

**ESTRATEGIA DE APRENDIZAJE NO SUPERVISADO PARA EL ANÁLISIS
DE IMÁGENES DE HISTOPATOLOGÍA EN CÁNCER DE MAMA**

MARÍA FERNANDA NAVAS BURGOS



**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA**

2022

**ESTRATEGIA DE APRENDIZAJE NO SUPERVISADO PARA EL ANÁLISIS
DE IMÁGENES DE HISTOPATOLOGÍA EN CÁNCER DE MAMA**

MARÍA FERNANDA NAVAS BURGOS

**Una tesis presentada en cumplimiento de los requisitos para el grado de:
Ingeniera de Sistemas e Informática**

Director:

David Romo Bucheli, Ph.D

Codirector:

Fabio Martínez Carrillo, Ph.D



**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA**

2022

AGRADECIMIENTOS

El autor expresa su agradecimiento:

Agradezco profundamente a mi familia que han sido un elemento importante en mi vida y sin ellos no habría sido posible la culminación de mis estudios profesionales, a mis padres Ludwig y Esperanza por darme su amor y apoyo incondicional, sin importar las circunstancias han estado para mi en todo momento y todo esto no hubiera sido posible sin ellos. A mi hermano William por ser una fuente de inspiración no solo en mi carrera universitaria si no en mi vida en general. Siempre estaran presentes en mi corazón, los amo mucho.

A mis directores de proyecto David Romo y Fabio Martínez por su apoyo y paciencia además de guiarme e impulsarme durante todo este trayecto, gracias por ser modelos a seguir no solo en el ambito de la ciencia si no también en el personal.

A mis compañeros del grupo de investigación Bivl2ab por sus consejos y sugerencias para culminar este proyecto

Por último quiero agradecer a mis amigos Juan David, Juan Pablo, Santiago, Manuel, Sebastian, Yara, Mariana y Jorge los cuales han sido personas excepcionales que me han acompañado en momentos de felicidad y tristeza, gracias a ellos soy la persona que soy. Estoy en eterno agradecimiento con la vida por haber puesto en mi camino a tan increíbles seres.

Índice general

	Pág
INTRODUCCIÓN	11
1. MARCO TEÓRICO Y TRABAJOS PREVIOS	14
1.1. Graduación Histológica en Cáncer de Mama	15
1.2. Análisis Automático en Imágenes de Cáncer de Mama	18
1.3. Aprendizaje no Supervisado	21
1.3.1. Autoencoders	23
1.4. Aprendizaje multi-instancia (multiple-instance learning (MIL))	25
2. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA	27
3. OBJETIVOS	28
4. REPRESENTACIÓN LATENTE DE IMÁGENES HISTOLÓGICAS DE CANCER DE MAMA VÍA APRENDIZAJE NO SUPERVISADO	29
4.1. Autoencoder para la representación de imágenes histológicas	29
4.1.1. Encoder	29
4.1.2. Decoder	29
4.2. Conjunto de datos MITOS ATYPIA 2014	31
4.3. Entrenamiento del autoencoder y diseño experimental	32
4.4. Métricas de evaluación para la reconstrucción en el autoencoder	33
4.5. Resultados de reconstrucción del autoencoder	34

5. APRENDIZAJE MULTI-INSTANCIA USANDO REPRESENTACIÓN LATENTE PARA EL PLEOMORFISMO NUCLEAR EN CÁNCER DE MAMA 38

5.1. Arquitectura de red neuronal para el aprendizaje Multi-instancia 38

5.2. Conjunto de datos: Embebidos - MITOS ATYPIA 39

5.3. Entrenamiento del modelo multi-instancia y diseño experimental 40

5.4. Métricas de evaluación para la clasificación multi-instancia 40

5.5. Evaluación y resultados 42

6. CONCLUSIONES Y PERSPECTIVAS 45

BIBLIOGRAFÍA 47

Índice de figuras

	Pág
Figura 1. Grado histológico del cáncer de mama según la evaluación del sistema de clasificación de Nottingham donde se observan cambios notables en la formación tubular Rakha et al (2010), Recuperado de ¹	17
Figura 2. Diferentes grados de formación tubular Moncayo et al, Recuperado de ²	17
Figura 3. (A) Campos histológicos que contienen células mitóticas. (B) Datos brutos de dos muestras diferentes que no contienen células mitóticas. Sommer et al (2012), Recuperado de ³	17
Figura 4. El tejido del cáncer de mama muestra una formación de túbulos altos (izquierda) y bajos (derecha). El lumen está delimitado por líneas azules. Los túbulos están delimitados por líneas naranjas. Romo et al (2016), Recuperado de ⁴	18
Figura 5. Estructura de Autoencoder	23
Figura 6. Principio del aprendizaje multi*instancia	25
Figura 7. Arquitectura de Autoencoder	30
Figura 8. Ejemplo de campos histológicos a 20X con grado pleomórfico asignado por un experto	31

¹ Emad A Rakha y col. “Breast cancer prognostic classification in the molecular era: the role of histological grade”. En: *Breast Cancer Research* 12.4 (2010), págs. 1-12.

² Ricardo Alexander Moncayo Martinez y col. “Quantitative estimation of nuclear pleomorphism in breast cancer histological images”. Tesis doct. Universidad Nacional de Colombia-Sede Bogotá.

³ Christoph Sommer y col. “Learning-based mitotic cell detection in histopathological images”. En: ene. de 2012, págs. 2306-2309.

⁴ David Romo-Bucheli y col. “Automated tubule nuclei quantification and correlation with oncotype DX risk categories in ER+ breast cancer whole slide images”. En: *Scientific reports* 6.1 (2016), págs. 1-9.

Figura 9. Graficas MSE	35
Figura 10.Graficas PSNR	35
Figura 11.Graficas UMAP de la representación latente	36
Figura 12.Gráficas Correlación para todos los tamaños de parches	37
Figura 13.Enfoque profundo de MIL	38
Figura 14.Estrategia MIL basada en embebidos para la clasificación de grados de pleo- morfismo	38
Figura 15.Matriz de embebidos correspondientes a un campo visual	39
Figura 16.Curva ROC, curva AUC y matrices de confusión para la clasificación de pleo- morfismo nuclear	44

Índice de cuadros

	Pág
Tabla 1. Tabla de resultados de las métricas MSE y MAE de la reconstrucción de imágenes	34
Tabla 2. Tabla de resultados de diferentes métodos en categorización de pleomorfismo nuclear	43

RESUMEN

TÍTULO: ESTRATEGIA DE APRENDIZAJE NO SUPERVISADO PARA EL ANÁLISIS DE IMÁGENES DE HISTOPATOLOGÍA EN CÁNCER DE MAMA *

AUTOR: MARÍA FERNANDA NAVAS BURGOS **

PALABRAS CLAVE: Aprendizaje no supervisado, histopatología, cáncer de mama

DESCRIPCIÓN: El cáncer de mama es el cáncer invasivo más común que afecta a las mujeres alrededor del mundo, con una probabilidad de letalidad del 2,6% . En Colombia, se estima que aproximadamente 13,380 mujeres padecieron este tipo de cáncer en el 2018, según datos del Observatorio Global de Cáncer. La graduación del cáncer de mama corresponde a la inspección visual de imágenes histológicas realizada por patólogos expertos. Este análisis permite estimar la agresividad del cáncer y también su recurrencia. Particularmente, tres características histológicas son evaluadas: la formación de túbulos, el pleomorfismo nuclear y la actividad mitótica. Debido a que existen diferencias en la selección de regiones de interés diagnóstico, así como diferencias en la interpretación de los protocolos, es frecuente que exista una variabilidad inter-observador significativa para el grado histológico. Diversas técnicas automáticas de aprendizaje supervisado basadas en ingeniería de características, y también en esquemas de aprendizaje profundo, han sido propuestas en el análisis automático de imágenes histológico de cáncer de mama. Estas técnicas han reportado resultados exitosos en conjuntos de imágenes limitados. En Este trabajo se exploró una estrategia de aprendizaje profundo no supervisado para la identificación de características visuales relevantes para el diagnóstico clínico de cáncer de mama. Este enfoque se evaluó haciendo uso del PSNR para medir la reconstrucción de las imágenes en donde se obtuvo un 88.24412dB en esta medida. En la tarea de clasificación se utilizaron imágenes provenientes de 11 pacientes donde 2 fueron para evaluación, los resultados conseguidos fueron 0.697 en el área bajo la curva ROC AUC.

* Trabajo de investigación

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: David Romo Bucheli, Ph.D Codirector: Fabio Martínez, Ph.D

ABSTRACT

TITLE: UNSUPERVISED LEARNING STRATEGY FOR HISTOPATHOLOGY IMAGE ANALYSIS IN BREAST CANCER. *

AUTHOR: MARÍA FERNANDA NAVAS BURGOS **

KEYWORDS: unsupervised learning, histopathology, breast cancer

DESCRIPTION: Breast cancer is the most common invasive cancer affecting women around the world, with a probability of lethality of 2.6% . In Colombia, it is estimated that approximately 13,380 women suffered from this type of cancer in 2018, according to data from the Global Cancer Observatory. The grading of breast cancer corresponds to visual inspection of histological images by expert pathologists. This analysis makes it possible to estimate the aggressiveness of the cancer and also its recurrence. In particular, three histological features are evaluated: tubule formation, nuclear pleomorphism and nuclear activity, three histological features are evaluated: tubule formation, nuclear pleomorphism and mitotic activity. Since there are differences in the selection of regions of diagnostic interest, as well as differences in the interpretation of protocols, there is often significant inter-observer variability in histologic grade. Several automatic supervised learning techniques based on feature engineering, and also on deep learning schemes, have been proposed in the automatic analysis of breast cancer histological images. These techniques have reported successful results on limited image sets. In this work we explored an unsupervised deep learning strategy for the identification of visual features relevant to the clinical diagnosis of breast cancer. breast cancer clinical diagnosis. This approach was evaluated by making use of PSNR to measure image reconstruction where 88.24412 dB was obtained on this measure. In the classification task we used images from 11 patients where 2 were for evaluation, the results obtained were 0.697 in the area under the ROC curve AUC.

* Research work

** Faculty of Physical-Mechanical Engineering. School of Systems and Computer Engineering. Advisor: David Romo Bucheli

INTRODUCCIÓN

El cáncer de mama es la segunda causa de muerte entre las mujeres a nivel mundial, después del cáncer de pulmón. Generalmente es causada por el crecimiento de células anormales incontroladas que por lo general surgen de los conductos lácteos internos o del lóbulo ¹. La probabilidad de que una mujer muera de cáncer de mama es de aproximadamente 1 de cada 38². En Colombia, según las cifras entregadas el 2018 por la Agencia Internacional para la Investigación del Cáncer (IARC), se presentan 13,380 casos nuevos de cáncer de mama al año y 3,702 mujeres mueren por esta enfermedad. Según el Ministerio de Salud y Protección Social se considera que este tipo de cáncer es la primera causa de muerte por cáncer entre las mujeres colombianas³.

Las muestras histológicas permiten realizar análisis clínicos para determinar la agresividad y el riesgo de recurrencia del cáncer de mama. El resultado de ese análisis informa el diagnóstico y la prognosis de los pacientes ⁴. Estos análisis son realizados siguiendo las pautas de protocolos para la asignación del grado histológico del tumor. Usualmente se utiliza el sistema modificado de Bloom-Richardson, que consiste en la evaluación del pleomorfismo nuclear, la formación de túbulos y la actividad mitótica. No obstante, el grado histológico es considerado como subjetivo y estos parámetros se toman como insuficientes para estratificar a los pacientes que necesitan

¹ American Cancer Society. *Cancer de Seno*. <https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html>. [Online; accessed 15-August-2020]. 2018.

² World Health Organization. *Colombia Globocan*. <https://gco.iarc.fr/today/data/factsheets/populations/170-colombia-fact-sheets.pdf>. [Online; accessed 03-July-2020]. 2018.

³ Liga Colombiana Contra el Cancer. *Cancer de Seno*. <https://www.ligacancercolombia.org/cancer-de-seno/>. [Online; accessed 11-August-2020]. 2018.

⁴ Andrew H Beck y col. "Systematic analysis of breast cancer morphology uncovers stromal features associated with survival". En: *Science translational medicine* 3.108 (2011), 108ra113-108ra113.

tratamientos invasivos como quimioterapia ⁵. El acuerdo entre los patólogos sobre el grado de un caso de cáncer es moderado. La diferencia en la interpretación de los criterios de clasificación, su multiplicidad, y la subjetividad de las mediciones, da lugar a una reproducibilidad baja ⁶. Esta situación representa una motivación para desarrollar sistemas de soporte al diagnóstico. Por ejemplo, en un estudio realizado en el año 2005, se evaluó la graduación histopatológica y el índice de proliferación de 631 pacientes con cáncer de mama ingresadas en el Hospital St Luke. El grado patológico fue obtenido siguiendo los lineamientos del protocolo Bloom-Richardson modificado. Tras analizar las graduaciones de los componentes del grado de Bloom-Richardson, la concordancia fue mínima para el pleomorfismo nuclear ($\kappa = 0.37-0.50$), mejor para la tubularidad ($\kappa = 0.57-0.83$) e intermedia para el recuento mitótico ($\kappa = 0.45-0.64$)⁶.

En la literatura diferentes trabajos han explorado metodos automáticos supervisados para la caracterización de cáncer de mama, por ejemplo en ⁷ se utilizo *support vector machine (SVM)* junto con otras técnicas de aprendizaje máquina en donde se proponen redes neuronales convolucionales para predecir la malignidad. En ⁸ se hace uso del *multiple instance learning (MIL)* con *K nearest neighbors (KNN)*, para el diagnóstico asistido por computadora de pacientes con cáncer de mama, basado en el análisis de imágenes histopatológicas. Otros trabajos utilizan tecnicas de aprendizaje no supervisado para encontrar una representación o una caracteriza-

⁵ World Health Organization. *Colombia Globocan*. https://www.cancer.gov.co/files/libros/archivos/95685f345e64aa9f0fece8a589b5acc3_BOLETIN20HECHOS20Y20ACCIONES20MAMA.PDF. [Online; accessed 10-August-2020]. 2019.

⁶ John S Meyer y col. “Breast carcinoma malignancy grading by Bloom–Richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index”. En: *Modern pathology* 18.8 (2005), págs. 1067-1078.

⁷ Neslihan Bayramoglu, Juho Kannala y Janne Heikkilä. “Deep learning for magnification independent breast cancer histopathology image classification”. En: *2016 23rd International conference on pattern recognition (ICPR)*. IEEE. 2016, págs. 2440-2445.

⁸ PJ Sudharshan y col. “Multiple instance learning for histopathological breast cancer image classification”. En: *Expert Systems with Applications* 117 (2019), págs. 103-111.

ción del cáncer de mama, utilizando métodos como *Convolutional Neural Networks* y *Transfer learning* para recuperar información en diferentes escalas, incluyendo tanto el núcleo como la organización general de los tejidos⁹. Con los trabajos previamente nombrados podemos observar que existe una aplicación exitosa del aprendizaje profundo en los WSI (por sus siglas en inglés whole slide image) y existe un potencial de crear nuevas herramientas que superen los enfoques clínicos actuales en términos de precisión, reproducibilidad y objetividad; también proporciona nuevos conocimientos sobre diversas patologías ¹⁰¹¹.

El presente trabajo explora estrategias de aprendizaje profundo no supervisado que permita identificar características visuales relevantes para el diagnóstico clínico de cáncer de mama. Estas características se compararán con el pleomorfismo nuclear o la formación tubular. Esta tarea sera evaluada con un conjunto de datos de pacientes con cáncer de mama.

⁹ Teresa Araújo y col. “Classification of breast cancer histology images using convolutional neural networks”. En: *PloS one* 12.6 (2017), e0177544.

¹⁰ Eric Cosatto y col. “Grading nuclear pleomorphism on histological micrographs”. En: *2008 19th International Conference on Pattern Recognition*. IEEE. 2008, págs. 1-4.

¹¹ Mitko Veta y col. “Breast cancer histopathology image analysis: A review”. En: *IEEE Transactions on Biomedical Engineering* 61.5 (2014), págs. 1400-1411.

1. MARCO TEÓRICO Y TRABAJOS PREVIOS

El diagnóstico asistido por computadora (CAD) ha sido un campo de investigación importante durante las últimas décadas para el análisis de imágenes médicas. Los métodos CAD utilizan el aprendizaje máquina para analizar datos de pacientes con o sin imágenes, realizando una evaluación de la condición del paciente, que luego puede soportar la toma de decisiones clínicas. El reciente éxito de la tecnología de aprendizaje profundo en el reconocimiento de imágenes naturales ha estimulado la investigación y el desarrollo de nuevos métodos CAD orientados a tareas clínicas complejas como el análisis cuantitativo de imágenes de heterogeneidad tumoral, correlación de fenotipos de imágenes con procesos genéticos y biológicos subyacentes, diferenciación de subtipos de cáncer y estadificación del cáncer ¹².

Durante la última década, la patología se ha beneficiado del rápido progreso de la tecnología de digitalización de imágenes. La mejora de estas tecnologías ha llevado a la creación de escáneres que pueden producir imágenes PVC (Placas virtuales completas) que son exploradas por visores de imágenes de una manera comparable a la del microscopio convencional. Como alternativa a las láminas histológicas, las PVC son generalmente aceptadas en actividades como investigación, diagnóstico, y especialmente en educación, donde están disponibles para un gran número de estudiantes con todas las posibilidades de anotaciones sin el problema de variación entre secciones seriadas ¹³. Últimamente, se han desarrollado diversas herramientas computacionales, basadas en aprendizaje máquina, para asistir el análisis de imágenes histopatológicas. La incorporación de estas herramientas en el trabajo rutinario clínico es una tarea que aún está en progreso y que promete brindar a los patólogos un importante soporte en sus evaluaciones

¹² Heang-Ping Chan, Lubomir M Hadjiiski y Ravi K Samala. “Computer-aided diagnosis in the era of deep learning”. En: *Medical Physics* 47.5 (2020), e218-e227.

¹³ Chetan L Srinidhi, Ozan Ciga y Anne L Martel. “Deep neural network models for computational histopathology: A survey”. En: *arXiv preprint arXiv:1912.12378* (2019).

diagnósticas ¹⁴.

1.1. Graduación Histológica en Cáncer de Mama

El cáncer de mama abarca muchos tipos de cáncer localizados en las glándulas mamarias. Distintas áreas de la mama pueden resultar afectadas: los conductos, los lóbulos o el tejido intermedio. Según el origen de las células, los cánceres de mama se pueden dividir en dos amplias clasificaciones, carcinomas y sarcomas. Los carcinomas surgen del componente epitelial de la mama, que están conformados por las células que recubren los lóbulos y los conductos terminales responsables de producir leche. Los sarcomas son una forma mucho más rara de cáncer de mama (< 1% del cáncer de mama primario) que surge de los componentes estromales de la mama. Estos componentes incluyen miofibroblastos y células de los vasos sanguíneos. Dichos grupos no siempre son categorías suficientes ya que, en algunos casos, un solo tumor de mama puede ser una combinación de diferentes tipos de células ¹⁵.

Amplios espectros de patrones morfológico en carcinomas de mama han dado lugar a muchos intentos de crear clasificaciones significativas. Greenough introdujo la clasificación del carcinoma de mama en 1925, donde siete características histológicas para clasificar los carcinomas de mama en tres grados de malignidad. En 1957, Bloom y Richardson propusieron un sistema simplificado que utilizaba solo tres de las variables de Greenough ⁶: (i) La formación de túbulos (estructura glandular de la mama), (ii) el pleomorfismo nuclear (cuantificación de variabilidad en el tamaño y forma de los núcleos), y (iii) la actividad mitótica (relacionada con la tasa de división celular)

¹⁴ Shaimaa Al-Janabi, André Huisman y Paul J Van Diest. “Digital pathology: current status and future perspectives”. En: *Histopathology* 61.1 (2012), págs. 1-9.

¹⁵ Christopher W Elston y Ian O Ellis. “Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up”. En: *Histopathology* 19.5 (1991), págs. 403-410.

¹⁵¹⁶. A continuación se procederá a describir cada una de las características morfológicas.

Formación tubular: Según el UK National Breast Screening Programme se escanean todas las partes del tumor y cuando más del 75 % está compuesta de estructuras tubulares reconocibles se suma un punto. En los casos en que una cantidad moderada de formación de túbulos es vista, se otorgan 2 puntos. Cuando se observa menos del 10 % de formación de túbulos, las células crecen en láminas o cordones, la puntuación es de 3 puntos ¹⁷.

Pleomorfismo nuclear: En esta categoría se evalúan las variaciones de tamaño y forma de los núcleos. Cuando los núcleos son pequeños y se tiñen de color oscuro, con poco aumento de tamaño, contornos regulares y uniformidad de apariencia, una puntuación de 1 punto es apropiada. Se asignan dos puntos cuando las células parecen más grandes de lo normal, tienen núcleos vesiculares abiertos con nucléolos visibles y hay una variación moderada en tamaño y forma. Una marcada variación en tamaño y forma, particularmente cuando hay núcleos muy grandes y extraños, obtiene 3 puntos. Los núcleos de esta categoría tienen una estructura vesicular abierta con nucléolos grandes prominentes.

Actividad mitótica: Analiza el número relativo tanto de núcleos hipercromáticos como de figuras mitóticas. En la práctica, los núcleos hipercromáticos son difíciles de distinguir de aquellos que han sufrido degeneración picnótica. Por lo tanto, se ignoran los núcleos hipercromáticos y las figuras mitóticas se cuentan con un aumento fijo de 300 veces. Menos de 1 mitosis por campo se considera 1 punto, 1-2 por campo son 2 puntos y 3 o más por campo son 3 puntos. Para obtener la calificación general, se suman las puntuaciones de cada categoría, dando un total posible de 3-9 puntos.

¹⁶ Leslie W Dalton, David L Page y William D Dupont. "Histologic grading of breast carcinoma. A reproducibility study". En: *Cancer* 73.11 (1994), págs. 2765-2770.

¹⁷ Sunil R Lakhani. "The pathology of familial breast cancer: morphological aspects". En: *Breast Cancer Research* 1.1 (1999), págs. 1-5.

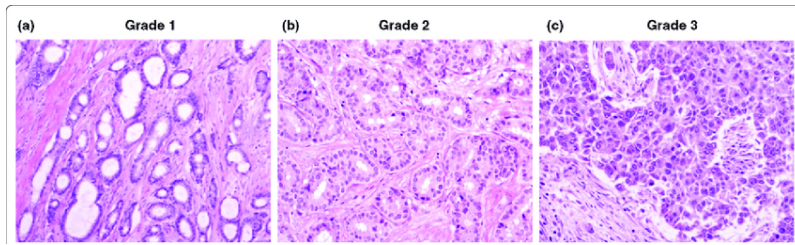


Figura 1. Grado histológico del cáncer de mama según la evaluación del sistema de clasificación de Nottingham donde se observan cambios notables en la formación tubular Rakha et al (2010), Recuperado de ¹⁸

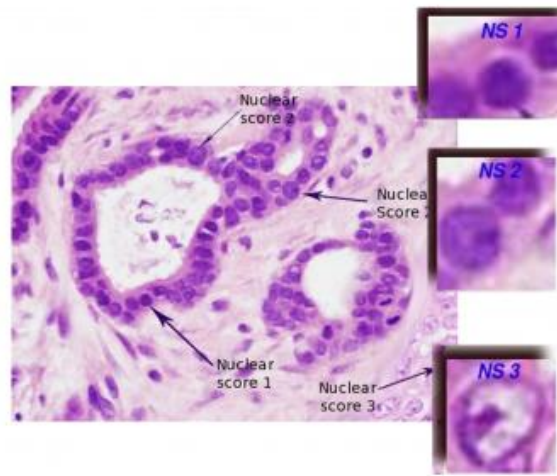


Figura 2. Diferentes grados de formación tubular Moncayo et al, Recuperado de ¹⁹

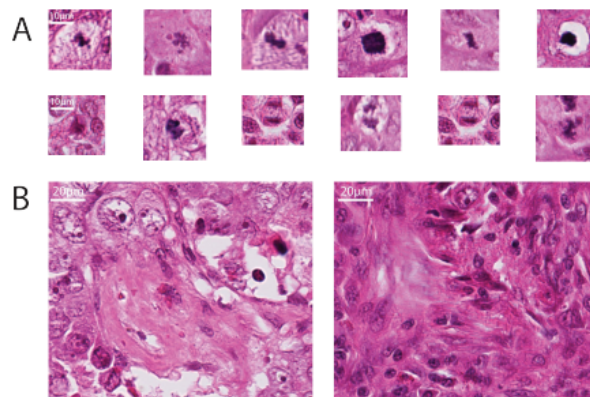


Figura 3. (A) Campos histológicos que contienen células mitóticas. (B) Datos brutos de dos muestras diferentes que no contienen células mitóticas. Sommer et al (2012), Recuperado de ²⁰

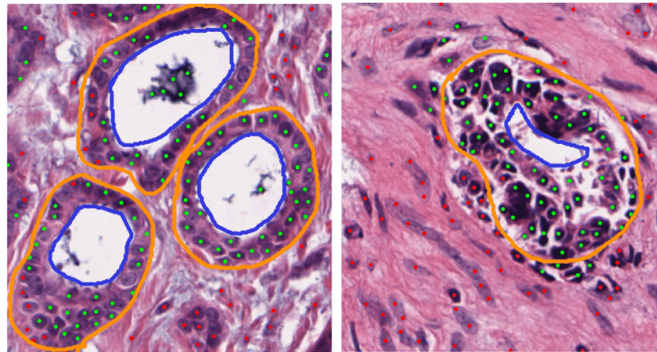


Figura 4. El tejido del cáncer de mama muestra una formación de túbulo alto (izquierda) y bajo (derecha). El lumen está delimitado por líneas azules. Los túbulo están delimitados por líneas naranjas. Romo et al (2016), Recuperado de ²²

El principio básico sigue siendo la suma de las puntuaciones para las tres variables, cada una de las cuales son asignadas de uno a tres puntos según el grado de desviación del epitelio mamario normal. Una puntuación total de 5 o menos define el grado 1, 6 a 7 puntos en el grado 2 y 8 a 9 puntos en el grado 3. La fuerza de la relación pronóstica no ha alcanzado los niveles necesarios para las decisiones terapéuticas críticas, y la reproducibilidad de la clasificación sigue siendo un motivo de preocupación ²¹.

1.2. Análisis Automático en Imágenes de Cáncer de Mama

La adopción de la patología digital se ha incrementado en años recientes, el análisis de imágenes automatizado de la morfología de los tejidos tiene el potencial de disminuir la carga de trabajo de los patólogos, reducir los tiempos de respuesta para informar y estandarizar las prácticas clínicas, las características histopatológicas se pueden cuantificar automáticamente, Además, gracias a el aprendizaje profundo se pueden emplear técnicas para reconocer patrones morfológicos dentro del espécimen para fines de diagnóstico y clasificación ²³.

²¹ Michael H Ross y Pawlina Wojciech. *Histología*. 2013.

²³ Neofytos Dimitriou, Ognjen Arandjelović y Peter D Caie. “Deep learning for whole slide image analysis: An overview”. En: *Frontiers in Medicine* 6 (2019).

En histología de cáncer de mama, diversos trabajos han abordado el desarrollo de técnicas automáticas para cuantificar y correlacionar características visuales observadas en las imágenes histológicas con el grado patológico manual u otras variables clínicas. Un análisis completo de las características morfológicas cuantificadas automáticamente podría identificar características de relevancia pronóstica, proporcionando un medio preciso y reproducible para evaluar el pronóstico a partir de datos de imágenes microscópicas. Recientemente, los modelos de aprendizaje profundo han hecho un progreso notable en la visión por computadora, específicamente en el procesamiento de imágenes biomédicas, debido a sus habilidades para aprender automáticamente características avanzadas de las imágenes, que inspiraron a varios investigadores a aprovechar estos modelos en la clasificación de las imágenes histopatológicas del cáncer de mama²³, a continuación se nombraran algunos trabajos, en ⁴ se desarrollo un un sistema CAD para medir un conjunto de características cuantitativas del epitelio y el estroma del cáncer de mama (En total 6642 características). Las características incluyen descriptores morfométricos estándar de objetos de imagen y características de imagen contextual, relacional y global de alto nivel. Estas características fueron comparadas con la supervivencia de pacientes en un estudio que incluyó 576 pacientes. En ²⁴ se busco evaluar la medición extraída por computadora de la orientación nuclear combinada con características relacionadas con la forma y textura nucleares, que son predictivas de agresividad y pronóstico en etapas tempranas,²⁵ propone una aplicación práctica del modelo de red neuronal de aprendizaje profunda para determinar la puntuación de pleomorfismo nuclear del tejido del cáncer de mama. La red neuronal está capacitada para clasificar una imagen de histología de cáncer de mama mediante el aprendizaje de la desviación media y estándar de las características morfológicas y de textura de toda la población de núcleos conte-

²⁴ Cheng Lu y col. “Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers”. En: *Laboratory Investigation* 98.11 (2018), págs. 1438-1448.

²⁵ P Maqlin y col. “Automated nuclear pleomorphism scoring in breast cancer histopathology images using deep neural networks”. En: *International Conference on Mining Intelligence and Knowledge Exploration*. Springer. 2015, págs. 269-276.

nida en una histología mamaria, se encontró que la precisión de clasificación del modelo en los conjuntos de datos de entrenamiento y prueba era del 96 % y 90 % respectivamente, ²⁶ se uso un enfoque de aprendizaje profundo para la clasificación definitiva de imágenes histopatológicas de cáncer de mama no canceroso y carcinoma. En donde se implementaron cuatro modelos diferentes basados en VGG16 y VGG19, los modelos ofrecieron una sensibilidad del 97,73 para la clase de carcinoma y una precisión general del 95,29 %. en este este trabajo ²⁷ propone un enfoque de red neuronal profunda utilizando el aprendizaje de transferencia para clasificar imágenes histológicas de cáncer de mama. La red utilizada es una Inception Resnet V2. Este trabajo es una solución sugerida para el BACH-Challenge ICIAR 2018 y la precisión es de 0,76 en el conjunto de pruebas ciegas. En ²⁸ se ha presentado un framework CNN basado en Deep Multiple Instance Learning (MIL) que hace uso de las bolsas de características que contienen instancias representativas de las imágenes. Una imagen PVC (es decir, una bolsa de características) se etiqueta como benigna si todos sus parches son benignos y como malignos incluso si un solo parche contiene células malignas. El aprendizaje se puede llevar a cabo al nivel de la bolsa de características incluso con etiquetas de parche ruidosas. El rendimiento de este método se evaluó utilizando los conjuntos de datos de cáncer de mama BreakHis, IUPHL y UCSB, donde se logró una precisión del 93,06 %, 96,63 % y 95,83 %, respectivamente. Por ultimo en ²⁹ el método CAD propuesto consta de 4 etapas; preprocesamiento de imágenes, extracción y fusión

²⁶ Asha Das, Madhu S Nair y S David Peter. “Computer-aided histopathological image analysis techniques for automated nuclear atypia scoring of breast cancer: a review”. En: *Journal of digital imaging* 33.5 (2020), págs. 1091-1121.

²⁷ Carlos A Ferreira y col. “Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet v2”. En: *International conference image analysis and recognition*. Springer. 2018, págs. 763-770.

²⁸ Kausik Das y col. “Detection of breast cancer from whole slide histopathological images using deep multiple instance CNN”. En: *IEEE Access* 8 (2020), págs. 213502-213511.

²⁹ Fatma Anwar y col. “Automatic breast cancer classification from histopathological images”. En: *2019 International Conference on Advances in the Emerging Computing Technologies (AECT)*. IEEE. 2020, págs. 1-6.

de características, reducción de características y clasificación. El CAD se basa en características de fusión extraídas con ResNet Deep Convolution Neural Network (DCNN) con características de descomposición de paquetes de wavelets (WPD) e histogramas de gradiente orientado (HOG). La dimensionalidad de los datos de las características se redujeron utilizando el análisis de componentes principales (PCA). Finalmente, las características reducidas se utilizan para entrenar diferentes clasificadores individuales. Donde se obtiene una precisión del 97,1 %.

1.3. Aprendizaje no Supervisado

Las técnicas de aprendizaje profundo han sido ampliamente utilizadas en el análisis automático de imágenes médicas para realizar tareas de clasificación de imagen, clasificación de lesión, detección de órganos o tejidos, segmentación de estructuras anatómicas, registro de imágenes entre otras ³⁰. Estas estrategias que en general comprenden métodos supervisados han sido ampliamente explorados. En contraposición, los métodos de aprendizaje no supervisados han sido menos explorados debido a que su eficacia ha sido menor.

Los métodos de aprendizaje automático generalmente se dividen en supervisados y algoritmos de aprendizaje no supervisados. En donde el aprendizaje supervisado presenta un modelo con un conjunto de datos $D = (x, y)_n^N$ donde $n = 1$ de características de entrada x y pares de etiquetas y . Esta y puede tomar varias formas, dependiendo de la tarea de aprendizaje; en una clasificación la configuración y es generalmente un escalar que representa una etiqueta de clase, mientras que en el caso de regresión puede ser un vector de variables. Los algoritmos de aprendizaje no supervisados procesan datos sin etiquetas y están entrenados para encontrar patrones, como subespacios latentes. Ejemplos de algoritmos de aprendizaje tradicionales no supervisados son el análisis de componentes principales y los métodos de agrupación. El entrenamiento sin supervisión se puede realizar bajo muchas funciones de pérdida diferentes. Un ejemplo es la

³⁰ Geert Litjens y col. "A survey on deep learning in medical image analysis". En: *Medical image analysis* 42 (2017), págs. 60-88.

pérdida de reconstrucción $L(x, \hat{x})$ donde el modelo tiene que aprender a reconstruir su entrada, a menudo a través de un ruido de menor dimensión o representación ³⁰.

El objetivo del aprendizaje no supervisado es aprender algo útil sobre una estructura de datos subyacente sin el uso de etiquetas. Este es un enfoque interesante para nuestro problema debido a que los datos que serán usados corresponden a imágenes sin etiquetas de identificación.

En el aprendizaje no supervisado, la tarea de aprendizaje es ambigua, ya que es posible mapear las entradas infinitamente en muchos subconjuntos, siempre que no haya restricciones.

La mayoría de los enfoques no supervisados tienen como objetivo maximizar la probabilidad distribución de los datos, sujeta a algunas restricciones, con el fin de limitar el espacio de la solución y lograr una agrupación en clúster para la tarea de destino ³⁰.

Para el problema presentado en este trabajo se utilizaron regiones de interés diagnósticas seleccionadas por expertos, que contienen la información esencial para el diagnóstico: los núcleos celulares, las mitosis, los túbulos y otros elementos del tejido de la mama, etc ³¹. Dada la presencia de todas estas estructuras tisulares en las imágenes histológicas se espera que el aprendizaje no supervisado logre encontrar una partición del espacio multidimensional en el cual sea relativamente sencillo separarlos. Adicionalmente se espera que dicha representación tenga una baja dimensionalidad ³².

³¹ Ishikawa S Komura D. “Machine Learning Methods for Histopathological Image Analysis”. En: 16.2 (2018), págs. 34-42.

³² Pendar Alirezazadeh y col. “Representation learning-based unsupervised domain adaptation for classification of breast cancer histopathology images”. En: *Biocybernetics and Biomedical Engineering* 38.3 (2018), págs. 671-683.

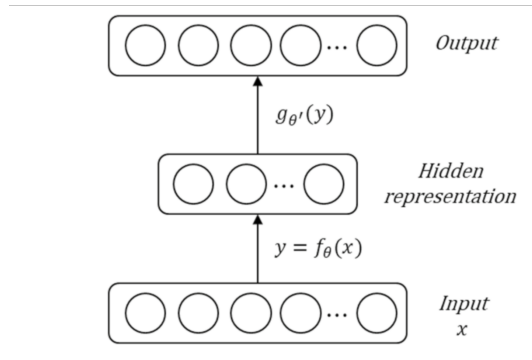


Figura 5. Estructura de Autoencoder. Pendar Alirezazadeh y col (2019), Recuperado de ³³.

1.3.1. Autoencoders En la figura 5 se muestra un ejemplo de la estructura básica de un autoencoder, que es un tipo de red neuronal artificial que se utiliza para aprender codificaciones eficientes de datos no etiquetados (aprendizaje no supervisado) ³⁴, el cual usa un algoritmo de retropropagación para el aprendizaje de características. Funciona en dos fases: codificación y decodificación. En la fase de codificación, los datos de entrada se trasladan a un espacio de representación de baja dimensión para obtener la característica más adecuada, que se traslada de nuevo al espacio de entrada en la fase de decodificación ³⁵. El problema, como se define formalmente en ³⁶ es aprender las funciones $A: \mathbb{R}^n \rightarrow \mathbb{R}^p$ (Encoder) y $B: \mathbb{R}^p \rightarrow \mathbb{R}^n$ (Decoder) que satisface

$$\operatorname{argmin}_{A,B} E[\Delta(x, B \circ A(x))],$$

³⁴ Mark A Kramer. “Nonlinear principal component analysis using autoassociative neural networks”. En: *AICHE journal* 37.2 (1991), págs. 233-243.

³⁵ M Bhuvaneshwari y col. “A comprehensive review on deep learning techniques for a BCI-based communication system”. En: *Demystifying Big Data, Machine Learning, and Deep Learning for Healthcare Analytics* (2021), págs. 131-157.

³⁶ Pierre Baldi. “Autoencoders, unsupervised learning, and deep architectures”. En: *Proceedings of ICML workshop on unsupervised and transfer learning*. JMLR Workshop y Conference Proceedings. 2012, págs. 37-49.

donde E es la expectativa sobre la distribución de X y Δ es la función de pérdida de reconstrucción, que mide la distancia entre la salida del decodificador y la entrada. La siguiente figura provee una ilustración del modelo de un autoencoder.

En la forma más popular de los autoencoders, A y B son redes neuronales ³⁷ en el caso especial de que A y B son operaciones lineales obtenemos un autoencoder lineal. en el caso del autoencoder lineal donde se eliminan las operaciones no lineales, el autoencoder lograría la misma representación latente que principal component analysis (PCA)³⁸. por lo tanto, un autoencoder es de hecho una generalización de PCA donde en lugar de encontrar un hiperplano de baja dimensión en el que se encuentran los datos, es capaz de aprender una variedad no lineal ³⁹.

³⁷ Marc'Aurelio Ranzato y col. "Unsupervised learning of invariant feature hierarchies with applications to object recognition". En: *2007 IEEE conference on computer vision and pattern recognition*. IEEE. 2007, págs. 1-8.

³⁸ Elad Plaut. "From principal subspaces to principal components with linear autoencoders". En: *arXiv preprint arXiv:1804.10253* (2018).

³⁹ Dor Bank, Noam Koenigstein y Raja Giryes. "Autoencoders". En: *arXiv preprint arXiv:2003.05991* (2020).

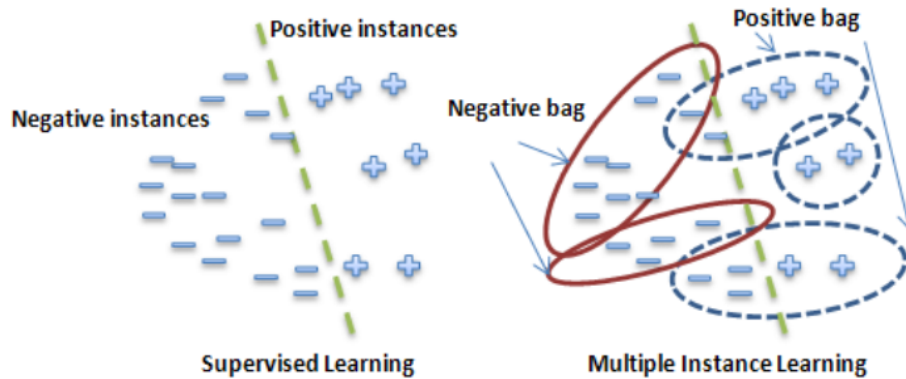


Figura 6. Comparación entre el aprendizaje supervisado y el aprendizaje multi-instancia. Kumar, J y col (2011), Recuperado de ⁴⁰.

1.4. Aprendizaje multi-instancia (multiple-instance learning (MIL))

El aprendizaje (MIL) es una forma de aprendizaje débilmente supervisado en el que las instancias de entrenamiento se organizan en conjuntos, llamados bolsas de características, y se proporciona una etiqueta para toda la bolsa de características. Esta formulación está ganando interés porque se ajusta de forma natural a varios problemas y permite aprovechar los datos débilmente etiquetados. En consecuencia, se ha utilizado en diversos campos de aplicación, como la visión por computador, entre otras. ⁴¹

En el problema clásico de aprendizaje supervisado (binario) el objetivo es encontrar un modelo que prediga un valor de la variable $y \in \{0, 1\}$ para una instancia determinada $x \in D$. Sin embargo en el MIL, en lugar de una única instancia hay una bolsa de instancias, $X = \{x_1, \dots, x_k\}$ que no muestran ni dependencia ni ordenamiento entre sí. Se asume que K puede variar para diferentes bolsas de características. También hay una única etiqueta binaria Y asociada a la bolsa. Además, se asume que existen etiquetas individuales para las instancias de una bolsa, es decir y_1, \dots, y_k y $y_k \in \{0, 1\}$, para $k = 1, \dots, k$ no obstante, no hay acceso a esas etiquetas y se mantendrán

⁴¹ Marc-André Carboneau y col. "Multiple instance learning: A survey of problem characteristics and applications". En: *Pattern Recognition* 77 (2018), págs. 329-353.

desconocidas durante el entrenamiento ⁴². El planteamiento del MIL también es descrito de la siguiente manera:

$$Y = \begin{cases} 0, & \text{Si } \sum_k y_k = 0, \\ 1, & \text{de lo contrario.} \end{cases}$$

Las dos afirmaciones de la ecuación puede reformularse en una forma compacta utilizando el operador máximo operador de la siguiente manera.

$$Y = \max_k y_k$$

Para clasificar correctamente las bolsas de características bajo el supuesto estándar, no es necesario identificar todas las características que componen la bolsa, siempre y cuando se encuentre al menos una característica positiva entonces la bolsa se clasificará como positiva. En ese caso, los falsos positivos (FP) y los falsos negativos (FN) no tienen ningún impacto en la precisión de la clasificación de la bolsa, pero siguen contando como errores de clasificación a nivel de instancia.

⁴² Maximilian Ilse, Jakub Tomczak y Max Welling. “Attention-based deep multiple instance learning”. En: *International conference on machine learning*. PMLR. 2018, págs. 2127-2136.

2. PLANTEAMIENTO Y JUSTIFICACIÓN DEL PROBLEMA

La evaluación del cáncer de mama depende en gran medida del análisis de las imágenes histopatológicas. Este análisis requiere de un tiempo considerable, y además solo puede ser realizado por personal médico con un previo entrenamiento especializado. Debido a las diferentes interpretaciones de los protocolos de evaluación histológica, así como a la diferencia en la selección de regiones de interés diagnóstico, no es raro que estos análisis presenten una variabilidad inter-observador alta. Esta variabilidad limita la reproducibilidad de la evaluación y dificulta un diagnóstico preciso.

Recientemente, varios algoritmos automáticos supervisados para la identificación y graduación de características histopatológicas conocidas han sido propuestos. Estos algoritmos prometen reducir la variabilidad inter-observador para el análisis histopatológico. Sin embargo, debido a su naturaleza supervisada, el entrenamiento y puesta a punto de estos algoritmos requiere de un demandante proceso de anotación por parte de los expertos. Como resultado, usualmente se tiene un conjunto limitado de datos anotados, lo cual limita la generalizabilidad de los algoritmos entrenados. Adicionalmente, estos algoritmos se centran en características histopatológicas ya conocidas por la comunidad médica.

En contraste, técnicas no supervisadas de aprendizaje que están siendo recientemente exploradas en el análisis automático de imágenes médicas prometen identificar características visuales relevantes sin una tarea previa de anotación manual. En el escenario del cáncer de mama, es de particular interés la exploración de estas metodologías no supervisadas para identificar características visuales que puedan estar asociadas a variables clínicamente relevantes, como por ejemplo el grado patológico del cáncer de mama.

3. OBJETIVOS

Objetivo general

Implementar y evaluar una estrategia de aprendizaje profundo no supervisada para la identificación automática de características visuales relevantes en imágenes histopatológicas de cáncer de mama.

Objetivos específicos

- Identificar y seleccionar un conjunto de datos públicos de histologías de cancer de mama teñidas en hematoxilina y eosina.
- Seleccionar e implementar una estrategia de aprendizaje no supervisado para aplicar en imágenes histológicas de cáncer de mama.
- Cuantificar la relación de las características visuales identificadas mediante la estrategia de aprendizaje no supervisado con marcadores patológicos conocidos.
- Validar las características aprendidas en tareas de clasificación.

4. REPRESENTACIÓN LATENTE DE IMÁGENES HISTOLÓGICAS DE CANCER DE MAMA VÍA APRENDIZAJE NO SUPERVISADO

El enfoque presentado en este trabajo trata del reconocimiento de características visuales, implementando y evaluando una red neuronal autoencoder, permitiendo obtener la abstracción de las imágenes histopatológicas.

4.1. Autoencoder para la representación de imágenes histológicas

El autoencoder es una técnica de aprendizaje no supervisado, en donde se le impone un cuello de botella a la red que obliga a comprimir la representación de la entrada original. Esta red consiste en la transformación de datos de alta dimensionalidad a un espacio de menor dimensionalidad, dando así una representación embebida de características significativas de la entrada inicial.

Los autoencoders se dividen en dos subredes, la subred encoder y la subred decoder, para nuestro problema se diseñó la siguiente arquitectura que se muestra en la figura 7.

4.1.1. Encoder o el codificador E es un conjunto de bloques convolucionales seguidos de módulos de agrupación que comprimen la entrada del modelo en este caso una imagen x reduciendo así la dimensión de x para encontrar la representación embebida de las características $z = E(x)$.

4.1.2. Decoder también denominado decodificador D compuesto de un conjunto de bloques convolucionales que realizan el proceso inverso del encoder, donde se utiliza la representación embebida z previamente obtenida y se descomprime $\hat{x} = D(z)$.

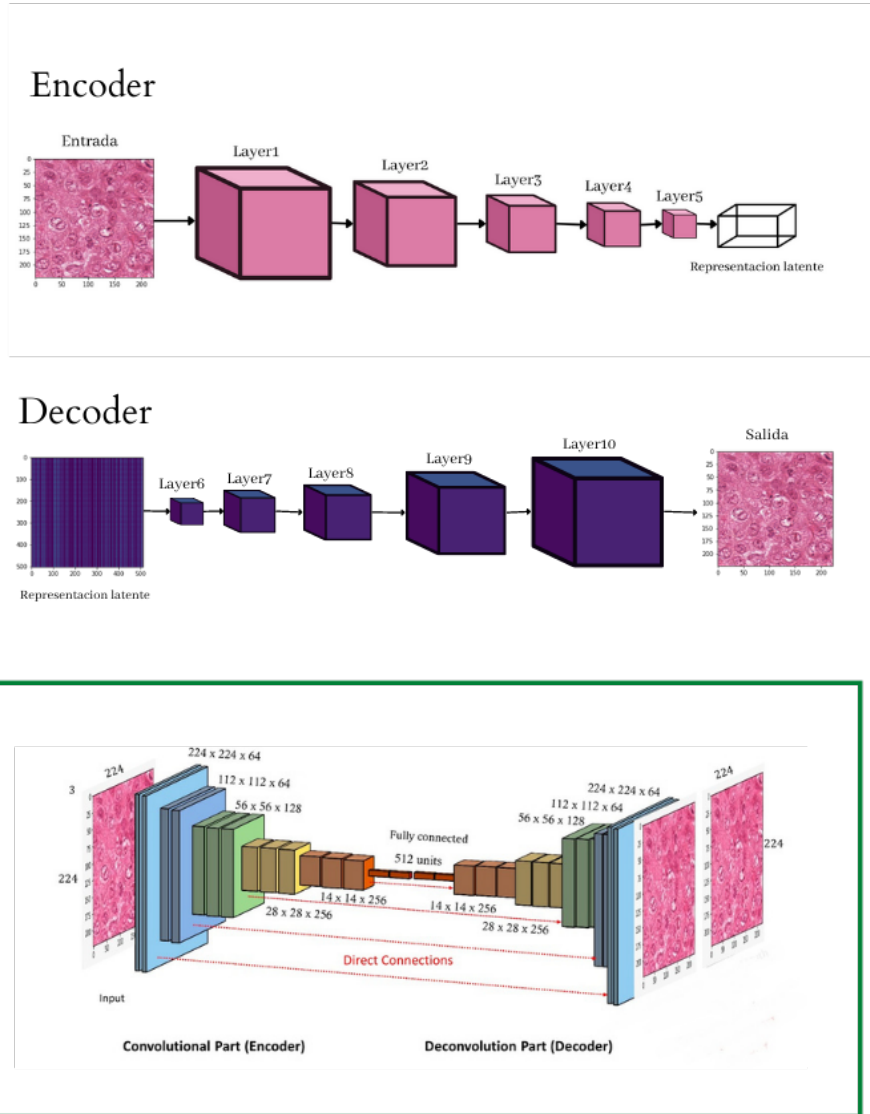


Figura 7. Arquitectura de la estrategia no supervisada con conexiones directas entre el encoder/decoder con embebidos de tamaño 512.

4.2. Conjunto de datos MITOS ATYPIA 2014

Para este estudio se utilizó el dataset proporcionado por el Mitos Atypia 14 challenge en donde se ha seleccionado y anotado un conjunto de imágenes de biopsia de cáncer de mama. Las imágenes están teñidas con tintes estándar de hematoxilina y eosina (HE) y han sido escaneados por dos dispositivos: Aperio Scanscope XT y Hamamatsu Nanozoomer 2.0-HT.

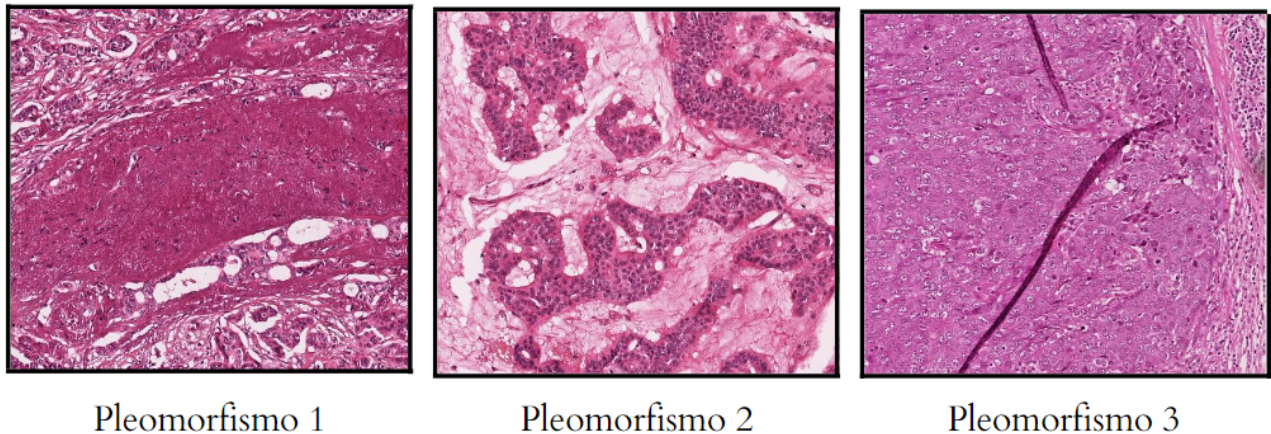


Figura 8. Ejemplo de campos histológicos a 20X con grado pleomórfico asignado por un experto ⁴³.

En cada imagen, los patólogos seleccionaron varios campos histológicos con un aumento de 20X situados en el interior de los tumores. Estos campos histológicos 20X se utilizan para puntuar la atipia nuclear. Los campos histológicos 20X se han subdividido en cuatro campos histológicos con aumento 40X. Los campos histológicos 40X se utilizan para anotar la mitosis y dar una puntuación a seis criterios relacionados con la atipia nuclear.

El número de campos histológicos varía de una diapositiva a otra. En el conjunto de datos de entrenamiento hay 284 campos histológicos con un aumento de 20X y 1.136 campos histológicos con un aumento de 40X. Los campos histológicos son imágenes de mapa de bits RGB en formato TIFF. Para los campos histológicos de aumento 20X, proporcionamos la puntuación de atipia nuclear como un número 1, 2 o 3. La puntuación 1 denota una atipia de bajo grado, la puntuación 2 una atipia de grado moderado y la puntuación 3 una atipia de alto grado. Esta puntuación

ha sido dada de forma independiente por dos patólogos diferentes de alto nivel ⁴⁴.

De este dataset se tomaron 11 pacientes con sus respectivos campos histológicos de aumento 20X, en donde cada cada imagen de 1000 x 1000 se denominará campo histológico, ahora bien, teniendo en cuenta que se tiene un conjunto de datos pequeño se decidió extraer parches de cada uno de estos campos en distintos tamaños de los cuales se escogieron 224x224,128x128 y 64x64 pixeles, por último, los parches fueron normalizados y organizados de nuevo por campo visual correspondiente. Para el entrenamiento del autoencoder se usaron 8 pacientes los cuales contienen los tres grados de pleomorfismo (1,2 y 3) y para su evaluación los dos pacientes restantes con solo dos grados de pleomorfismo (2,3).

4.3. Entrenamiento del autoencoder y diseño experimental

Para el entrenamiento del modelo se usaron como datos de entrada los parches de los campos visuales previamente descritos, para configurar el autoencoder se usaron 50 epochs, batchsize de 64 donde se consiguió un espacio latente de 512 unidades. Ya finalizada esta fase, se procede a la evaluación del modelo utilizando a los pacientes de testeo teniendo en cuenta la pérdida descrita como.

$$l = ||x - \hat{x}||^2$$

donde

$$\hat{x} = z = D(E(x))$$

El objetivo principal de este experimento fue obtener un mapeo de las características representativas de los parches de entrada x a un espacio latente z .

⁴⁴ *MITOS-ATYPIA-14 - Dataset.*

4.4. Métricas de evaluación para la reconstrucción en el autoencoder

En esta sección se discutirá a profundidad las métricas de validación que fueron utilizadas para estimar la reconstrucción de los vectores embebidos, para esto se tuvo en cuenta los siguientes términos:

- **Mean Squared Error (MSE):** El error medio cuadrático se calcula entre la imagen original y la imagen procesada se utiliza a menudo como criterio de calidad en el procesamiento de imágenes. Esta mide la similitud entre la imagen imagen original y la procesada, y se define como.

$$MSE = \frac{1}{MN} \sum_{n=0}^M \sum_{m=1}^N [\hat{x}(n, m) - x(n, m)]^2$$

- **Mean Absolute Error (MAE):** El error absoluto medio de un modelo con respecto a un conjunto de pruebas es la media de los valores absolutos de los errores de predicción individuales sobre todas las instancias del conjunto de pruebas. Cada error de predicción es la diferencia entre el valor real y el valor predicho para la instancia.

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n}$$

- **Peak signal-to-noise ratio (PSNR):** Se utiliza habitualmente para medir la calidad de la reconstrucción para la compresión de imágenes. En este caso, la señal son los datos originales y el ruido es el error introducido por la compresión.

$$PSNR = 10 \log_{10}(P^2/MSE)$$

Donde P es la fluctuación máxima del tipo de datos de la imagen de entrada.

- **Pearson Correlation Coefficient (PCC):** El coeficiente de correlación de Pearson es la

covarianza de las dos variables dividida por el producto de sus desviaciones estándar.

$$r_1 = \frac{\sum_i (x_i - x_m)(y_i - y_m)}{\sqrt{\sum_i (x_i - x_m)^2} \sqrt{\sum_i (y_i - y_m)^2}}$$

donde x_i es la intensidad del píxel i^n de la imagen 1, y_i es la intensidad del píxel i^n de la imagen 2, x_m es la intensidad media de la imagen 1, y y_m es la intensidad media de la imagen 2.

4.5. Resultados de reconstrucción del autoencoder

A continuación se exponen los resultados logrados para la representación comprimida (latente) de características visuales obtenidas por la arquitectura tipo autoencoder. En nuestro trabajo esta sección es importante ya que podemos cuantificar cuán buena es la representación resultante del modelo aplicado. En la siguiente tabla se muestran los resultados de errores de reconstrucción promedio tanto para los pacientes de entrenamiento como los de validación.

Tamaño	MSE Train	MSE Test	MAE Train	MAE Test	PSNR Train	PSNR Test
224 × 224	0.00032	0.00032	0.00342	0.00880	88.258	88.244
128 × 128	0.00076	0.00076	0.00671	0.00886	84.408	84.407
64 × 64	0.00072	0.00072	0.01258	0.01240	84.413	84.401

Tabla 1. Tabla de resultados de las métricas MSE y MAE de la reconstrucción de imágenes

Así mismo también se calculó el PSNR para la medición de la calidad de reconstrucción a partir de los vectores embebidos. En las figuras se muestran los resultados de las métricas MSE y PSNR para los diferentes tamaños de parches usados.

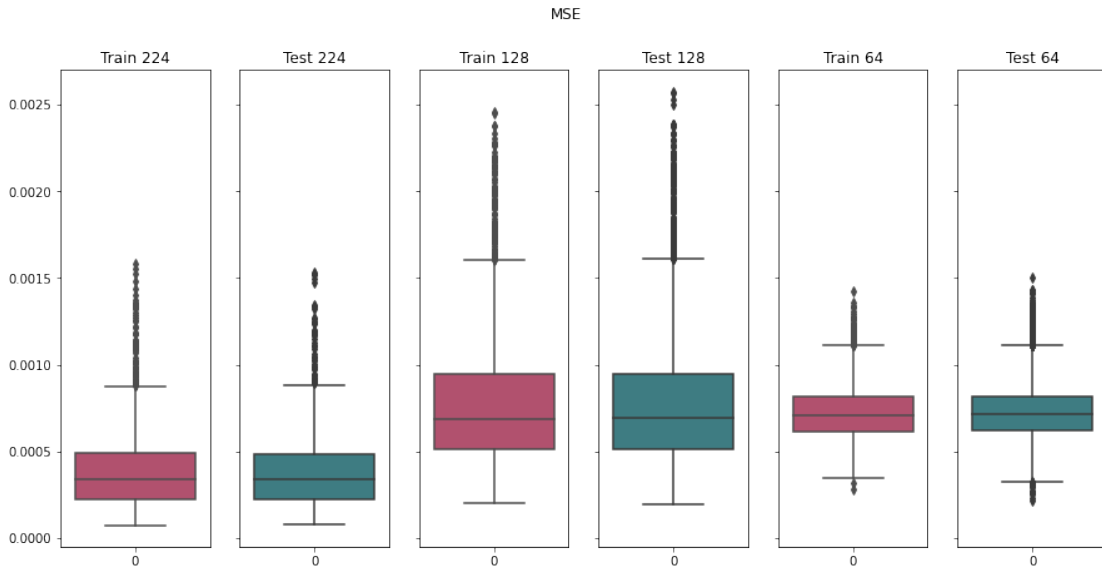


Figura 9. Diagrama de cajas de MSE para todos los tamaños de parches para los cuales se calculó el promedio (desviación estandar) para los parches de tamaño 224x224: $1.4e-3$ ($3.2e-3$), 128x128: $3.1e-3$ ($7.6e-3$), 64x64: $1.4e-3$ ($7.2e-3$)

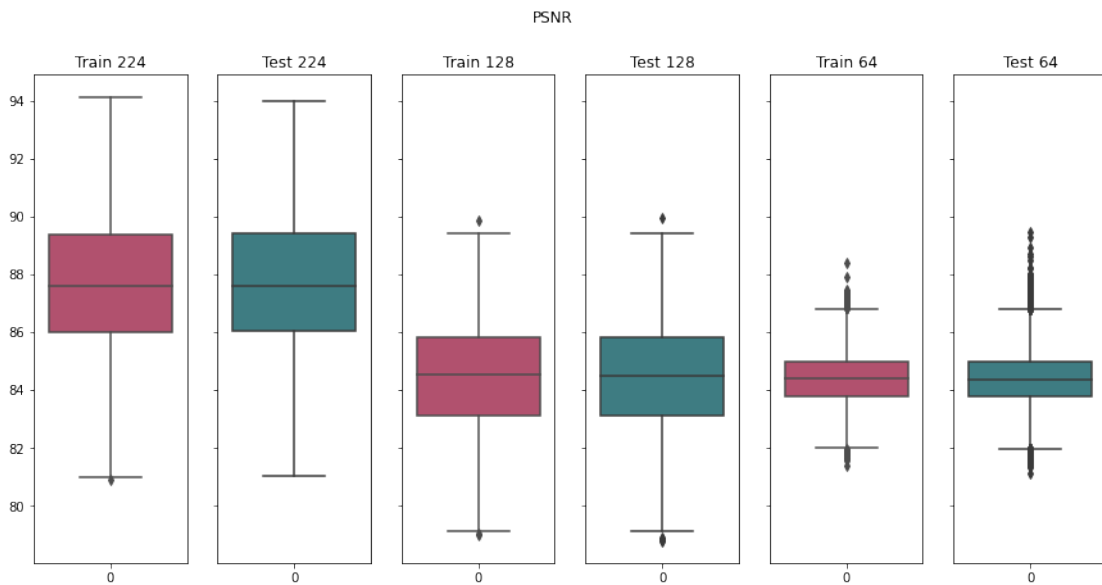


Figura 10. Diagrama de cajas de PSNR para todos los tamaños de parches para los cuales se calculó el promedio (desviación estandar) para los parches de tamaño 224x224: 1.87 (88.24) , 128x128: 1.65 (84.40), 64x64: 0.90 (84.40)

En un análisis adicional de los vectores embebidos se calculó la visualización UMAP para observar la distribución de los datos (ver figura 11) donde se visualizó que los datos ocupan un espacio similar en la visualización 2D para todas las configuraciones, dándonos a conocer que no hay solapamiento a nivel de pacientes entre los conjuntos train/validation/test. Además se calculó la correlación con el coeficiente de Pearson para las 512 características, esto con la finalidad de examinar la fuerza de asociación entre las características. Podemos concluir que no existe una correlación fuerte entre las variables, por ejemplo para el tamaño de parches 224x224 el valor máximo que toma el coeficiente es de aproximadamente 0,1511 y mínimo de $-0,1407$, en 128x128 los coeficientes toman valores de 0,0778 y $-0,1146$ y en 64x64 los coeficientes toman valores de 0,0653 y $-0,0674$ (ver figura 12).

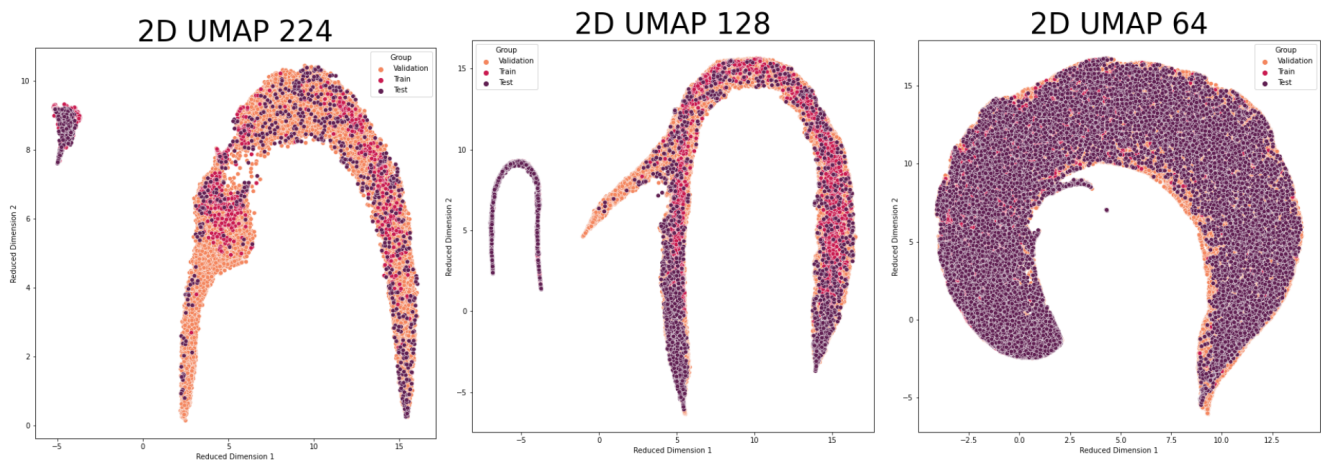


Figura 11. Gráficas UMAP para los embebidos (parches de tamaño 224x224, 128x128 y 64x64) diferenciando las particiones train/validation/test. Obsérvese que las representaciones train/validation/test ocupan un espacio similar en la visualización 2D para todas las configuraciones. Es importante subrayar que no hay solapamiento a nivel de pacientes entre los conjuntos train/validation/test.

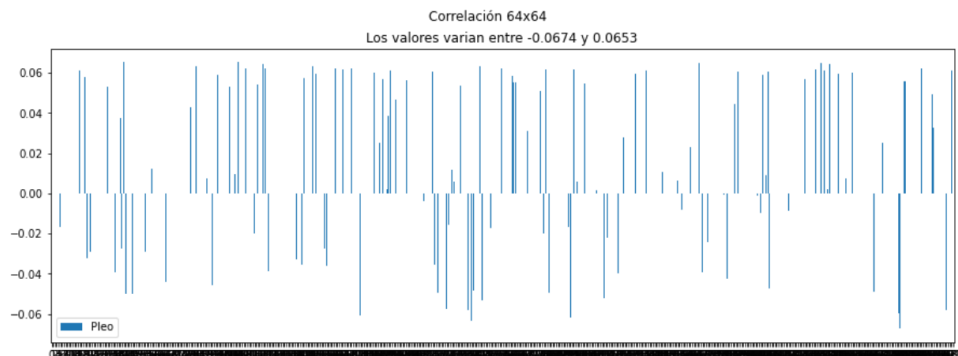
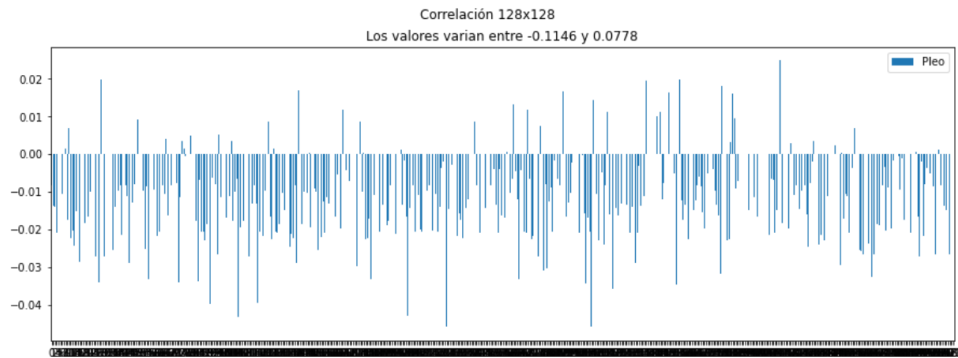
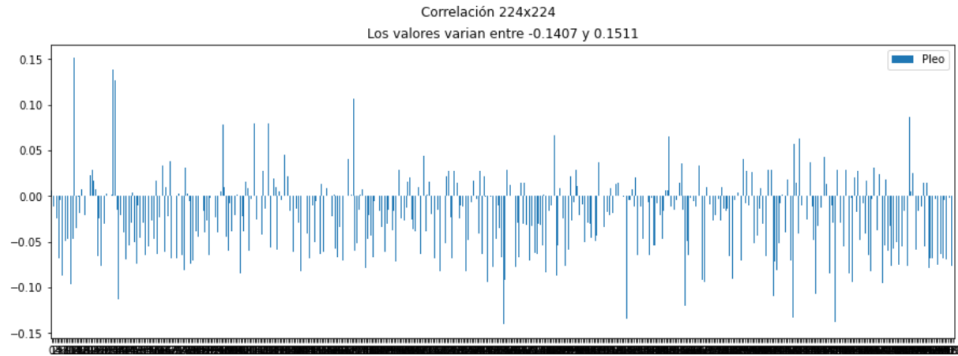


Figura 12. Gráficas de correlación para todos los tamaños de parches en el eje X se encuentran las características y en el eje Y la correlación. Obsérvese que la correlación encontrada entre el grado pleomórfico y las características es cercana a 0

5. APRENDIZAJE MULTI-INSTANCIA USANDO REPRESENTACIÓN LATENTE PARA EL PLEOMORFISMO NUCLEAR EN CÁNCER DE MAMA

5.1. Arquitectura de red neuronal para el aprendizaje Multi-instancia

Uno de los objetivos propuestos consiste en validar las características aprendidas en tareas de clasificación, para esto se decidió utilizar la estrategia de Multiple instance learning (MIL). Para esta sección se hizo uso de los embebidos que fueron posteriormente convertidos a bolsas de características, con esto se procedió a las tareas de clasificación en donde se hizo uso los modelos que se muestran en las siguientes figuras.

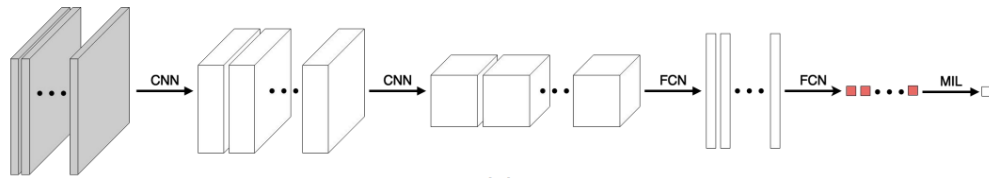


Figura 13. Enfoque profundo de MIL. Ilse y col (2018), Recuperado de ⁴⁵.

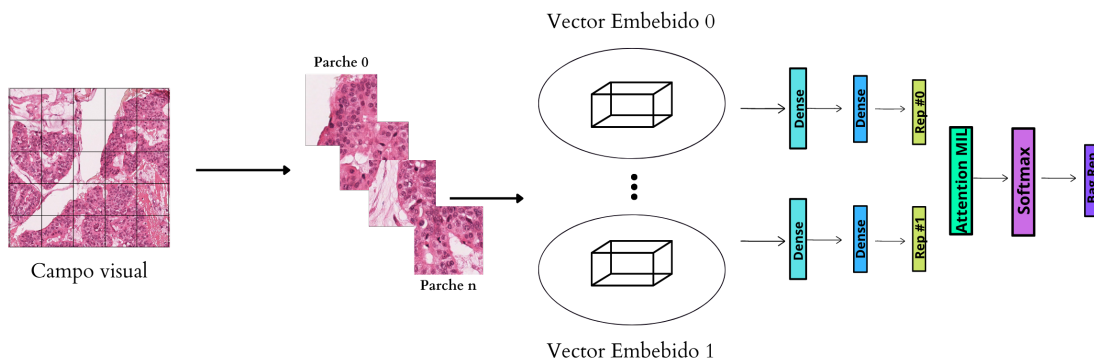


Figura 14. estrategia MIL basada en embebidos para la clasificación de grados de pleomorfismo

Este modelo está basado en la arquitectura propuesta en el artículo Attention-based Deep Multiple Instance Learning por MIL. Ilse y col el cual consiste de estos tres pasos: (i) transformación de las instancias a una incrustación de baja dimensión, (ii) una función de agregación (simétrica), y (iii) transformación final a la probabilidad de la bolsa. Esto traducido a la figura 14, se toman las bolsas las cuales pasan por una red neuronal que reduce su dimensión de 512 unidades a 64 unidades, luego llegan a la capa de attention-based Deep MIL, finalmente por una capa de softmax que entrega la probabilidad de la etiqueta de la bolsa.

5.2. Conjunto de datos: Embebidos - MITOS ATYPIA

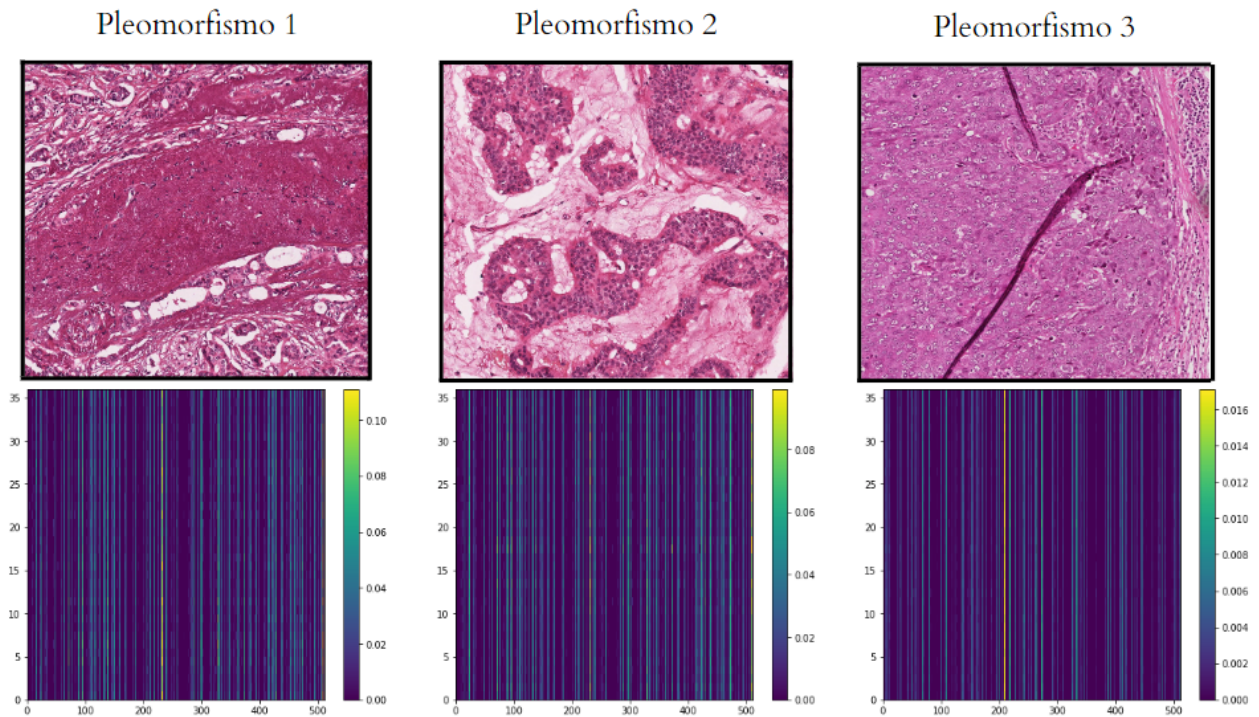


Figura 15. Estos embebidos corresponden a parches de 224x224. Obsérvese que para cada campo visual se utilizan 36 vectores embebidos en su representación

Para esta sección del proyecto se utilizaron los vectores embebidos que anteriormente se obtuvieron del modelo Autoencoder contemplado en la sección 4. La formación de bolsas se realizó partiendo de que cada campo visual es considerado como una única bolsa, las etiquetas asignadas

a los campos visuales fueron binarizadas de tal forma que toda etiqueta que fuese de pleomorfismo tipo 1 o 2 es igual a 0 y si es pleomorfismo tipo 3 es igual a 1. Sin importar el tamaño de los parches la cantidad de bolsas se mantuvo de la siguiente manera: para entrenamiento se tuvieron 151 bolsas en entrenamiento y 32 bolsas para evaluación. No obstante el tamaño de los parches si alteró la cantidad de instancias o características presentes en cada bolsa es decir que para los tamaños 224x224 cada bolsa contiene 36 instancias, 128x128 contiene 120 instancias por bolsa y 64x64 contiene 504 instancias por bolsa. También es importante recalcar que cada instancia es un embebido de 512 unidades por consiguiente la configuración de cada bolsa según el tamaño del parche es de la siguiente manera: para 224x224 (1,36,512), 128x128 (1,120,512) y 64x64 (1,504,512).

5.3. Entrenamiento del modelo multi-instancia y diseño experimental

En el entrenamiento del modelo, partimos de la idea de que una bolsa es clasificada como "positiva" si contiene la etiqueta de pleomorfismo nuclear tipo 3. Para este caso los datos de entrada son las bolsas de características asignadas para esta sección anteriormente descritas, para la estructuración del Attention MIL se utilizaron 300 epoch, con un batchsize de 256. ya terminada esta fase, procedemos a la fase de evaluación en donde se usaron un total de 32 bolsas y se procedió a su clasificación.

5.4. Métricas de evaluación para la clasificación multi-instancia

Seguidamente se hará la descripción de las métricas de validación que se emplearon en el presente trabajo para la cuantificación y análisis del modelo propuesto. Estas métricas denotan una contabilización de las predicciones para clasificación binaria, se tuvieron en cuenta los siguientes términos:

- **Verdaderos positivos (VP):** Son aquellos resultados que se predijeron correctamente como positivos. En nuestro contexto corresponde a los embebidos que fueron clasificados con su pleomorfismo tipo 3.

- **Falsos positivos (FP):** Son aquellos resultados que se predijeron incorrectamente como negativos. En nuestro contexto corresponde a los embebidos que fueron clasificados con un pleomorfismo tipo 1 o 2 siendo tipo 3.
- **Verdaderos Negativos (VN):** Son aquellos resultados que se predijeron correctamente como negativos. En nuestro contexto corresponde a los embebidos que fueron clasificados con su pleomorfismo tipo 1 o 2.
- **Falsos negativos (FN):** Son aquellos resultados que se predijeron incorrectamente como negativos. En nuestro contexto corresponde a los embebidos que fueron clasificados con un pleomorfismo tipo 3 siendo tipo 1 o 2.
- **Matriz de Confusión:** Es una herramienta que permite visualizar el rendimiento de un algoritmo, en donde cada fila representa las instancias de una clase real, mientras tanto las columnas tienen instancias de una clase predicha.
- **Curva ROC:** (receiver operating characteristic curve) es un gráfico que muestra la capacidad de rendimiento de un modelo en todos los umbrales de clasificación.

Esta curva se genera usando la sensibilidad VS la tasa de falsos positivos que se calcula como (1 - especificidad). Para hallar el ROC de una clasificación, se utiliza la medida de área bajo la curva (AUC). El AUC proporciona una medida agregada del rendimiento para todos los umbrales de clasificación posibles.

- **Sensibilidad:** Mide la aptitud del modelo para detectar eventos de la clase positiva.

$$Sensibilidad = \frac{VP}{VP + FN}$$

- **Especificidad:** Mide la exactitud del modelo para la asignación a la clase positiva

$$Especificidad = \frac{VN}{FP + VN}$$

5.5. Evaluación y resultados

A continuación se presentan los resultados alcanzados en la validación de características aprendidas en tareas de clasificación usando como base una arquitectura tipo Multiple instance learning. La experimentación se basó en la clasificación de las bolsas compuestas por las representaciones embebidas elaboradas por el Autoencoder. El enfoque propuesto se comparó con otros trabajos del estado del arte (ver Tabla 2) como Lu y col ⁴⁶ en donde proponen una segmentación en todas las regiones de los núcleos. Extraen un conjunto de características morfológicas y de textura de las regiones de los núcleos presegmentados. Aplicaron un clasificador de máquina de soporte vectorial para clasificar una imagen de campo de alta potencia en diferentes clases de atipia nuclear. En Khan y col⁴⁷, se presentó una media geométrica generalizada del descriptor de covarianza de la región como un descriptor a nivel de imagen junto con el clasificador GkNN para la puntuación de atipia nuclear en imágenes histopatológicas de cáncer de mama. Por último en Xu y col⁴⁸ utilizaron una red convolucional multirresolución (MR-CN) con Votación Plural (MR-CN-PV) para la puntuación de atipia nuclear automatizada. MR-CN-PV consiste en tres redes convolucionales de resolución simple (SR-CN) con voto mayoritario (SR-CN-MV) para obtener una puntuación de atipia nuclear. Para el Attention deep MIL se obtuvo un F1 Score promedio entre todos los tamaños de parches.

En un análisis más detallado se recurrió a la curva ROC para observar el comportamiento de la arquitectura, además se calculó el AUC y el índice de Youden que indica el rendimiento

⁴⁶ Cheng Lu y col. “Automated image analysis of nuclear atypia in high-power field histopathological image”. En: *journal of microscopy* 258.3 (2015), págs. 233-240.

⁴⁷ Adnan Mujahid Khan, Korsuk Sirinukunwattana y Nasir Rajpoot. “A global covariance descriptor for nuclear atypia scoring in breast histopathology images”. En: *IEEE journal of biomedical and health informatics* 19.5 (2015), págs. 1637-1647.

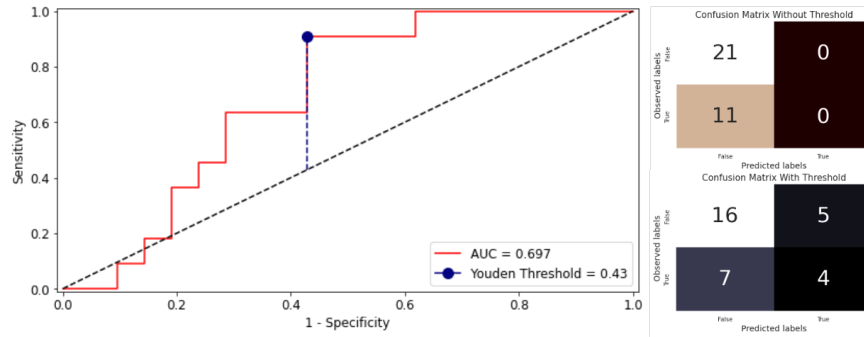
⁴⁸ Jun Xu y col. “Deep learning for histopathological image analysis: Towards computerized diagnosis on cancers”. En: *Deep Learning and Convolutional Neural Networks for Medical Image Computing*. Springer, 2017, págs. 73-95.

	<i>Algoritmo</i>	<i>F1 Score</i>	<i>Tipo de Clasificación</i>
Lu y col.	SVM-Sigmoid	0.1789	clasificación multicategoría. Grados 1, 2, y 3
Rezailouyen y col.	Neural Network	0.1789	clasificación multicategoría. Grados 1, 2, y 3
Xu y col.	Deep learning	0.5509	clasificación multicategoría. Grados 1, 2, y 3
Modelo propuesto	Attention MIL	0.3157	clasificación binaria. Grados 1,2 vs 3

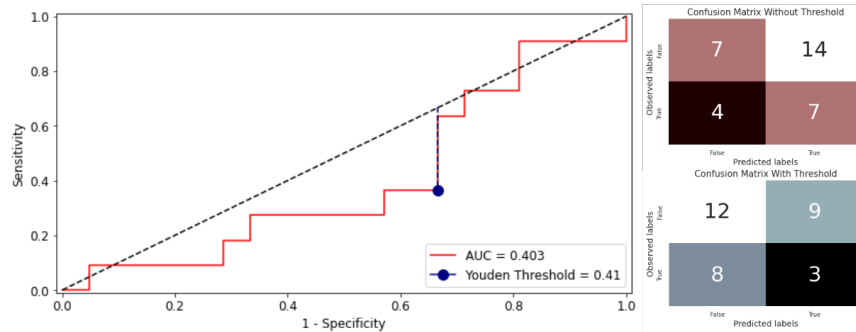
Tabla 2. Tabla de resultados de diferentes métodos en categorización de pleomorfismo nuclear

que nos permite tener un umbral de precisión del modelo, el cual optimizó la clasificación de pleomorfismo en los pacientes de prueba, estos resultados fueron ilustrados en una matriz de confusión. En las siguientes figuras se muestran estos cálculos por cada tamaño de parche.

a) 224



b) 128



c) 64

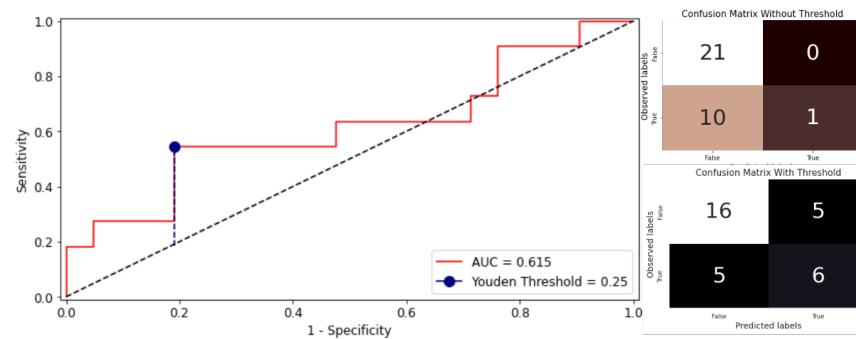


Figura 16. Curva ROC, curva AUC y matrices de confusión para la clasificación de pleomorfismo nuclear, en el panel superior (a) se encuentra el resultado de clasificación para los embebidos de los parches con tamaño (224x224), en el panel (b) los resultados para los parches de tamaño(128x128) y el panel (c) son los resultados para los parches de tamaño (64x64).

6. CONCLUSIONES Y PERSPECTIVAS

En este trabajo se realizó un reconocimiento de características visuales para imágenes histopatológicas de cáncer de mama, implementando una estrategia de aprendizaje profundo no supervisado. La metodología propuesta se fundamenta en la red neuronal Autoencoder, la cual generó una representación comprimida de características relevantes. Sin embargo, se evidenció que la fase de entrenamiento fue favorecida con los experimentos que contienen un mayor volumen de datos de entrada. Es por esto que se usó como recurso la generación de parches de diferentes tamaños por campo visual para aumentar de manera significativa los datos iniciales que ingresan en las capas de entrada con el fin de mejorar la reconstrucción de las imágenes. Al evaluar la reconstrucción se pudo visualizar valores de PSNR superiores a 88,244, donde la configuración de 224x224 obtuvo en promedio 88,244 en 128x128 84,407 y 64x64 con 84,401, esto se debe a que la arquitectura presenta skip connections en el encoder/decoder.

Para analizar las representaciones embebidas se procedió a utilizar una técnica de reducción de dimensionalidad denominada UMAP, de lo cual se concluyó que visualmente los datos de entrenamiento, validación y evaluación ocupan el mismo espacio en una visualización 2D, lo cual indica que el embebido generaliza adecuadamente es decir que no existe un solapamiento entre los conjuntos a nivel de pacientes. Además de esto, también se evaluó de manera individual los embebidos en donde se encontró una correlación muy baja entre los embebidos y el grado pleomórfico.

En el transcurso del ciclo de validación, se implementó una arquitectura de clasificación automática MIL, la cual antes de generar la clasificación reducía de manera significativa las características pasando de tener 512 a 64 que benefició las predicciones obtenidas. Además de utilizar 10 pacientes que contienen los tres tipos de pleomorfismo para entrenamiento y solo tener pacientes con evidencias de dos tipos de pleomorfismo para la evaluación. En este estudio se comprobó

la hipótesis que a mejor reconstrucción hay una mejor clasificación, con el modelo de parches 224x224 se verificó dicha hipótesis, puesto que logró un resultado de prueba del 0,697 en la métrica curva ROC AUC a comparación de las configuraciones de 128x128 y 64x64 que lograron 0,403 0,615 respectivamente, además de tener el PSNR más alto, lo cual no se encuentra tan lejos de las implementaciones presentadas en el estado del arte. Con esto concluimos que el enfoque MIL combinado con la representación latente fueron una estrategia adecuada para diferenciar el pleomorfismo nuclear en niveles 1 vs niveles 2/3.

Con este trabajo podemos concluir que la representación de los autoencoders es prometedora y que es necesario profundizar en la investigación de estas representaciones y en el aprendizaje multi-instancia para validar su capacidad en la representación del pleomorfismo nuclear.

BIBLIOGRAFÍA

- Al-Janabi, Shaimaa, André Huisman y Paul J Van Diest. “Digital pathology: current status and future perspectives”. En: *Histopathology* 61.1 (2012), págs. 1-9 (vid. pág. 15).
- Alirezazadeh, Pendar y col. “Representation learning-based unsupervised domain adaptation for classification of breast cancer histopathology images”. En: *Biocybernetics and Biomedical Engineering* 38.3 (2018), págs. 671-683 (vid. pág. 22).
- Anwar, Fatma y col. “Automatic breast cancer classification from histopathological images”. En: *2019 International Conference on Advances in the Emerging Computing Technologies (AECT)*. IEEE. 2020, págs. 1-6 (vid. pág. 20).
- Araújo, Teresa y col. “Classification of breast cancer histology images using convolutional neural networks”. En: *PloS one* 12.6 (2017), e0177544 (vid. pág. 13).
- Baldi, Pierre. “Autoencoders, unsupervised learning, and deep architectures”. En: *Proceedings of ICML workshop on unsupervised and transfer learning*. JMLR Workshop y Conference Proceedings. 2012, págs. 37-49 (vid. pág. 23).
- Bank, Dor, Noam Koenigstein y Raja Giryes. “Autoencoders”. En: *arXiv preprint arXiv:2003.05991* (2020) (vid. pág. 24).
- Bayramoglu, Neslihan, Juho Kannala y Janne Heikkilä. “Deep learning for magnification independent breast cancer histopathology image classification”. En: *2016 23rd International conference on pattern recognition (ICPR)*. IEEE. 2016, págs. 2440-2445 (vid. pág. 12).

- Beck, Andrew H y col. “Systematic analysis of breast cancer morphology uncovers stromal features associated with survival”. En: *Science translational medicine* 3.108 (2011), 108ra113-108ra113 (vid. págs. 11, 19).
- Bhuvaneshwari, M y col. “A comprehensive review on deep learning techniques for a BCI-based communication system”. En: *Demystifying Big Data, Machine Learning, and Deep Learning for Healthcare Analytics* (2021), págs. 131-157 (vid. pág. 23).
- Cancer, Liga Colombiana Contra el. *Cancer de Seno*. <https://www.ligacancercolombia.org/cancer-de-seno/>. [Online; accessed 11-August-2020]. 2018 (vid. pág. 11).
- Carbonneau, Marc-André y col. “Multiple instance learning: A survey of problem characteristics and applications”. En: *Pattern Recognition* 77 (2018), págs. 329-353 (vid. pág. 25).
- Chan, Heang-Ping, Lubomir M Hadjiiski y Ravi K Samala. “Computer-aided diagnosis in the era of deep learning”. En: *Medical Physics* 47.5 (2020), e218-e227 (vid. pág. 14).
- Choi, Hyunseung y col. “Unsupervised learning approach for network intrusion detection system using autoencoders”. En: *The Journal of Supercomputing* 75.9 (2019), págs. 5597-5621 (vid. pág. 23).
- Cosatto, Eric y col. “Grading nuclear pleomorphism on histological micrographs”. En: *2008 19th International Conference on Pattern Recognition*. IEEE. 2008, págs. 1-4 (vid. pág. 13).
- Dalton, Leslie W, David L Page y William D Dupont. “Histologic grading of breast carcinoma. A reproducibility study”. En: *Cancer* 73.11 (1994), págs. 2765-2770 (vid. pág. 16).
- Das, Asha, Madhu S Nair y S David Peter. “Computer-aided histopathological image analysis techniques for automated nuclear atypia scoring of breast cancer: a review”. En: *Journal of digital imaging* 33.5 (2020), págs. 1091-1121 (vid. pág. 20).

- Das, Kausik y col. “Detection of breast cancer from whole slide histopathological images using deep multiple instance CNN”. En: *IEEE Access* 8 (2020), págs. 213502-213511 (vid. pág. 20).
- Dimitriou, Neofytos, Ognjen Arandjelović y Peter D Caie. “Deep learning for whole slide image analysis: An overview”. En: *Frontiers in Medicine* 6 (2019) (vid. págs. 18, 19).
- Elston, Christopher W y Ian O Ellis. “Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up”. En: *Histopathology* 19.5 (1991), págs. 403-410 (vid. págs. 15, 16).
- Ferreira, Carlos A y col. “Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet v2”. En: *International conference image analysis and recognition*. Springer. 2018, págs. 763-770 (vid. pág. 20).
- Ilse, Maximilian, Jakub Tomczak y Max Welling. “Attention-based deep multiple instance learning”. En: *International conference on machine learning*. PMLR. 2018, págs. 2127-2136 (vid. págs. 26, 31, 38).
- Khan, Adnan Mujahid, Korsuk Sirinukunwattana y Nasir Rajpoot. “A global covariance descriptor for nuclear atypia scoring in breast histopathology images”. En: *IEEE journal of biomedical and health informatics* 19.5 (2015), págs. 1637-1647 (vid. pág. 42).
- Komura D, Ishikawa S. “Machine Learning Methods for Histopathological Image Analysis”. En: 16.2 (2018), págs. 34-42 (vid. pág. 22).
- Kramer, Mark A. “Nonlinear principal component analysis using autoassociative neural networks”. En: *AIChE journal* 37.2 (1991), págs. 233-243 (vid. pág. 23).

- Kumar, Jayant, Jaishanker Pillai y David Doermann. “Document image classification and labeling using multiple instance learning”. En: *2011 International Conference on Document Analysis and Recognition*. IEEE. 2011, págs. 1059-1063 (vid. pág. 25).
- Lakhani, Sunil R. “The pathology of familial breast cancer: morphological aspects”. En: *Breast Cancer Research* 1.1 (1999), págs. 1-5 (vid. pág. 16).
- Litjens, Geert y col. “A survey on deep learning in medical image analysis”. En: *Medical image analysis* 42 (2017), págs. 60-88 (vid. págs. 21, 22).
- Lu, Cheng y col. “Automated image analysis of nuclear atypia in high-power field histopathological image”. En: *journal of microscopy* 258.3 (2015), págs. 233-240 (vid. pág. 42).
- Lu, Cheng y col. “Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers”. En: *Laboratory Investigation* 98.11 (2018), págs. 1438-1448 (vid. pág. 19).
- Maqlin, P y col. “Automated nuclear pleomorphism scoring in breast cancer histopathology images using deep neural networks”. En: *International Conference on Mining Intelligence and Knowledge Exploration*. Springer. 2015, págs. 269-276 (vid. pág. 19).
- Meyer, John S y col. “Breast carcinoma malignancy grading by Bloom–Richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index”. En: *Modern pathology* 18.8 (2005), págs. 1067-1078 (vid. págs. 12, 15).
- MITOS-ATYPIA-14 - Dataset* (vid. pág. 32).
- Moncayo Martinez, Ricardo Alexander y col. “Quantitative estimation of nuclear pleomorphism in breast cancer histological images”. Tesis doct. Universidad Nacional de Colombia-Sede Bogotá (vid. pág. 17).

Organization, World Health. *Colombia Globocan*. <https://gco.iarc.fr/today/data/factsheets/populations/170-colombia-fact-sheets.pdf>. [Online; accessed 03-July-2020]. 2018 (vid. pág. 11).

— *Colombia Globocan*. <https://www.cancer.gov.co/files/libros/archivos/95685f345e64aa9f0feceBOLETIN20HECHOS20Y20ACCIONES20MAMA.PDF>. [Online; accessed 10-August-2020]. 2019 (vid. pág. 12).

Plaut, Elad. “From principal subspaces to principal components with linear autoencoders”. En: *arXiv preprint arXiv:1804.10253* (2018) (vid. pág. 24).

Rakha, Emad A y col. “Breast cancer prognostic classification in the molecular era: the role of histological grade”. En: *Breast Cancer Research* 12.4 (2010), págs. 1-12 (vid. pág. 17).

Ranzato, Marc’Aurelio y col. “Unsupervised learning of invariant feature hierarchies with applications to object recognition”. En: *2007 IEEE conference on computer vision and pattern recognition*. IEEE. 2007, págs. 1-8 (vid. pág. 24).

Romo-Bucheli, David y col. “Automated tubule nuclei quantification and correlation with oncotype DX risk categories in ER+ breast cancer whole slide images”. En: *Scientific reports* 6.1 (2016), págs. 1-9 (vid. pág. 18).

Ross, Michael H y Pawlina Wojciech. *Histología*. 2013 (vid. pág. 18).

Society, American Cancer. *Cancer de Seno*. <https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html>. [Online; accessed 15-August-2020]. 2018 (vid. pág. 11).

Sommer, Christoph y col. “Learning-based mitotic cell detection in histopathological images”. En: ene. de 2012, págs. 2306-2309 (vid. pág. 17).

- Srinidhi, Chetan L, Ozan Ciga y Anne L Martel. “Deep neural network models for computational histopathology: A survey”. En: *arXiv preprint arXiv:1912.12378* (2019) (vid. pág. 14).
- Sudharshan, PJ y col. “Multiple instance learning for histopathological breast cancer image classification”. En: *Expert Systems with Applications* 117 (2019), págs. 103-111 (vid. pág. 12).
- Veta, Mitko y col. “Breast cancer histopathology image analysis: A review”. En: *IEEE Transactions on Biomedical Engineering* 61.5 (2014), págs. 1400-1411 (vid. pág. 13).
- Xu, Jun y col. “Deep learning for histopathological image analysis: Towards computerized diagnosis on cancers”. En: *Deep Learning and Convolutional Neural Networks for Medical Image Computing*. Springer, 2017, págs. 73-95 (vid. pág. 42).