



## DETECCIÓN DE ANOMALÍAS RELACIONADAS CON EL CÁNCER DE MAMA UTILIZANDO MODELOS DE APRENDIZAJE PROFUNDO E IMÁGENES TÉRMICAS

Guillermo Pinto Ruiz

Ingeniería de Sistemas, Universidad Industrial de Santander, Colombia,  
[guillermo2210069@correo.uis.edu.co](mailto:guillermo2210069@correo.uis.edu.co)

Miguel Fernando Pimiento Escobar

Ingeniería de Sistemas, Universidad Industrial de Santander, Colombia,  
[miguel2210054@correo.uis.edu.co](mailto:miguel2210054@correo.uis.edu.co)

Cristian Diaz Franco

Ingeniería de Sistemas, Universidad Industrial de Santander, Colombia,  
[cristian2211251@correo.uis.edu.co](mailto:cristian2211251@correo.uis.edu.co)

Hoover Fabián Rueda Chacón, PhD.

Ingeniería de Sistemas, Universidad Industrial de Santander, Colombia,  
[hfarueda@uis.edu.co](mailto:hfarueda@uis.edu.co)

### Resumen

**Objetivo:** Desarrollar un enfoque de visión por computadora para la detección de anomalías relacionadas con el cáncer de mama usando transformadores de visión y redes neuronales convolucionales en imágenes térmicas. **Metodología:** Utilizamos el conjunto de datos *Database for Research Mastology with Infrared Image* para entrenar y evaluar modelos preentrenados como *Vision Transformers (ViT)*, *ResNet* y *Xception*. Implementamos validación cruzada y técnicas de atención visual, como *GradCAM*, para mejorar la explicabilidad. **Resultados:** Los modelos *ViT\_B\_16* y *ResNet152* lograron una exactitud del 95,30% y 95,19% respectivamente. La validación cruzada fortalece nuestros resultados en comparación con estudios previos que no aplicaron esta metodología. *GradCAM* reveló que las predicciones se basaron en regiones relevantes de las imágenes. **Conclusión:** Los resultados subrayan el potencial de los modelos de aprendizaje profundo y las termografías, abriendo oportunidades para aplicaciones clínicas en la detección temprana del cáncer de mama. El código está disponible en <https://github.com/guillepinto/Fine-Tuned-Thermal-Breast-Cancer-Diagnosis>

**Palabras clave:** Cáncer de mama, Redes neuronales convolucionales, Transformador de visión, Termografía mamaria, Transferencia de aprendizaje.

## 1. Introducción

El cáncer de mama (CM) es el cáncer más comúnmente diagnosticado en mujeres a nivel mundial, representando más del 23% de los casos registrados (Giaquinto et al., 2022). En 2022, más de 666,000 mujeres murieron por CM, con una incidencia creciente (Grigorean & Cristian, 2022) especialmente en economías emergentes, donde la mortalidad supera a la de países desarrollados (Bray et al., 2024). En Colombia, la tasa de mortalidad estandarizada por edad (ASR) por cada 100,000 mujeres ha aumentado de 8.4 en 2008 a 9.6 en 2020 (IARC<sup>1</sup>). El diagnóstico temprano es crucial para la supervivencia: la tasa a cinco años es del 99% en casos localizados<sup>2</sup>, pero cae al 30% cuando la enfermedad es distante (Giaquinto et al., 2022). Aunque puede detectarse en autoexámenes, a menudo la masa es demasiado pequeña para percibirse al tacto. Por ello, se recomiendan estudios médicos, aunque estos enfrentan limitaciones como el acceso complicado o el alto costo (Bray et al., 2024).

Las técnicas de imagenología como la mamografía, el ultrasonido y la resonancia magnética son actualmente las principales

herramientas para la detección temprana de CM, aunque presentan limitaciones importantes. La mamografía expone a radiación, tiene un costo elevado y baja especificidad en mujeres con tejido mamario denso (Giaquinto et al., 2022). El ultrasonido ofrece imágenes menos detalladas, no siempre cubre toda la mama y depende de la experiencia del operador, mientras que la resonancia magnética, aunque efectiva, está limitada en pacientes con implantes metálicos y puede provocar claustrofobia debido a la duración y la necesidad de inmovilidad prolongada del procedimiento (Din et al., 2022). A pesar de estas modalidades, la biopsia sigue siendo el único método definitivo para confirmar CM (Figueiredo et al., 2019). Sin embargo, el análisis de imágenes histopatológicas es difícil, requiere tiempo y está sujeto a la experiencia del especialista, lo que puede derivar en diagnósticos inexactos (Zhang et al., 2019).

Las imágenes térmicas (termografías) han emergido como una técnica prometedora por ser no invasiva, sin contacto, libre de radiación y económica, lo que la hace viable para la detección temprana del CM. Esta técnica mide los patrones de calor en la piel, que pueden indicar tumores malignos debido al aumento de la perfusión sanguínea y el metabolismo en las células cancerosas (Ring & Ammer, 2012). A diferencia de otras modalidades, la termografía es eficaz en

---

<sup>1</sup> <https://gco.iarc.fr/en>

<sup>2</sup> Los casos localizados se refieren al cáncer confinado a la mama, mientras que los casos distantes indican que el cáncer se ha propagado a otras partes del cuerpo, como órganos o huesos (metástasis).

tejido denso o en etapas iniciales de la enfermedad (McCormack & dos Santos Silva, 2006). A pesar de sus ventajas, las termografías no pueden detectar CM (Din et al., 2022) y la interpretación sigue siendo un reto, ya que el análisis manual puede ser propenso a errores (Figueiredo et al., 2019). En este sentido, métodos automáticos como las redes neuronales convolucionales (CNNs), han mostrado resultados prometedores en la mejora de la detección de anomalías en imágenes médicas. Sin embargo, su uso en termografías para el diagnóstico del CM aún se encuentra en sus primeras etapas, principalmente debido al costo computacional (Zuluaga-Gomez et al., 2021).

Nuestro estudio se enfoca en afrontar estas limitaciones aplicando modelos preentrenados de aprendizaje profundo, proponiendo los ViT, que han demostrado eficacia en la clasificación de imágenes en diversos dominios<sup>3</sup>. Aprovechamos transferencia de aprendizaje, reutilizando el conocimiento adquirido por modelos entrenados en grandes conjuntos de datos y adaptándolo a nuevos datos más específicos, como las termografías, con el objetivo de obtener resultados competitivos frente a métodos convencionales.

---

<sup>3</sup> Se hace referencia a dominios como la visión por computadora en aplicaciones generales (p. ej., clasificación de objetos).

## 2. Metodología

### 2.1 Adquisición de imágenes

En este trabajo se utilizó el conjunto de datos DMR-IR<sup>4</sup> (Silva et al., 2014), que forma parte de la *Database for Mastology Research*, una plataforma en línea que almacena imágenes mastológicas para la detección temprana de CM. Esta plataforma incluye termografías y mamografías, obtenidas por el Departamento de Ciencia de la Computación de la Universidad Federal Fluminense (IC/UFF). El conjunto de datos utilizado contiene termografías de 56 pacientes, de los cuales 19 son sanos y 37 presentan una mama maligna.

Las imágenes fueron capturadas con una cámara FLIR SC620, que registra temperaturas entre -50 °C y 400 °C, con un margen de error de 2 °C. El protocolo de adquisición basado en (Ng, 2009), asegura la calidad de las termografías mediante varias directrices: los pacientes deben evitar café, alcohol, comidas grandes y fumar antes de la prueba. La cámara debe operar 15 minutos antes a 100 mK y 30 °C, con una resolución mínima de 120<sup>2</sup> puntos térmicos. Además, la sala debe mantenerse entre 18-25 °C, con una humedad del 40-75%, piso alfombrado y evitar fuentes de calor. Los diagnósticos fueron confirmados previamente por expertos mediante mamografías y biopsias.

---

<sup>4</sup> <https://visual.ic.uff.br/dmi/>

## 2.2 Preprocesamiento y aumento de datos

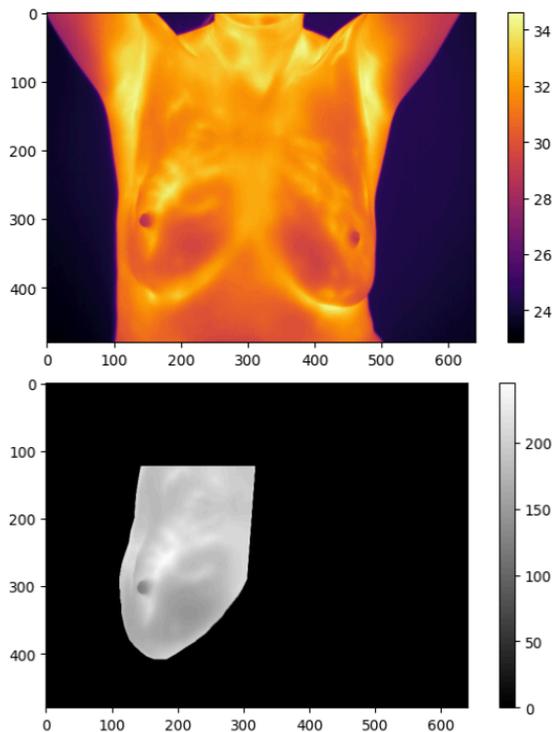
Se realiza el cargue de una matriz de temperatura y las imágenes segmentadas correspondientes a la zona de interés (p. ej., Figura 1) con Python, para un total de 1522 imágenes de los 56 pacientes. Se aplicó segmentación binaria sobre la imagen para obtener una máscara. Posteriormente, se multiplicaron las matrices de temperatura por estas máscaras para obtener la imagen segmentada con valores de temperatura. Se siguió la metodología propuesta en (Bhowmik et al., 2018), incluyendo redimensionamiento, recorte y normalización respecto a la temperatura máxima de todas las matrices, asegurando la coherencia de las termografías en términos de intensidad. Como resultado tenemos una imagen térmica de 224 x 224 reduciendo el coste computacional en un 75% y manteniendo suficiente detalle. Para el aumento de datos se aplicaron cuatro tipos de generación, cada una con una probabilidad del 50% de realizarse: (i) volteo horizontal, (ii) rotación entre 0 y 45 grados, (iii) desenfoque gaussiano con un kernel de tamaño cinco (5) y (iv) transformación afín aleatoria (hasta con el 5% del tamaño de la imagen).

## 2.3 Modelos de vanguardia

Utilizamos arquitecturas de aprendizaje profundo que han demostrado ser clave en el campo de la visión por computadora, destacando su rendimiento en el conjunto de datos *ImageNet* (Deng et al., 2009), una

referencia esencial para avances en inteligencia artificial.

**Figura 1.** *Matriz de temperatura e imagen de la zona de interés de paciente enfermo*



Fuente: Base de datos DMR-IR

Los modelos implementados en este estudio provienen de *torchvision*<sup>5</sup> y *timm*<sup>6</sup>, con pesos preentrenados en diferentes versiones de *ImageNet*. *ResNet*, propuesto por (He et al., 2015), se distingue por sus conexiones residuales, que permiten entrenar redes profundas al evitar la degradación del gradiente. Usamos su versión de 152 capas, *ResNet152*. *Xception*, desarrollada por (Chollet, 2016), mejora la eficiencia mediante

<sup>5</sup> <https://pytorch.org/vision/stable/models.html#classification>

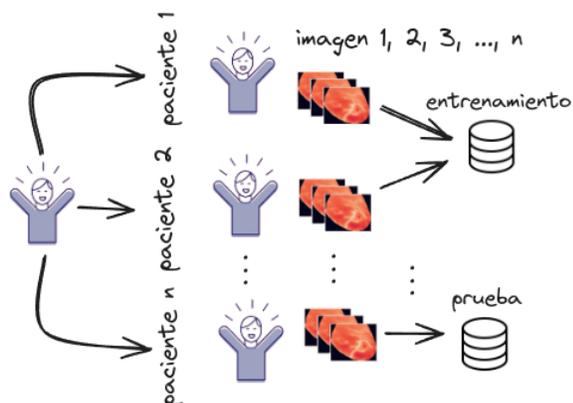
<sup>6</sup> <https://pprp.github.io/timm/models/xception/>

convoluciones separables en profundidad, lo que la hace menos costosa computacionalmente. Por último, ViT (Dosovitskiy et al., 2020) emplea autoatención para procesar imágenes como secuencias, similar al procesamiento de texto. Usamos su versión base, con 16x16 píxeles por cada parche de imagen, ViT\_B\_16.

## 2.4 División del conjunto de datos

La partición del conjunto de datos se realizó a nivel de pacientes, evitando el enfoque tradicional basado en imágenes. Esto previene que imágenes del mismo paciente se encuentren en ambos conjuntos, lo cual podría sesgar los resultados. Se asignaron 48 pacientes al conjunto de entrenamiento y 8 pacientes al conjunto de prueba (ver Figura 2). Además, para las etapas de entrenamiento y prueba, se implementó una división en 7  *folds* , cada uno compuesto por un 15% (8) de los pacientes, permitiendo una validación cruzada consistente.

**Figura 2.** Ejemplo de división de los datos



Fuente: Elaborada por autores

## 2.5 Optimización de hiperparámetros

Se ajustaron hiperparámetros de los modelos *Xception*, *ResNet* y *ViT* preentrenados mediante optimización bayesiana, utilizando la plataforma *W&B*<sup>7</sup>. La optimización bayesiana construye un modelo estadístico para predecir el rendimiento en diferentes configuraciones de hiperparámetros. Este modelo probabilístico genera una distribución que estima el error esperado para cada configuración y se actualiza en cada evaluación. En cada iteración, se sugiere una nueva combinación de hiperparámetros con el objetivo de maximizar (o minimizar) métricas, en nuestro caso, el *recall*. Los hiperparámetros ajustados incluyeron el tamaño de lote (*batch\_size*), tasa de aprendizaje (*learning\_rate*), optimizador (*optimizer*) y grado de ajuste fino (*fine\_tuning*).

La transferencia de aprendizaje consistió en congelar o descongelar capas preentrenadas y ajustar el modelo en un proceso conocido como ajuste fino. En nuestros experimentos, probamos dos enfoques: (1) descongelar solo el clasificador y la primera capa convolucional, mientras las capas intermedias permanecen congeladas; y (2) descongelar todas las capas para ajustar el modelo completamente. Debido al alto costo computacional, se seleccionó aleatoriamente uno de los siete  *folds*  en cada iteración,

<sup>7</sup> <https://wandb.ai/>

evitando realizar una validación cruzada completa, la cual toma aproximadamente cuatro horas por iteración para ViT.

## 2.6 Validación cruzada

Utilizamos los mejores modelos obtenidos en el proceso de optimización de hiperparámetros y realizamos una validación cruzada para evaluar su desempeño, con la división introducida en la subsección 2.4. En cada iteración, se utilizan 6 *folds* para entrenar el modelo, y el *fold* restante se reserva para probarlo. Este proceso se repite hasta que cada uno de los *folds* haya sido utilizado como conjunto de prueba. Al finalizar las iteraciones, se calcularon la media y desviación estándar de las métricas (ver Figura 3): exactitud (*accuracy*), valor-F1 (*F1-score*), precisión (*precision*) y sensibilidad (*recall*), que ofrecen una visión detallada del rendimiento de los modelos.

## 2.7 Explicabilidad

La explicabilidad en modelos de aprendizaje profundo es crucial en medicina. Saber qué regiones de una imagen son relevantes para las predicciones refuerza la confianza de los médicos en los sistemas automatizados (Hassija et al., 2024). Por ello, implementamos *Gradient-weighted Class Activation Mapping (GradCAM)* (Selvaraju et al., 2016), un método que genera mapas de activación ponderados por gradientes, destacando las áreas de la imagen que

influyen en las decisiones del modelo. Para nuestra investigación, usamos la implementación disponible en la librería *pytorch-grad-cam*<sup>8</sup>. Cargamos los pesos entrenados de cada modelo y seleccionamos las capas previas a la clasificación. Posteriormente, generamos los mapas de activación que resaltan las regiones más relevantes para la predicción.

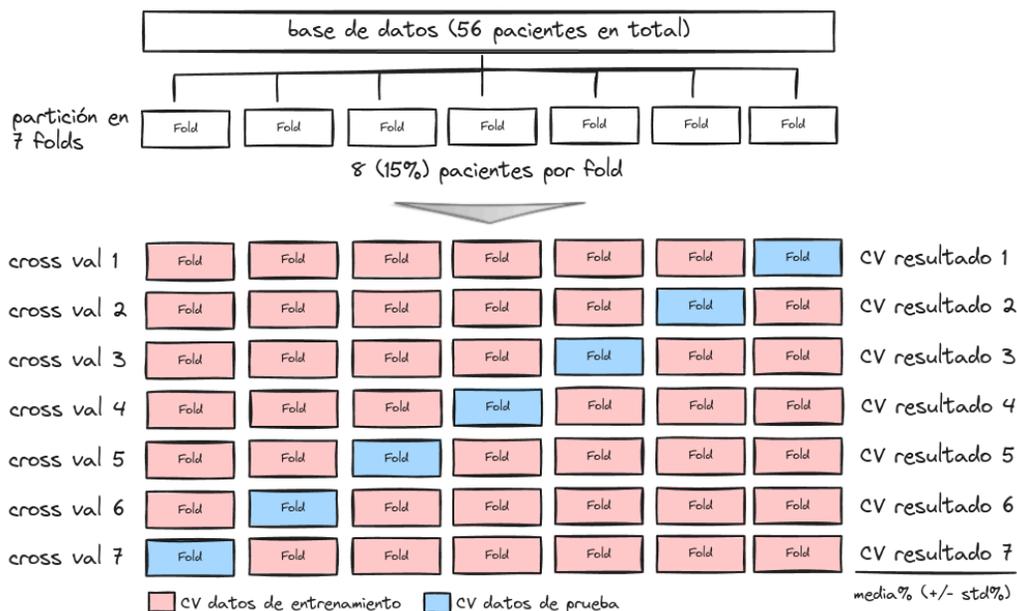
## 3. Resultados

Los modelos *ResNet152* y *ViT\_B\_16* arrojaron los mejores resultados para la clasificación del CM. El modelo *ViT\_B\_16* obtuvo una exactitud de 95.30% y una sensibilidad de 91.10%. Por su parte, *ResNet152* alcanzó una exactitud de 95.19% y una sensibilidad de 91.13%. Los resultados completos para cada modelo se muestran en la Tabla 1. Los resultados al aplicar *GradCAM* para evaluar la explicabilidad mostraron que los modelos con buenos resultados, como *ViT\_B\_16* y *ResNet152*, activaron correctamente las áreas afectadas por la enfermedad, confirmando que las predicciones se basaban en regiones relevantes de las imágenes. En contraste, modelos seleccionados con bajo rendimiento, presentaron activaciones dispersas en áreas no relacionadas con la enfermedad. Estas diferencias se ilustran en la Figura 4.

---

<sup>8</sup> <https://jacobgil.github.io/pytorch-gradcam-book/>

**Figura 3. Validación cruzada por grupos**



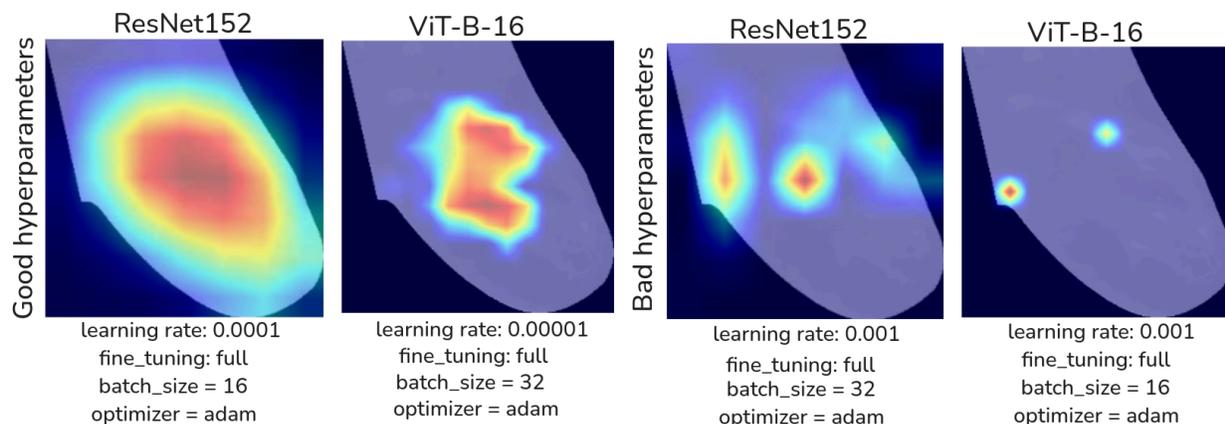
Fuente: Elaborada por autores

**Tabla 1. Resultados para cada modelo**

Modelo	Accuracy	F1-score	Precision	Recall
ViT-B-16 (Full)	<b>0.9530</b> ( $\pm$ 0.0381)	<b>0.9483</b> ( $\pm$ 0.0453)	0.9957 ( $\pm$ 0.0075)	0.9110 ( $\pm$ 0.0788)
ResNet-152 (Full + Adam)	0.9519 ( $\pm$ 0.0414)	0.9433 ( $\pm$ 0.0560)	0.9953 ( $\pm$ 0.0055)	0.9113 ( $\pm$ 0.0943)
ResNet-152 (Full + SGD)	0.9231 ( $\pm$ 0.0525)	0.9093 ( $\pm$ 0.0722)	<b>1.0000</b> ( $\pm$ 0.0000)	0.8439 ( $\pm$ 0.1165)
ViT-B-16 (Classifier+conv1)	0.9087 ( $\pm$ 0.0568)	0.9167 ( $\pm$ 0.0523)	0.9063 ( $\pm$ 0.0969)	<b>0.9390</b> ( $\pm$ 0.0591)
Xception (Classifier+conv1)	0.8623 (+/- 0.1212)	0.8340 (+/- 0.1418)	0.8984 (+/- 0.1516)	0.7883 (+/- 0.1531)

Fuente: Elaborada por autores

**Figura 4. Comparación activaciones GradCAM**



Fuente: Elaborada por autores

#### 4. Discusión

Nuestros hallazgos destacan el rendimiento de los modelos *ViT\_B\_16* y *ResNet152* en la clasificación de termografías de CM, con exactitudes superiores al 95%. Comparado con estudios previos, como el de (Chatterjee et al., 2022), que reportó una exactitud del 100% utilizando *VGG16* y el algoritmo *Dragonfly* para selección de características, u trabajos como el de (Ekici & Jawzal, 2020), alcanzando una exactitud del 98.95% mediante optimización bayesiana con CNNs, nuestro enfoque destaca por su capacidad de generalización, gracias a la validación cruzada para evitar el sobreajuste y el uso de modelos preentrenados. Además, la aplicación de técnicas avanzadas como *GradCAM* mejoran la interpretabilidad de las predicciones, un aspecto que no fue abordado en la mayoría de estos trabajos.

Estudios como el de (Farooq & Corcoran, 2020) utilizaron el modelo *Inception-V3* con preprocesamiento especializado para la normalización y mejora de contraste, obteniendo una exactitud del 80%, lo que refleja la variabilidad en el rendimiento según el preprocesamiento aplicado. De manera similar, modelos más complejos como *InceptionResNetV2* o *SeResNet18* solo lograron un 90% y 65% de exactitud, respectivamente. (Zuluaga-Gomez et al., 2021). Esto destaca la robustez de nuestro enfoque de transferencia de aprendizaje con

*ViTs*, siendo la primera vez que se prueba en este conjunto de datos.

#### 5. Conclusiones

Esta investigación destaca la importancia del preprocesamiento de las imágenes y la selección adecuada de hiperparámetros en la optimización de modelos de aprendizaje profundo para detectar anomalías en termografías. Aunque las termografías por sí solas no pueden detectar CM, combinadas con modelos de aprendizaje profundo ofrecen una herramienta valiosa como apoyo a los médicos para la detección temprana. La investigación presenta algunas limitaciones, como el uso exclusivo de una sola base de datos (DMR-IR), lo que restringe la capacidad de generalización de los modelos. Resaltando la necesidad urgente de invertir en la creación y disponibilidad de nuevas bases de datos públicas que puedan ampliar el alcance de los modelos de aprendizaje profundo en este campo. Para trabajos futuros, se sugiere explorar enfoques multimodales. La integración de termografías con mamografías disponibles en el mismo conjunto de datos podría resultar en un sistema más robusto en la detección del CM.

#### Agradecimientos

Agradecemos al Semillero de Investigación Hands-On Computer Vision por su valiosa dirección y apoyo durante el desarrollo de este proyecto.

## Referencias

- Bhowmik, M. K., Gogoi, U. R., Majumdar, G., Bhattacharjee, D., Datta, D., Ghosh, A. K., Bhowmik, M. K., Gogoi, U. R., Majumdar, G., Bhattacharjee, D., Datta, D. & Ghosh, A. K. (2018). Designing of ground-truth-annotated DBT-TU-JU breast thermogram database toward early abnormality prediction. *IEEE Journal of Biomedical and Health Informatics*, 22(4), 1238–1249.
- Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I. & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3), 229–263.
- Chatterjee, S., Biswas, S., Majee, A., Sen, S., Oliva, D. & Sarkar, R. (2022). Breast cancer detection from thermal images using a Grunwald-Letnikov-aided Dragonfly algorithm-based deep feature selection method. *Computers in Biology and Medicine*, 141(105027), 105027.
- Chollet, F. (2016). Xception: Deep learning with depthwise separable convolutions. In *arXiv [cs.CV]*. arXiv. <http://arxiv.org/abs/1610.02357>
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.
- Din, N. M. U., Dar, R. A., Rasool, M. & Assad, A. (2022). Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. *Computers in Biology and Medicine*, 149, 106073.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. & Hounsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. In *arXiv [cs.CV]*. arXiv. <http://arxiv.org/abs/2010.11929>
- Ekici, S. & Jawzal, H. (2020). Breast cancer diagnosis using thermography and convolutional neural networks. *Medical Hypotheses*, 137(109542), 109542.
- Farooq, M. A. & Corcoran, P. (2020). Infrared imaging for human thermography and breast tumor classification using thermal images. *2020 31st Irish Signals and Systems Conference (ISSC)*, 1–6.
- Figueiredo, A. A. A., do Nascimento, J. G., Malheiros, F. C., da Silva Ignacio, L. H., Fernandes, H. C. & Guimaraes, G. (2019). Breast tumor localization using skin surface temperatures from a 2D anatomic model without knowledge of the thermophysical properties. *Computer Methods and Programs in Biomedicine*, 172, 65–77.
- Giaquinto, A. N., Sung, H., Miller, K. D., Kramer, J. L., Newman, L. A., Minihan, A., Jemal, A. & Siegel, R. L. (2022).

- Breast Cancer Statistics, 2022. *CA: A Cancer Journal for Clinicians*, 72(6), 524–541.
- Grigorean, V. T. & Cristian, D. A. (2022). Cancer-Yesterday, Today, Tomorrow. *Medicina*, 59(1). <https://doi.org/10.3390/medicina59010098>
- Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M. & Hussain, A. (2024). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation*, 16(1), 45–74.
- He, K., Zhang, X., Ren, S. & Sun, J. (2015). Deep residual learning for image recognition. In *arXiv [cs.CV]*. arXiv. <http://arxiv.org/abs/1512.03385>
- McCormack, V. A. & dos Santos Silva, I. (2006). Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 15(6), 1159–1169.
- Ng, E. Y.-K. (2009). A review of thermography as promising non-invasive detection modality for breast tumor. *International Journal of Thermal Sciences*, 48(5), 849–859.
- Ring, E. F. J. & Ammer, K. (2012). Infrared thermal imaging in medicine. *Physiological Measurement*, 33(3), R33–R46.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. & Batra, D. (2016). Grad-CAM: Visual explanations from deep networks via Gradient-based localization. In *arXiv [cs.CV]*. arXiv. <https://doi.org/10.1007/s11263-019-01228-7>
- Silva, L. F., Saade, D. C. M., Sequeiros, G. O., Silva, A. C., Paiva, A. C., Bravo, R. S. & Conci, A. (2014). A new database for breast research with infrared image. *Journal of Medical Imaging and Health Informatics*, 4(1), 92–100.
- Zhang, X., Zhang, Y., Qian, B., Liu, X., Li, X., Wang, X., Yin, C., Lv, X., Song, L. & Wang, L. (2019). Classifying Breast Cancer Histopathological Images Using a Robust Artificial Neural Network Architecture. *Bioinformatics and Biomedical Engineering*, 204–215.
- Zuluaga-Gomez, J., Al Masry, Z., Benagoune, K., Meraghni, S. & Zerhouni, N. (2021). A CNN-based methodology for breast cancer diagnosis using thermal images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 9(2), 131–145.