is to enable the development of models for automatic

classification of tumors (malignant and benign). The main classes of the set are distributed in 5,429 and 2,480 images for the malignant and benign classes, respectively. Data imbalance in classes, a common problem in medical image analysis, can negatively impact the performance of machine learning models. In this work, a generative model (GAN) is used for image balancing. Subsequently, the behavior and performance of deep learning-based classifiers are tested using the unbalanced and the balanced datasets. The obtained results are compared using four different metrics.

Keywords: BreakHis dataset, data balancing, classification, data augmentation, GAN.

1. Introducción

En la actualidad, los casos de cáncer de mama han ido en aumento, no obstante, gracias a los avances en los métodos de detección temprana, la tasa de mortalidad ha decrementedo considerablemente. De los diversos métodos para el diagnóstico de cáncer de mama, la examinación de biopsias histopatológicas suele ser infalible; este es llevado a cabo por patólogos al analizar láminas de tejido mamario con varios niveles de aumento microscópicos para resaltar las zonas de interés; sin embargo, la interpretación de los patólogos es susceptible a errores humanos por factores como cansancio, distracción, etc [3].

En el área de medicina cada vez es más frecuente la adopción de sistemas para el diagnóstico asistido por computadora con el objetivo de disminuir los riesgos de un mal diagnóstico. En particular, para el caso de detección de cáncer de mama a partir de imágenes histopatológicas existe una base de datos llamada BreakHis [24], la cuál recopila imágenes de biopsias con diferentes niveles de aumento microscópico de tumores de mama benignos y malignos.

Esta base de datos es de uso público con el objetivo de proveer un marco de referencia y para fines de investigación; sin embargo, a pesar de la riqueza de su contenido clínico, BreakHis aún posee anomalías comunes en todos los conjuntos de datos médicos debido a la naturaleza de la enfermedad y limitaciones en las técnicas de adquisición de datos médicos [3].

El conjunto BreakHis tiene una proporción de imágenes del 68,6 % y 31,4 % para las clases maligno y benigno, respectivamente; este desbalance de datos entre las clases es una de las anomalías que se suelen presentar en los conjuntos de datos médicos.

El desbalanceo de datos puede tener efectos no deseados en un sistema de diagnóstico asistido por computadora, como sesgar su capacidad discriminativa hacia las clases más representativa en el conjunto de datos, es decir, la clase menos representativa puede ser considerada como datos atípicos o ruido [3, 22].

Para lidiar con el problema del desbalance de clases, existen diferentes técnicas o métodos. En la literatura se pueden encontrar técnicas como duplicar imágenes pertenecientes a la clase minoritaria, aplicar transformaciones a imágenes existentes (rotaciones, cambios de color o añadir ruido), combinar imágenes de la clase minoritaria para crear nuevas instancias (SMOTE por sus siglas en inglés -Synthetic Minority Oversampling Technique-) [22].

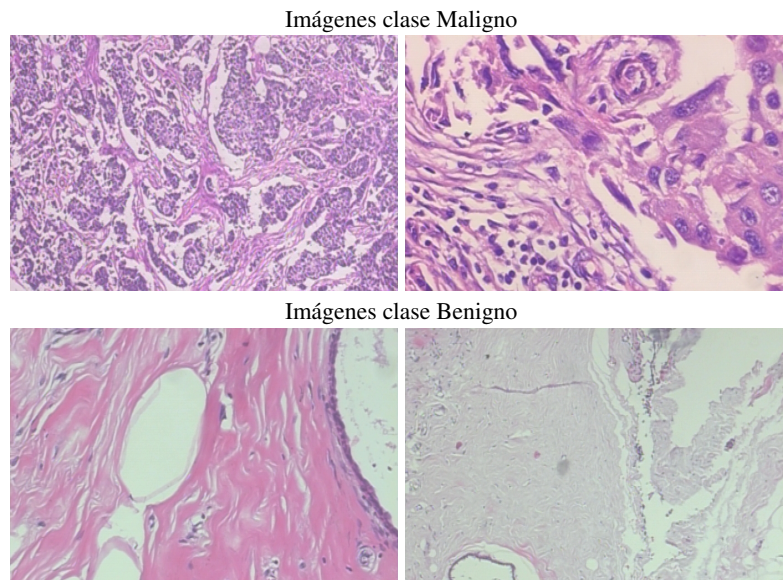


Fig. 1. Muestra de las imágenes del conjunto BreakHis.

Un método de los más recurridos últimamente es el aumento de datos artificiales mediante modelos generativos, en particular usando GAN [9] (por sus siglas en inglés -Generative Adversarial Networks-); las cuales son tratadas dentro de este artículo. Las GAN han sido utilizadas exitosamente en diversos ámbitos, como en el médico [26, 23]; por ejemplo, en [6] se logró mejorar los valores de métricas como la sensibilidad en la clasificación de lesiones en hígado.

No obstante, el uso de GAN en aplicaciones médicas para aumentar la cantidad de datos y mejorar los modelos de clasificación debe ser tratado con cuidado, ya que, los datos generados podrían añadir ruido a los modelos, ocasionando el efecto contrario al de mejorar su desempeño [4].

En el presente artículo se estudia el efecto de desarrollar modelos de clasificación usando el conjunto BreakHis original y una versión balanceada del mismo a través de una GAN; es decir, solo se agregarán imágenes artificiales de la clase benigna. Se realizará una experimentación exhaustiva del desempeño de modelos entrenados con ambos conjuntos de datos, midiendo métricas de exactitud, precisión, recuperación y especificidad, con las cuales se observará el comportamiento de los clasificadores basados en aprendizaje profundo.

La estructura del artículo es la siguiente. La sección 2 describe las características del conjunto BreakHis. El modelo de aprendizaje profundo que se utiliza para la tarea de clasificación, CNN (por sus siglas en inglés -Convolutional Neural Network-), se detalla en la sección 3.

El modelo generativo, GAN, para lograr el balanceo de clases es explicada en la sección 4. En la sección 5 se describe la experimentación realizada y los resultados obtenidos. Finalmente, las conclusiones y trabajo a futuro se mencionan en la sección 6.

Tabla 1. Distribución de las imágenes del conjunto BreakHis.

Aumento	Maligno	Benigno	Total
40×	1,370	625	1,995
100×	1,437	644	2,081
200×	1,390	623	2,013
400×	1,232	588	1,820
Total	5,429	2,480	7,909

2. Conjunto de datos BreakHis

El conjunto de datos BreakHis [24] es una base de datos de imágenes histopatológicas de tejidos mamarios, la cual fue recopilada con el propósito de ser un marco de referencia para el desarrollo de sistemas de clasificación para el diagnóstico asistido por computadora.

El conjunto de datos está dividido principalmente en dos clases de tumores, malignos y benignos; la Figura 1 muestra algunas de las imágenes contenidas en BreakHis para estas clases. Cada clase principal tiene cuatro subclases, siendo carcinoma ductal, carcinoma lobulillar, carcinoma mucinoso y carcinoma papilar para la clase maligno y adenosis, fibroadenoma, tumores filodes y adenoma tubular para la clase benigno; sin embargo, en este trabajo se limitará a trabajar únicamente con las clases principales.

El conjunto de datos está organizado en cuatro niveles de aumento. La Tabla 1 muestra la distribución de datos por nivel de aumento y clases. Enfocándose en las clases principales se puede observar el conjunto BreakHis sufre de desbalance de datos; siendo así que la clase predominante es la de tumores malignos con un total de 5,429 imágenes y la clase con menor presencia la de tumores benignos con 2,480 imágenes.

Dicho desbalance de datos puede repercutir en la generación de modelos de clasificación con comportamientos no deseados, como el hecho de que la clase menos representativa sea vista como casos raros por el modelo y llegue a ser clasificada incorrectamente [22].

Existen trabajos que han usado modelos generativos para lograr un balance de los datos en las clases y efectuar la tarea de clasificación de tumores reportando un incremento en la exactitud de los modelos [15, 7]; sin embargo, a diferencia de los trabajos previamente mencionados, en esta investigación se busca conocer el efecto que tiene el agregar datos artificiales mediante otras métricas como la precisión y recuperación que se enfocan en la clase positiva (maligno) y que no ha sido modificada y la especificidad que se enfoca en la clase negativa (benigno) la cual será balanceada mediante el uso de una GAN.

3. CNN: Convolutional Neural Network

Las CNN son un tipo de arquitectura de redes neuronales ampliamente usadas en tareas de visión por computadora, reconocimiento y clasificación de imágenes [25]; siendo la última el tipo de tarea que se resolverán en este trabajo. La Fig. 2 muestra un bosquejo general de la arquitectura y funcionamiento de una CNN.

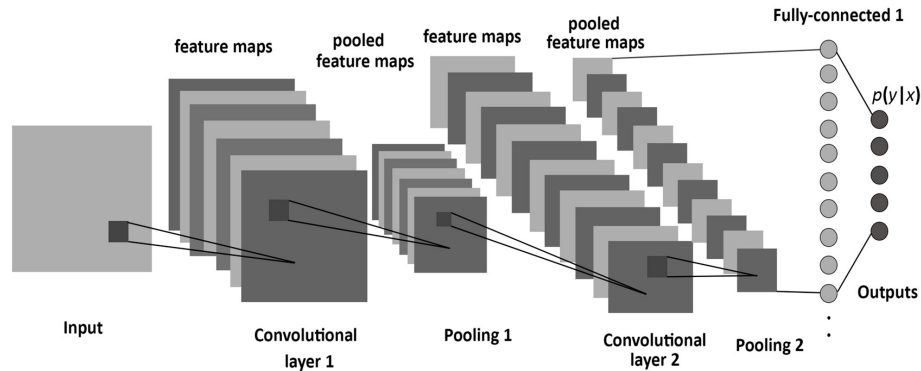


Fig. 2. Arquitectura de CNN para clasificación [1].

Estas redes trabajan mediante un proceso conocido como convolución, donde se aplican filtros para identificar y extraer patrones y características de las imágenes. Estos filtros son matrices de pesos que se aplican a lo largo de cada posición de la imagen, multiplicando los elementos de la imagen con los valores del filtro y sumando los resultados.

Este proceso se repite para cada posición dentro de la imagen, generando una nueva matriz de características; conocido como mapa de características. A medida que se aplican capas convolucionales sucesivas, la red va aprendiendo representaciones cada vez más complejas de la imagen. Las capas posteriores suelen utilizar filtros más grandes para capturar características globales, como formas grandes o patrones complejos [8, 18].

Es común que una capa convolucional sea precedida por una capa conocida como pooling, la cuál sirve para extraer las características más relevantes del mapa generado por la capa convolucional; como resultado, el pooling proporciona un mapa de características importantes más pequeño. Finalmente, en la última capa se obtienen las probabilidades para cada clase y con base a ellas se realiza la clasificación [8].

4. GAN: Generative Adversarial Network

Las GAN [10] pertenecen a los modelos generativos, estas están formadas por dos redes que compiten entre sí; un generador y un discriminador. El generador tiene la tarea de tratar de engañar al discriminador, sintetizando información a partir de ruido aleatorio; es decir, el generador debe crear muestras que son lo suficientemente parecidas a las reales como para engañar al discriminador.

Mientras que el discriminador tiene la tarea de identificar si los datos son reales o falsos; es decir, creados artificialmente por el generador. Durante el proceso de entrenamiento de una GAN, se usa como función de pérdida la Ecuación 1, propuesta originalmente en [10]:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))], \quad (1)$$

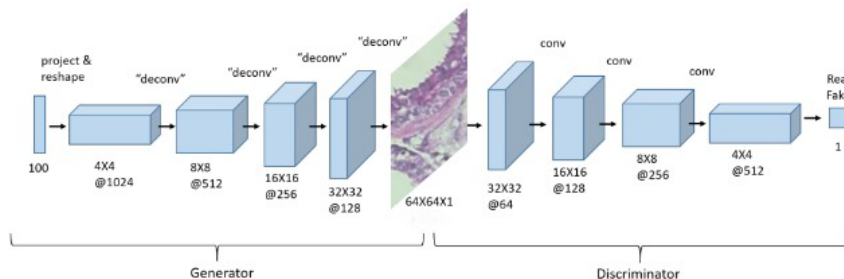


Fig. 3. Arquitectura de la DCGAN. (Imagen modificada de [6]).

donde $D(x)$ representa la estimación del discriminador para una muestra que proviene de una distribución de datos reales, mientras que $D(G(z))$ es la estimación del discriminador de que una muestra generada por $G(z)$ es real.

El generador busca minimizar la función, mientras que el discriminador busca maximizarla. La primera parte de la función $\mathbb{E}x \sim p_{data}(x)[\log D(x)]$ corresponde a la esperanza matemática del logaritmo de la probabilidad de que el valor x sea real. La segunda parte, $\mathbb{E}z \sim p_z(z)[\log(1 - D(G(z)))]$, corresponde al valor esperado del logaritmo de la probabilidad asignada a la muestra generada $G(z)$ de que no sea real.

Las GAN se han utilizado en una amplia variedad de tareas y han dado lugar a múltiples variantes para tareas específicas [12, 17]. En el campo de la visión por computadora, se han utilizado principalmente para generar imágenes y tienen aplicaciones como la traducción de imagen a imagen [16], la generación de rostros [19], la generación de señales, entre otras que no están necesariamente relacionadas con imágenes [17].

En el ámbito médico, las GAN han tenido diversos usos [26], como la generación de imágenes, la reconstrucción, la segmentación, entre otras. La generación de imágenes es una de las categorías más populares [20], ya que permite generar nuevas imágenes en comparación con otras técnicas de aumento de datos.

Una de las múltiples aplicaciones de las GAN consiste en la generación de imágenes para balancear conjuntos de datos [22], lo cual puede ser de gran utilidad para compensar la falta de datos específicos de una clase en particular. Este enfoque puede ser relevante en conjuntos de datos médicos, donde la obtención de nueva información puede resultar complicada.

4.1. DCGAN: Deep Convolutional Generative Adversarial Network

Las DCGAN (por sus siglas en inglés -Deep Convolutional Generative Adversarial Networks-) son una modificación de la GAN original, la cual ha sido utilizada en arquitecturas actuales [21]. En lugar de emplear una capa completamente conectada, la DCGAN utiliza capas convolucionales, lo que permite la generación mejores imágenes.

En la Figura 3, se muestra en la parte izquierda la arquitectura del generador utilizado en la DCGAN, comenzando con una entrada de tamaño 100, que corresponde al tamaño del vector de ruido, se utilizan capas deconvolucionales para aumentar el tamaño de las salidas en cada capa hasta obtener una salida de $64 \times 64 \times 3$ [11, 21].

Layer (type)	Output Shape	Param #	Layer (type)	Output Shape	Param #
ConvTranspose2d-1	[-1, 512, 4, 4]	819,200	Conv2d-1	[-1, 64, 32, 32]	3,072
BatchNorm2d-2	[-1, 512, 4, 4]	1,024	LeakyReLU-2	[-1, 64, 32, 32]	0
LeakyReLU-3	[-1, 512, 4, 4]	0	Conv2d-3	[-1, 128, 16, 16]	131,072
ConvTranspose2d-4	[-1, 256, 8, 8]	2,097,152	BatchNorm2d-4	[-1, 128, 16, 16]	256
BatchNorm2d-5	[-1, 256, 8, 8]	512	LeakyReLU-5	[-1, 128, 16, 16]	0
LeakyReLU-6	[-1, 256, 8, 8]	0	Conv2d-6	[-1, 256, 8, 8]	524,288
ConvTranspose2d-7	[-1, 128, 16, 16]	524,288	BatchNorm2d-7	[-1, 256, 8, 8]	512
BatchNorm2d-8	[-1, 128, 16, 16]	256	LeakyReLU-8	[-1, 256, 8, 8]	0
LeakyReLU-9	[-1, 128, 16, 16]	0	Conv2d-9	[-1, 512, 4, 4]	2,097,152
ConvTranspose2d-10	[-1, 64, 32, 32]	131,072	BatchNorm2d-10	[-1, 512, 4, 4]	1,024
BatchNorm2d-11	[-1, 64, 32, 32]	128	LeakyReLU-11	[-1, 512, 4, 4]	0
LeakyReLU-12	[-1, 64, 32, 32]	0	Conv2d-12	[-1, 1, 1, 1]	8,192
ConvTranspose2d-13	[-1, 3, 64, 64]	3,072	Sigmoid-13	[-1, 1, 1, 1]	0
Tanh-14	[-1, 3, 64, 64]	0			
Total params: 3,576,704			Total params: 2,765,568		
Trainable params: 3,576,704			Trainable params: 2,765,568		
Non-trainable params: 0			Non-trainable params: 0		

(a) Arquitectura del generador

(b) Arquitectura del discriminador

Fig. 4. Arquitecturas de los modelos de la GAN.

El discriminador, mostrado en la parte derecha de la Figura 3, tiene una arquitectura similar, pero de forma inversa, ya que en este caso, el discriminador recibe una imagen de tamaño $64 \times 64 \times 3$ y de salida tiene la predicción de si dicha entrada es real o falsa; esta red es básicamente una CNN como se explicó en la Sección 3.

5. Experimentos y resultados

En esta sección se describe el procedimiento experimental y los resultados obtenidos. El desarrollo de los modelos fue realizado usando el lenguaje de programación Python y la librería de PyTorch.

5.1. Generación de imágenes para balanceo de datos

El balanceo de datos se enfocará a igualar la cantidad de imágenes de la clase benigno a las que conforman la clase maligno. Dicho proceso consistirá en la creación de 2,949 imágenes artificiales que asemejen lo más posible las imágenes reales contenidas en la clase benigno mediante el entrenamiento y uso de una DCGAN.

La DCGAN usada está basada principalmente en el modelo publicado en [21] con modificaciones propuestas en [9]. Las Figuras 4a y Figura 4b muestran la configuración de los modelos que conforman la DCGAN; generador y discriminador respectivamente.

Para el entrenamiento de la DCGAN, será necesario preprocesar las imágenes, esto incluye su normalización a través del método *Z-Score*. La Tabla 2 muestra los hiperparámetros de los modelos usados para la DCGAN.

En la Figura 5 se muestran algunas imágenes creadas a partir del generador de la DCGAN utilizada.

5.2. Clasificación de tumores

Para la clasificación de tumores, se proponen dos configuraciones experimentales. La primera configuración consiste en obtener modelos a partir del conjunto BreakHis desbalanceado, es decir, el conjunto original.

Tabla 2. Hiperparámetros de la DCGAN.

DCGAN	Tasa de aprendizaje	Adam, Beta	Épocas	Batch
Generador	0.0001	0.5, 0.999	700	512
Discriminador	0.0001	0.5, 0.999	700	512

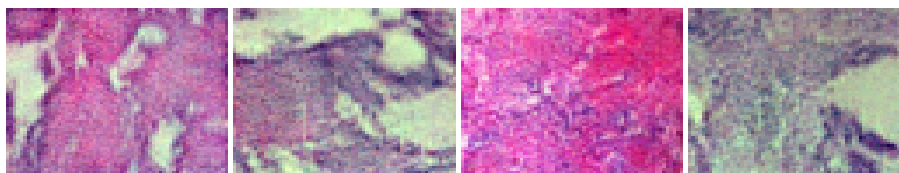


Fig. 5. Imágenes artificialmente generadas de la clase benigna.

Para la segunda configuración se utilizará el conjunto balanceado con las imágenes complementarias de la clase benigno creadas por la DCGAN.

Para ambas configuraciones experimentales se realizaron 30 entrenamientos independientes, en los que para cada uno se crearon particiones aleatorias de conjuntos de entrenamiento, validación y prueba; usando el 64 %, 16 % y 20 % respectivamente para cada conjunto.

La cantidad de 10 épocas de entrenamiento para ambas configuraciones experimentales fue seleccionada empíricamente analizando el comportamiento de la función de pérdida para el conjunto de entrenamiento y validación.

Para poder medir el desempeño de los clasificadores basados en DCNN se usarán cuatro métricas, para las cuales es necesario definir que en este trabajo la clase positiva corresponde a los tumores malignos y la clase negativa a los tumores benignos. Para la cantidad de datos positivos cuyas predicciones son positivas se le denomina TP (por sus siglas en inglés -True Positive-), la cantidad de datos positivos con predicciones negativas entonces se conoce como FN (por sus siglas en inglés -False Negative-).

Por otra parte, la cantidad de datos negativos, con predicciones negativas, se le conoce como TN (por sus siglas en inglés -True Negative-), de lo contrario, la cantidad de datos negativos con predicciones positivas se le nombra como FP (por sus siglas en inglés -False Positive-) [2]. A continuación, se describen las métricas a usar y el objetivo de las mismas [5].

La exactitud (*acc*, Ecuación 2) indica la proporción de los datos correctamente clasificados. Esta métrica nos permite conocer de manera general el rendimiento del modelo sin poner énfasis en alguna clase en particular:

$$acc = \frac{TP + TN}{TP + FN + FP + TN}. \quad (2)$$

La precisión (*prec*, Ecuación 3) indica la proporción de datos positivos reales entre los datos que el modelo identifica como positivos:

$$prec = \frac{TP}{TP + FP}. \quad (3)$$

Tabla 3. Tabla de experimentos.

Exp	No balanceado				Balanceado			
	acc	prec	rec	spec	acc	prec	rec	spec
1	0.848	0.861	0.931	0.658	0.875	0.863	0.897	0.851
2	0.839	0.887	0.881	0.745	0.868	0.814	0.961	0.771
3	0.851	0.857	0.942	0.643	0.860	0.795	0.977	0.737
4	0.851	0.886	0.903	0.734	0.864	0.907	0.817	0.912
5	0.848	0.881	0.904	0.722	0.887	0.858	0.932	0.839
6	0.832	0.823	0.967	0.525	0.879	0.859	0.914	0.843
7	0.846	0.888	0.891	0.743	0.878	0.877	0.885	0.870
8	0.859	0.896	0.902	0.761	0.883	0.860	0.922	0.843
9	0.829	0.910	0.836	0.811	0.883	0.841	0.951	0.813
10	0.856	0.872	0.929	0.689	0.878	0.834	0.950	0.803
11	0.836	0.827	0.965	0.539	0.864	0.823	0.936	0.789
12	0.849	0.876	0.912	0.705	0.874	0.822	0.963	0.782
13	0.853	0.883	0.910	0.724	0.859	0.923	0.789	0.931
14	0.848	0.899	0.880	0.774	0.887	0.875	0.908	0.864
15	0.856	0.895	0.898	0.759	0.876	0.900	0.851	0.901
16	0.837	0.896	0.866	0.770	0.882	0.874	0.897	0.865
17	0.851	0.868	0.927	0.678	0.884	0.849	0.940	0.825
18	0.846	0.854	0.940	0.633	0.875	0.869	0.888	0.861
19	0.848	0.894	0.886	0.759	0.881	0.839	0.948	0.811
20	0.813	0.925	0.795	0.853	0.870	0.840	0.920	0.817
21	0.850	0.891	0.894	0.751	0.844	0.780	0.968	0.715
22	0.846	0.874	0.909	0.701	0.874	0.832	0.945	0.800
23	0.842	0.838	0.957	0.579	0.862	0.899	0.823	0.903
24	0.839	0.879	0.891	0.720	0.872	0.907	0.835	0.911
25	0.841	0.907	0.860	0.799	0.878	0.834	0.950	0.803
26	0.834	0.915	0.840	0.822	0.872	0.891	0.854	0.891
27	0.851	0.858	0.941	0.645	0.881	0.894	0.870	0.893
28	0.847	0.894	0.884	0.761	0.874	0.828	0.950	0.794
29	0.841	0.837	0.957	0.575	0.889	0.872	0.917	0.860
30	0.853	0.875	0.919	0.701	0.881	0.841	0.946	0.814
min	0.813	0.823	0.795	0.525	0.844	0.780	0.789	0.715
max	0.859	0.925	0.967	0.853	0.889	0.923	0.977	0.931
μ	0.844	0.878	0.902	0.708	0.874	0.856	0.909	0.836
σ	0.011	0.027	0.044	0.088	0.011	0.037	0.053	0.057

La recuperación (rec, Ecuación 4) indica la proporción de datos positivos correctamente clasificados. Puede entenderse como un estimador de la probabilidad de clasificar correctamente un dato positivo arbitrario:

$$rec = \frac{TP}{TP + FN}. \quad (4)$$

La especificidad (spec, Ecuación 5) indica la proporción de datos negativos correctamente clasificados. Puede entenderse como un estimador de la probabilidad de clasificar correctamente un dato negativo arbitrario:

$$spec = \frac{TN}{FP + TN}. \quad (5)$$

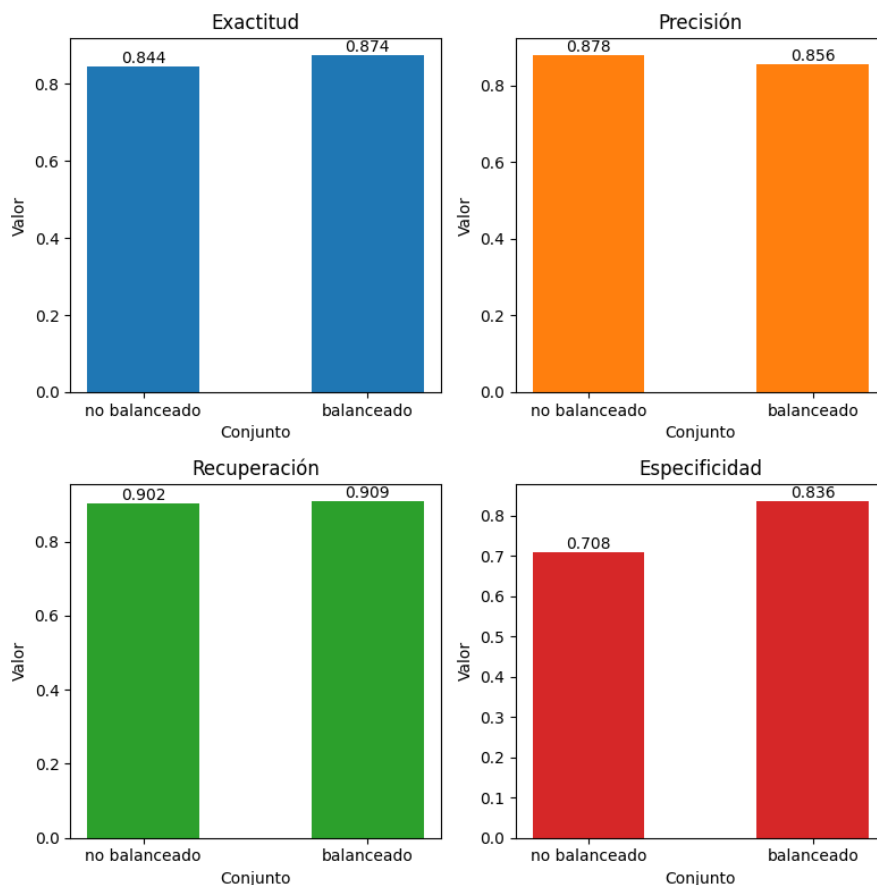


Fig. 6. Comparación de medias de las métricas de los conjuntos no balanceados y balanceados.

La Tabla 3 muestra los resultados de 30 ejecuciones utilizando el conjunto BreakHis original (parte izquierda) y el conjunto balanceado mediante GAN (parte derecha). Para cada ejecución se reportan los valores obtenidos por las métricas mencionadas usando conjuntos de pruebas, también para cada una de ellas se calculan los estadísticos mínimo, máximo, media y desviación estándar.

Con respecto al estado del arte, generalmente se reportan porcentajes de exactitud entre el 80 % y 90 % [24, 3], otras métricas como las mencionadas en este trabajo no son utilizadas en la literatura revisada. En la Tabla 3 se pueden observar exactitudes promedio 84,4 % y 87,4 % para modelos entrenados con los conjuntos no balanceado y balanceado, respectivamente.

La Figura 6 muestra gráficas de barra para comparar los valores medios de las métricas utilizadas para ambas configuraciones experimentales. La gráfica a destacar es la que involucra la especificidad, ya que en esta es donde se ve reflejado de cierta manera el impacto del balanceo de la clase negativa, teniendo un incremento aproximado del 13 % en su valor con respecto a los resultados del conjunto no balanceado.

Los resultados de exactitud y recuperación no se ven mayormente afectados y los de la precisión muestran un ligero decremento porcentual de la experimentación con datos balanceados con respecto al de los datos no balanceados.

6. Conclusiones y trabajo futuro

El uso del DCGAN puede producir imágenes que ayuden a balancear conjuntos de datos, aunque estas pueden introducir problemas de ruido [22], como se pudo observar en los resultados de la precisión de la Figura 6.

A pesar de que las imágenes sintéticas pueden no llegar a tener la misma calidad o características a las imágenes reales, estas pueden contribuir a mejorar en el rendimiento de clasificadores cuando existe una clase con pocos datos, caso común en el ámbito médico [17]; como se pudo observar en los resultados de la exactitud, recuperación y especificidad de la Figura 6.

Aunque los resultados obtenidos en el trabajo demuestran que el uso de una DCGAN para el balanceo de datos del conjunto BreakHis puede llegar a mejorar la exactitud, la especificidad y la recuperación, existen áreas en las que se puede trabajar para intentar conseguir mejores resultados a la hora de realizar clasificación.

Una de las posibilidades es explorar otras técnicas como SMOTE, así como arquitecturas emergentes de GAN específicas para el balanceo, como lo es BAGAN [14], que se enfoca en generar imágenes para el balanceo de conjuntos de datos, o arquitecturas GAN condicionales, en las que se puede generar imágenes de una clase en específico. Además de explorar métricas que permitan evaluar la calidad de las imágenes generadas como la Fréchet Inception Distance [13].

Otra área de trabajo a futuro podría ser la generación de imágenes sintéticas de subclases específicas en lugar de trabajar únicamente con un enfoque binario. Por ejemplo, en tareas de clasificación de tumores, sería más útil generar imágenes de diferentes tipos de tumores y en diferentes resoluciones, lo que podría mejorar la capacidad del modelo para generalizar y clasificar con mayor rendimiento los diferentes tipos de tumores presentes en el conjunto de datos BreakHis [3].

Referencias

1. Albelwi, S., Mahmood, A.: A framework for designing the architectures of deep convolutional neural networks. *Entropy*, vol. 19, no. 6 (2017) doi: 10.3390/e19060242
2. Alpaydin, E.: *Introduction to machine learning*, The MIT Press (2010)
3. Benhammou, Y., Achchab, B., Herrera, F., Tabik, S.: BreakHis based breast cancer automatic diagnosis using deep learning: Taxonomy, survey and insights. *Neurocomputing*, vol. 375, pp. 9–24 (2020) doi: 10.1016/j.neucom.2019.09.044
4. Bissoto, A., Valle, E., Avila, S.: GAN-based data augmentation and anonymization for skin-lesion analysis: A critical review. *Computer Vision and Pattern Recognition*, pp. 1847–1856 (2021) doi: 10.1109/cvprw53098.2021.00204
5. Flach, P.: *Machine learning: The art and science of algorithms that make sense of data* (2012)
6. Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: Synthetic data augmentation using GAN for improved liver lesion classification. In: *2018 IEEE 15th International Symposium on Biomedical Imaging*, pp. 289–293 (2018) doi: 10.1109/isbi.2018.8363576

7. Gheshlaghi, S. H., Enoch-Kan, C. N., Ye, D.: Breast cancer histopathological image classification with adversarial image synthesis. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3387–3390 (2021) doi: 10.1109/EMBC46164.2021.9630678
8. Glassner, A.: Deep learning: A visual approach. No Starch Press, (2021) <https://books.google.com.mx/books?id=NgTyDwAAQBAJ>
9. Goodfellow, I.: NIPS 2016 tutorial: Generative adversarial networks. Cornell University, (2016) doi: 10.48550/arXiv.1701.00160
10. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. *Advances in Neural Information Processing Systems*, vol. 27 (2014)
11. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T.: Recent advances in convolutional neural networks. *Pattern Recognition*, vol. 77, pp. 354–377 (2018) doi: 10.1016/j.patcog.2017.10.013
12. Gui, J., Sun, Z., Wen, Y., Tao, D., Ye, J.: A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3313–3332 (2023) doi: 10.1109/TKDE.2021.3130191
13. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local nash equilibrium. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, vol. 30, pp. 6629–6640 (2017)
14. Huang, G., Jafari, A.: Enhanced balancing GAN: Minority-class image generation. *Neural Computing and Applications*, vol. 35, no. 7, pp. 5145–5154 (2020) doi: 10.1007/s00521-021-06163-8
15. Huynh-Thuy, M. B., Hoang, V. T.: Fusing of deep learning, transfer learning and GAN for breast cancer histopathological image classification. *Advanced Computational Methods for Knowledge Engineering*, vol. 1121, pp. 255–266 (2019) doi: 10.1007/978-3-030-38364-023
16. Isola, P., Zhu, J. Y., Zhou, T., Efros, A. A.: Image-to-image translation with conditional adversarial networks. *Computer Vision and Pattern Recognition*, pp. 5967–5976 (2017) doi: 10.1109/cvpr.2017.632
17. Jabbar, A., Li, X., Omar, B.: A survey on generative adversarial networks: Variants, applications, and training. *ACM Computing Surveys*, vol. 54, no. 157, pp. 1–49 (2021) doi: 10.1145/3463475
18. Jernelv, I. L., Hjelme, D. R., Matsuura, Y., Aksnes, A.: Convolutional neural networks for classification and regression analysis of one-dimensional spectral data (2020) doi: 10.48550/arxiv.2005.07530
19. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. *Computer Vision and Pattern Recognition*, pp. 4396–4405 (2019) doi: 10.1109/cvpr.2019.00453
20. Li, X., Jiang, Y., Rodriguez-Andina, J. J., Luo, H., Yin, S., Kaynak, O.: When medical images meet generative adversarial network: Recent development and research opportunities. *Discover Artificial Intelligence*, vol. 1, no. 5 (2021) doi: 10.1007/s44163-021-00006-0
21. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks (2015) doi: 10.48550/arXiv.1511.06434
22. Sampath, V., Maurtua, I., Aguilar-Martín, J. J., Gutierrez, A.: A survey on generative adversarial networks for imbalance problems in computer vision tasks. *Journal of Big Data*, vol. 8, no. 27 (2021) doi: 10.1186/s40537-021-00414-0
23. Shorten, C., Khoshgoftaar, T. M.: A survey on image data augmentation for deep learning. *Journal of Big Data*, vol. 6, no. 60 (2019) doi: 10.1186/s40537-019-0197-0

24. Spanhol, F. A., Oliveira, L. S., Petitjean, C., Heutte, L.: A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462 (2016) doi: 10.1109/TBME.2015.2496264
25. Yamashita, R., Nishio, M., Gian-Do, R. K., Togashi, K.: Convolutional neural networks: An overview and application in radiology. *Insights Into Imaging*, vol. 9, no. 4, pp. 611–629 (2018) doi: 10.1007/s13244-018-0639-9
26. Yi, X., Walia, E., Babyn, P.: Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, vol. 58 (2019) doi: 10.1016/j.media.2019.101552