



UNIVERSIDAD DE GUANAJUATO

DIVISION DE CIENCIAS ECONOMICO ADMINISTRATIVAS

“INTELIGENCIA ARTIFICIAL GENERATIVA BASADA EN POCAS MUESTRAS PARA LA SÍNTESIS DE IMÁGENES MÉDICAS: UN ESTUDIO EXPLORATORIO”

INFORME DE RESULTADOS

PRESENTA

EMILIO ALEJANDRO PEREZ CERRILLO

ASESOR

DR. ANDRÉS ESPINAL JIMÉNEZ

GUANAJUATO, GTO.

AGOSTO 2024

Introducción

Desde sus inicios en las décadas de 1950 a 1970, la Inteligencia Artificial (IA) buscaba apoyar en la toma de decisiones basados en datos que previamente solo los humanos eran capaces de hacer [3].

A pesar de que debido a limitaciones técnicas por décadas no fue viable aplicar herramientas de IA, la medicina es uno de los campos de la ciencia que más se ha interesado y beneficiado de aplicar dichos conocimientos. Desde predicción de comportamiento de proteínas complejas [2], creación de modelos que ayudan a administradores de sistemas de salud a tomar decisiones que mejoren la calidad de vida de los usuarios [4] y análisis de imágenes para el diagnóstico de padecimientos [1].

El cáncer de mama es el tipo cáncer que más provoca muertes en mujeres en México y el tipo de cáncer más común en mujeres a nivel mundial [6]. Por esto, la detección oportuna es vital y la IA podría ser una herramienta formidable para lograrlo.

Para apoyar el entrenamiento de futuros modelos de IA, en 2016 se publicó el conjunto de datos conocido como BreakHis. Este conjunto se compone por 7909 imágenes clínicamente representativas de biopsias de tejidos tumorosos utilizando diferentes magnitudes. Del total de imágenes se cuenta con 2,480 muestras de tejido del tipo benigno y 5,429 del tipo maligno [5].

A pesar de la utilidad que BreakHis ofrece, este conjunto de datos cuenta con una clase sobre representada, problema conocido como desbalanceo de clases. Esta problemática puede generar sesgos en los modelos resultando en errores en las predicciones de las clases minoritarias [7].

Una de las técnicas que han mostrado mejoras en el desempeño de los modelos al corregir este problema es el aumento de datos. Este proceso consiste en generar muestras nuevas de la clase minoritaria para alcanzar un equilibrio en la representación [8]. En el caso de conjuntos de datos compuestos por imágenes, estas pueden ser transformadas mediante rotaciones o volteos, o bien, generando imágenes sintéticas mediante otros modelos de Aprendizaje Profundo (AP).

Un modelo de AP utilizado para generar imágenes sintéticas es el conocido como Generative Adversarial Networks (GAN). Este consiste en dos redes neuronales que compiten entre sí, una para generar imágenes sintéticas a partir de vectores de ruido y otra cuyo objetivo es detectar imágenes sintéticas de las reales. Bajo esta competencia ambos modelos mejoran y posteriormente puede utilizarse la red generadora para obtener imágenes nuevas [10].

Sin embargo, las GAN presentan un problema crucial y es que para un entrenamiento que proporcione resultados de calidad se requiere una cantidad de muestras elevada, misma que no siempre se encuentra disponible. Para sobrellevar este obstáculo posteriormente se propuso el modelo conocido como Few Shot GAN (FSGAN) que mediante una muestra más reducida y los pesos de una red previamente entrenada ha demostrado resultados satisfactorios [9]

Este proyecto se centra en el análisis de resultados de la aplicación de un FSGAN con el objetivo de lograr el balanceo de clases del conjunto BreakHis.

Marco teórico

El conjunto de datos utilizado fue el antes descrito BreakHis, mismo que fue procesado para facilitar su uso. Se trabajó con una implementación de la FSGAN de Robb antes mencionada, utilizando el lenguaje Python 3.7 con el apoyo de librerías como tensorflow, keras y scikit-learn que ofrecen herramientas previamente implementadas y permiten centrarse en el análisis del experimento. A través de la librería de keras se utilizó la red neuronal conocida como EfficientNet [11]. Esta red se encuentra entrenada por miles de imágenes y mediante la técnica de transferencia de conocimiento permite crear clasificadores de imágenes con la cantidad de clases que se requiera.

Metodología

La metodología que se siguió una vez que se contaba con el conjunto de datos en las condiciones necesarias fue la siguiente:

- 1) Calcular la cantidad de imágenes a generar según la clase. Como se mencionó durante la introducción, debido a que la clase benigna se encuentra subrepresentada el objetivo fue generar artificialmente más muestras para mitigar esta situación. Para esto, se calculó la cantidad de muestras artificiales necesarias para que las subclases del tipo benigno contaran con la misma cantidad.

Originalmente se contaban con la siguiente cantidad de muestras.

	Subclase del tipo benigno
--	---------------------------

Magnificación	Adenosis	Fibroadenoma	Phyllodes tumor	Tubular adenoma
40X	114	253	109	149
100X	113	260	121	150
200X	111	264	108	140
400X	106	225	115	130

Por lo que la cantidad de imágenes a generar sigue la siguiente distribución.

	Subclase del tipo benigno			
Magnificación	Adenosis	Fibroadenoma	Phyllodes tumor	Tubular adenoma
40X	139	0	144	104
100X	147	0	139	110
200X	153	0	156	124
400X	119	0	110	95

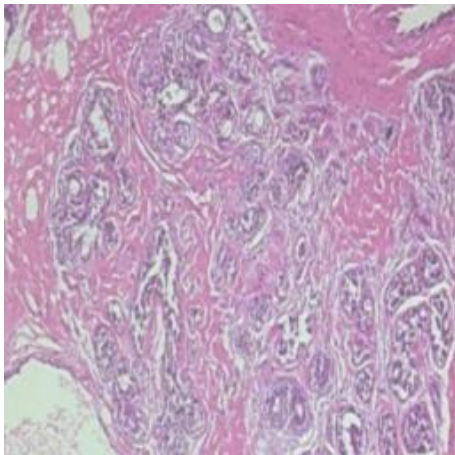
De manera que cada una de las subclases del tipo benigno cuenten con la misma representación.

- 2) Generar imágenes mediante la FSGAN. Haciendo uso de una implementación de la red antes mencionada, se logró generar las muestras necesarias y se incluyeron junto a las muestras originales para lograr un conjunto de datos al que para propósitos de este reporte llamamos BreakHisV2, la cantidad de imágenes de cada subclase se detalla a continuación.

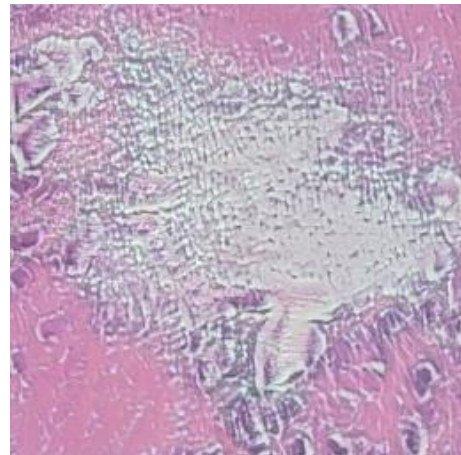
	Subclases del tipo benigno	Subclases del tipo maligno
--	----------------------------	----------------------------

Magnificación	Adenosis	Fibroadenoma	Phylloides tumor	Tubular adenoma	Ductal Carcinoma	Lobular Carcinoma	Mucinous Carcinoma	Papillary Carcinoma
40X	253	253	253	253	864	156	205	145
100X	260	260	260	260	903	170	222	142
200X	264	264	264	264	896	163	196	135
400X	225	225	225	225	788	137	169	138

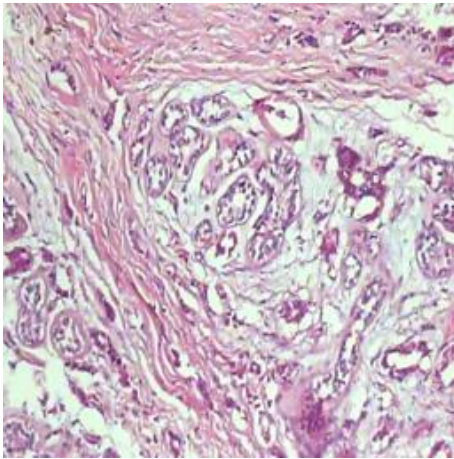
A continuación, se muestra un par de ejemplos de imágenes generadas junto a sus contrapartes originales.



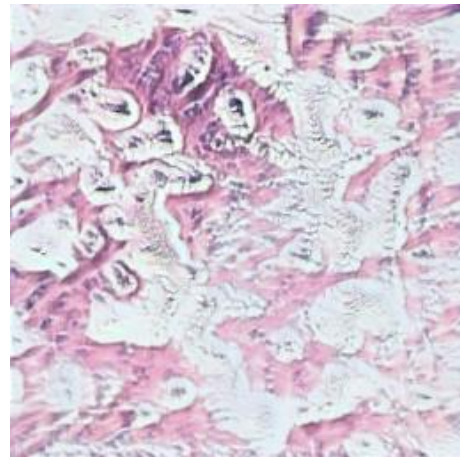
Clase benigna, subclase adenoma, magnificación 40X. Muestra original.



Clase benigna, subclase adenoma, magnificación 40X. Muestra artificial.



Clase benigna, subclase tubular
adenoma, magnificación 100X.
Muestra original.



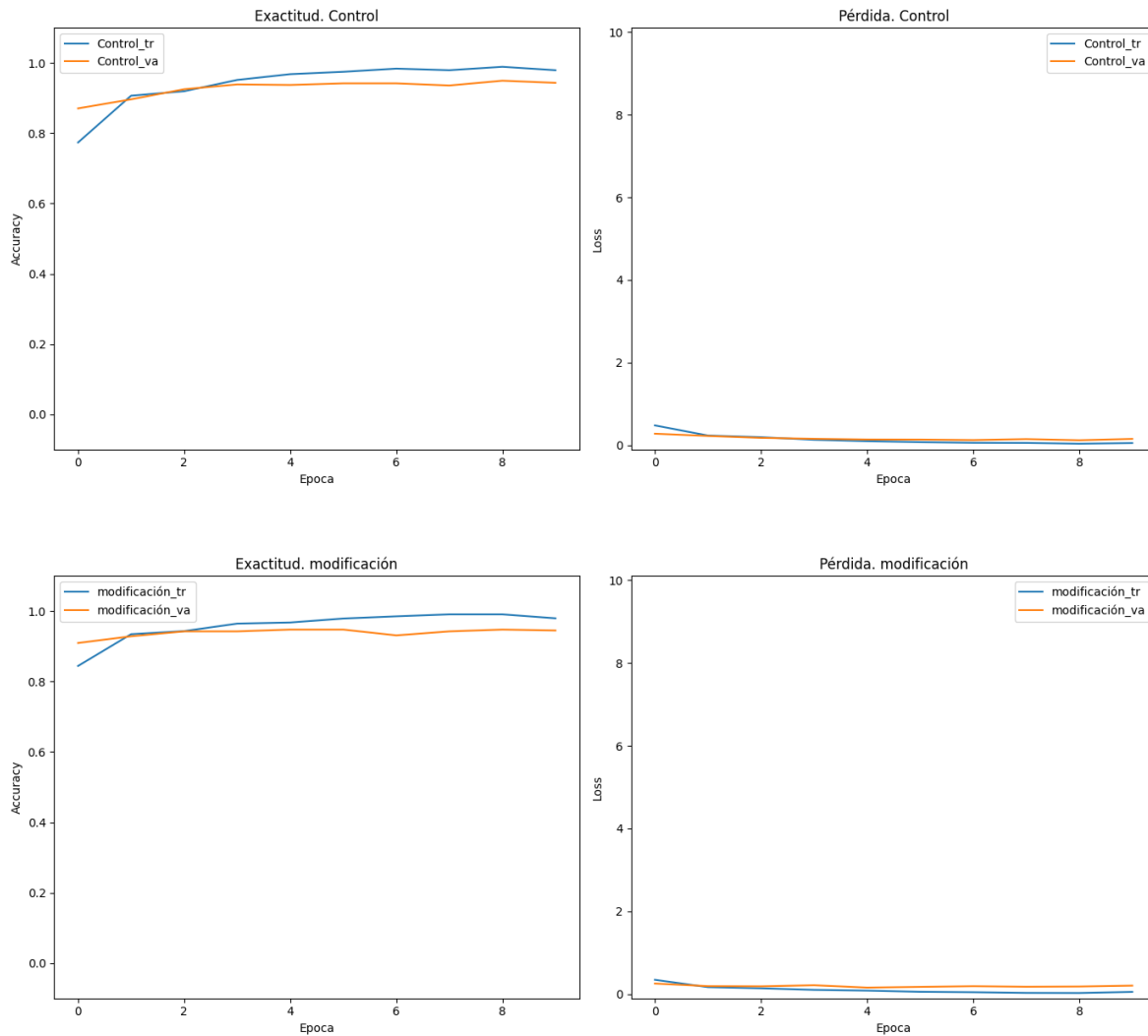
Clase benigna, subclase tubular
adenoma, magnificación 100X.
Muestra artificial.

- 3) Obtener clasificador de imágenes. Gracias a la librería de Keras la implementación del clasificador es muy sencilla a través de EfficientNet. Se importaron los pesos de dicha red para utilizar la transferencia de ese aprendizaje y aplicarlo a la clasificación de muestras de nuestro conjunto de datos BreakHisV2.
- 4) Generar pruebas. Utilizando diversas métricas, evaluamos el desempeño del clasificador tanto con la primera versión del conjunto de datos como con la modificación que incluye las imágenes generadas, enfocándonos en dos casos principales, uno en el que se le pide decidir si se trata de una muestra benigna o maligna y otro en el que se le pide distinguir entre cada una de las subclases de ambos tipos.

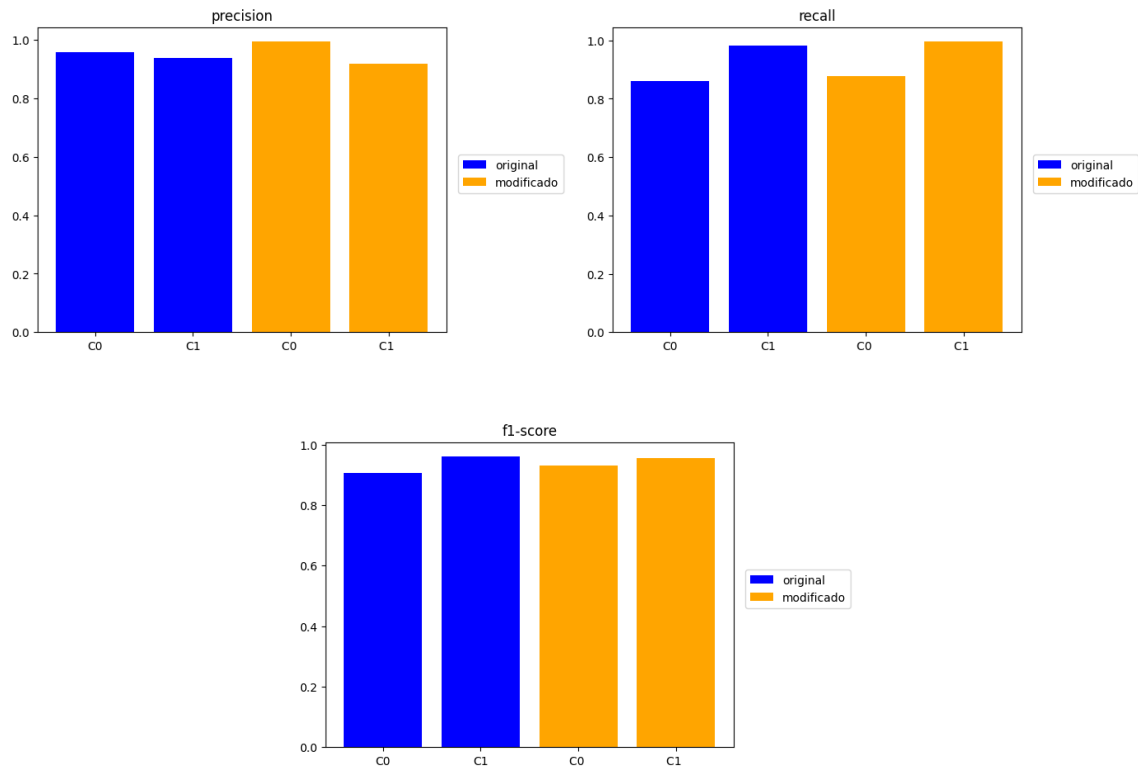
Resultados

Caso biclase.

En las siguientes gráficas se puede apreciar que tanto el conjunto de datos de control como nuestra modificación tienen un buen desempeño en la exactitud y la pérdida durante el entrenamiento.



Sin embargo, en otras métricas como la precisión, el ratio de verdaderos positivos y el puntaje de confianza f-1 nuestra modificación presenta un mejor desempeño.

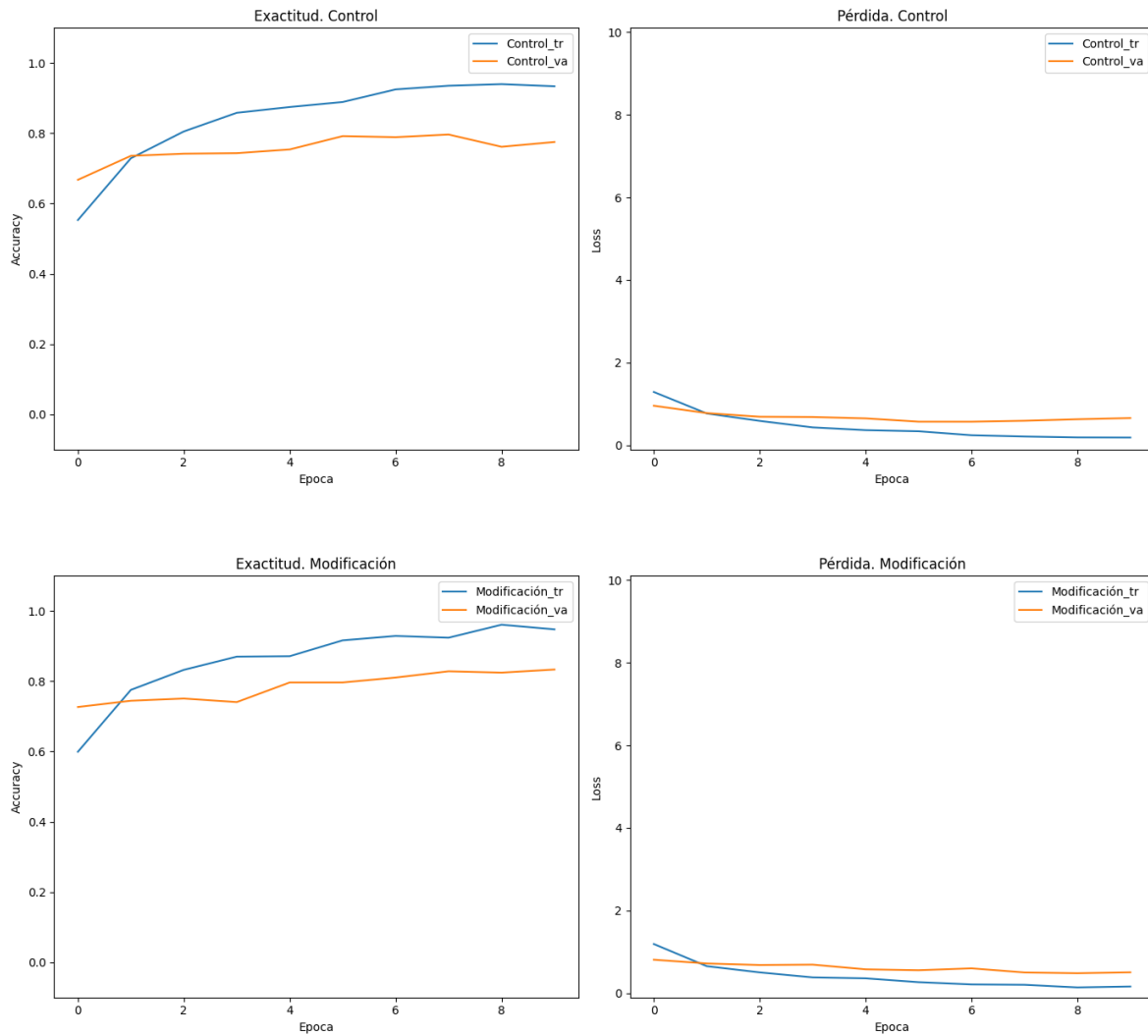


Cuantitativamente, la media ponderada de estas métricas toma la siguiente distribución.

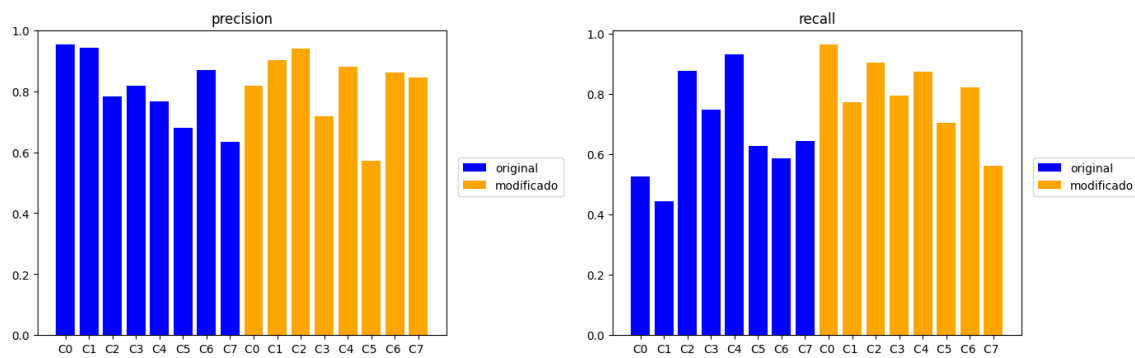
	Original	Modificación
Precisión	0.9444	0.9491
Ratio de verdaderos positivos	0.9438	0.9453
Puntaje de confianza f-1	0.9429	0.9447

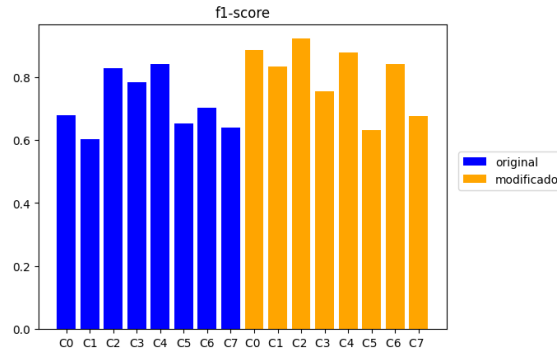
Caso multiclase.

De manera similar, para el caso multiclase, ambos conjuntos de datos presentan niveles de exactitud y pérdida similares.



Y la mejora de nuestra modificación con respecto al conjunto original se presenta en las métricas de precisión, ratio de verdaderos positivos y puntaje de confianza f-1.





Cuantitativamente, la media ponderada de estas métricas se ve de la siguiente manera.

	Original	Modificación
Precisión	0.7887	0.8410
Ratio de verdaderos positivos	0.7754	0.8335
Puntaje de confianza f-1	0.7668	0.8336

Conclusiones

Como se puede apreciar en la sección de resultados, la modificación propuesta en que enriqueció el conjunto de datos con imágenes generadas artificialmente obtuvo resultados ligeramente mejores, lo que puede inferir que este tipo de balanceo puede traer beneficios a otros conjuntos de datos con este problema.

Cabe mencionar que el entrenamiento de la FS-GAN es tardado y computacionalmente costoso. Se utilizó un servidor dedicado en el que tomó alrededor de tres horas el entrenamiento del generador para cada subconjunto de imágenes, dicho proceso se repitió 12 veces.

Para completamente erradicar el desbalanceo de clases sería necesario que todas las subclases de ambos tipos benigno y maligno contaran con la misma cantidad de imágenes, sin embargo, en el caso del conjunto de datos BreakHis esto es poco recomendable ya que algunas de estas subclases contienen alrededor de 800 muestras frente a otras que tienen menos de 200, por lo que complementar las subclases con menos muestras para que alcancen a aquellas con más podría disminuir la calidad del conjunto de datos al componerse mayoritariamente de muestras sintéticas.

Referencias

- [1] R. R. Barbieri *et al.*, “Reimagining leprosy elimination with AI analysis of a combination of skin lesion images with demographic and clinical data,” *The Lancet Regional Health - Americas*, vol. 9, no. 100192, May 2022, doi: <https://doi.org/10.1016/j.lana.2022.100192>.
- [2] K. Theofilatos *et al.*, “Predicting protein complexes from weighted protein–protein interaction graphs with a novel unsupervised methodology: Evolutionary enhanced Markov clustering,” *Artificial Intelligence in Medicine*, vol. 63, no. 3, pp. 181–189, Mar. 2015, doi: <https://doi.org/10.1016/j.artmed.2014.12.012>.
- [3] V. Kaul, S. Enslin, and S. A. Gross, “History of artificial intelligence in medicine,” *Gastrointestinal Endoscopy*, vol. 92, no. 4, pp. 807–812, Oct. 2020, doi: <https://doi.org/10.1016/j.gie.2020.06.040>.
- [4] B. G. Silverman, N. Hanrahan, G. Bharathy, K. Gordon, and D. Johnson, “A systems approach to healthcare: Agent-based modeling, community mental health, and population

well-being,” *Artificial Intelligence in Medicine*, vol. 63, no. 2, pp. 61–71, Feb. 2015, doi: <https://doi.org/10.1016/j.artmed.2014.08.006>.

[5] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “A Dataset for Breast Cancer Histopathological Image Classification,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016, doi: <https://doi.org/10.1109/tbme.2015.2496264>.

[6] Fatin Izzati Ramli, Praveen Thokala, T. Tong, and K. Unger-Saldaña, “Understanding breast cancer patient pathways and their impact on survival in Mexico,” *Journal of Cancer Policy*, vol. 40, no. 100482, 2024, doi: <https://doi.org/10.1016/j.jcpo.2024.100482>.

[7] M. A. Hambali, T. O. Oladele, and K. S. Adewole, “Microarray cancer feature selection: Review, challenges and research directions,” *International Journal of Cognitive Computing in Engineering*, vol. 1, pp. 78–97, Jun. 2020, doi: <https://doi.org/10.1016/j.ijcce.2020.11.001>.

[8] A. A. Khan, O. Chaudhari, and R. Chandra, “A Review of Ensemble Learning and Data Augmentation Models for Class Imbalanced problems: Combination, Implementation and Evaluation,” *Expert Systems with Applications*, vol. 244, p. 122778, Dec. 2023, doi: <https://doi.org/10.1016/j.eswa.2023.122778>.

[9] E. Robb, W.-S. Chu, A. Kumar, and J.-B. Huang, “Few-shot adaptation of generative adversarial networks,” *ArXiv*, vol. abs/2010.11943, 2020, Available: <https://api.semanticscholar.org/CorpusID:225039998>

[10] I. Goodfellow *et al.*, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: <https://doi.org/10.1145/3422622>.

[11] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” *ArXiv*, vol. abs/1905.11946, 2019, Available: <https://api.semanticscholar.org/CorpusID:167217261>