

Categorización de anomalías cancerígenas en mastografías digitales aplicando aprendizaje profundo

José Aurelio Carrera Melchor¹, Eddy Sánchez-DelaCruz¹, Rajesh Roshan Biswal¹,
María Victoria Carreras Cruz²

¹ Instituto Tecnológico Superior de Misantla, Veracruz,
México

² Universidad Panamericana, Ciudad de México,
México

{162t0076, esanchezd, rroshanb}@itsm.edu.mx, mvcruz@up.edu.mx

Resumen. El cáncer es una enfermedad considerada grave desde hace siglos y a nivel global es uno de los padecimientos de mayor incidencia el cual se ha reforzado a lo largo de los últimos años, el cáncer de mama es el tipo de cáncer más frecuente en las mujeres de México y la segunda causa de muerte por cáncer a nivel mundial. Esta tasa de mortandad se ha reducido gracias a diversas técnicas de detección temprana, principalmente mastografías, sumado a un análisis correcto. Actualmente, las mastografías digitales pueden ser asistidas por computadora y esta investigación toma como referencia la aplicación del preprocesamiento de imágenes y diversos algoritmos ensamblados en conjunto con Aprendizaje Profundo para mejorar la eficiencia de la detección. A través de datasets generados y aplicando los algoritmos *LogitBoost* y *AttributeSelectedClassifier* en conjunto con Aprendizaje profundo, se analiza el histograma de las imágenes pertenecientes al Dataset de dominio público MIAS, obteniendo resultados competitivos de 88.37%.

Palabras clave: cáncer de mama, microcalcificación, clasificación, aprendizaje profundo.

Categorization of Carcinogenic Abnormalities in Digital Mastography using Deep Learning Algorithms

Abstract. Cancer has been considered a serious disease for centuries and globally is one of the most prevalent conditions, which has been reinforced in recent years; breast cancer is the most common type of cancer in women in Mexico and the second leading cause of cancer death worldwide. This mortality rate has been reduced thanks to various early detection techniques, mainly mastography and correct analysis. Currently, digital mastography can be computer assisted and this research takes as a reference the application of image preprocessing and various assembled algorithms in conjunction with Deep learning to improve the efficiency of detection. Through datasets generated and applying *LogitBoost* and

AttributeSelectedClassifier algorithms in conjunction with Deep Learning, it analyzes the histogram of the images belonging to MIAS Dataset, obtaining competitive results of 88.37%.

Keywords: breast cancer, micro-calcification, classification, deep learning.

1. Introducción

El cáncer es un proceso de crecimiento y diseminación células de manera incontrolada, una célula normal se divide de una célula madre, ésta, a su vez, se divide una vez más; si durante el proceso alguna se daña o envejece es remplazada y el ciclo se reinicia. Sin embargo, en las células cancerígenas crecen anormalmente y sobreviven para, también, volver a dividirse. Debido a que el cuerpo humano se compone de millones de células, este proceso puede aparecer en cualquier, incluso el tumor suele invadir el tejido circundante y puede provocar metástasis en puntos distantes del organismo.

El cáncer de mama es la forma más frecuente de cáncer entre las mujeres y también se considera asociado con la tasa de mortalidad más alta. Se ha atribuido a la mastografía la opción más viable para su detección temprana debido a la relación costo beneficio que ésta ofrece. La detección del cáncer es su etapa primaria conduce a un tratamiento efectivo en los pacientes, sin embargo, en México, existen pocos especialistas en el área de identificación de posibles micro-calcificaciones malignas en mastografías digitales, lo que conduce a predicciones inexactas de esta anomalía específica.

Las micro-calcificaciones en los senos (ver Fig. 1.) son hallazgos frecuentes en la mastografía digital, la mayoría de ellas originadas por patologías benignas, las cuales pueden llegar a ser malignas, especialmente en carcinomas, cuya detección es difícil debido a su pequeño tamaño y a la falta de pericia para tomar y considerar lecturas precisas de la forma, textura, tamaño y ubicación. Sin embargo, la tasa de detección puede mejorar con programas asistidos por computadora que implementen algoritmos de clasificación de Aprendizaje automático.

2. Trabajos relacionados

Los siguientes trabajos muestran el desarrollo reciente en el campo médico y social del cáncer de mama, haciendo hincapié en la aplicación de técnicas de inteligencia artificial para proporcionar soluciones como: predicción y clasificación. Estas investigaciones se basaron en técnicas de inteligencia artificial para aportar soluciones a los problemas mencionados.

Moradkhani et al, [6] basado en la extracción de imágenes del MIAS, cortó y removió la información adicional para luego usar un filtro en la imagen, obteniendo datos para después ser clasificados y obtener un método que brinda el 97% de correcta clasificación.

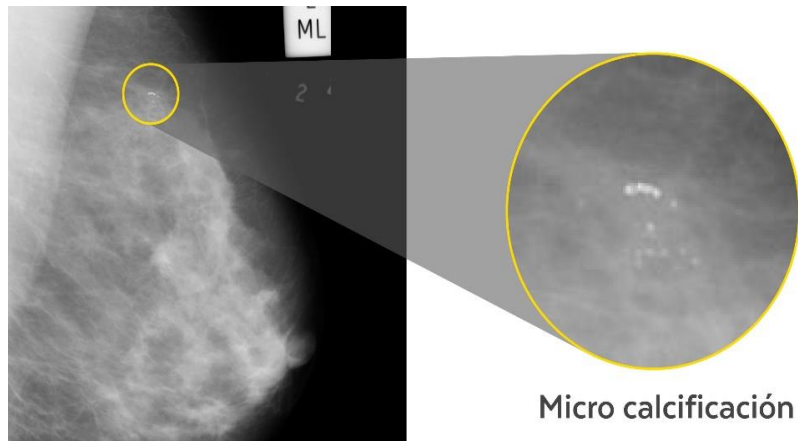


Fig. 1. Micro-calcificación presente en una Mastografía Digital.

Arafi et al, [7] implementaron un método para la detección de cáncer basado en Support Vector Machine, como técnica de aprendizaje supervisado para clasificar datos empíricos. Así optimizaron el rendimiento del clasificador resultante y obtuvieron un 94.74%.

Carreras et al, [3] para atender la problemática de clasificar anomalías cancerígenas, utilizaron el dataset MIAS, en el cual implementaron un algoritmo de agrupamiento parcial k-means y como resultado un único falso positivo, de la imagen mdb026. El resultado fue 95% de confianza en la clasificación de tipos de cáncer en imágenes mastográficas.

Neto et al, [8] para automatizar la segmentación de masas en mastografías, se utilizó la optimización de enjambre de partículas (PSO) y graph clusters, logrando un 95.2% de efectividad.

Arevalo et al, [9] utilizó un enfoque híbrido donde las redes neuronales convolucionales se utilizan para aprender la representación de forma supervisada, obteniendo un porcentaje de 82%.

Lévy et al, [10] implementaron un modelo integral de aprendizaje profundo para clasificar las masas mamarias pre-detectadas de mastografías, se utilizó la arquitectura AlexNet y GoogLeNet, obteniendo con la última el mayor porcentaje exactitud de 92.9%.

Gerazov et al, [11] aplicaron métodos de aprendizaje profundo a un conjunto de datos de dominio de tiempo en mama homogénea del tejido adiposo. Emplearon redes neuronales convolucionales, así como el clasificador de entrada Support Vector Machine para una precisión de 93.44%.

Al-Masni et al, [12] utilizaron el sistema de diseño asistido por computadora (por sus siglas CAD en inglés), para la detección de masas de seno y clasificación de cáncer, implementaron una Red Neuronal Convolucional logrando una eficacia de 93.20% al clasificar imágenes benignas, mientras que las malignas un 78% de efectividad, y su porcentaje global de 85.52% para clasificar anomalías.

Cruz et al, [5] el enfoque utilizado fue evaluar la exactitud y robustez de un método basado en el aprendizaje profundo para identificar automáticamente la extensión del tumor invasivo en las imágenes digitalizadas, este arrojó un de 75.86% de instancias detectadas y un valor predictivo positivo de 71.62 %.

Camacho et al, [4] implementaron el método heurístico basado en minería de datos para extraer información esencial de las imágenes mamográficas y transformarlas en patrones.

Pedraza et al., entrenaron una red neuronal convolucional basada en la arquitectura GoogLeNet, para desarrollar el modelo, después se llevó a cabo un proceso de validación cruzada. Así, el algoritmo proporciona una precisión del 95.62% para un conjunto de 5750 instancias [13].

Dalmi et al, [14] utilizaron el algoritmo ensamblado Random forests, combinándolo con una CNN, para la clasificación de lesiones, obteniendo el 85% de efectividad al distinguir los distintos tipos de anomalías.

3. Motivación y problemática

Los puntos importantes que motivaron esta investigación fueron: 1) Según la Organización Mundial de la Salud (OMS) [19] en México, la población femenina tiene una alta tasa de cáncer de mama ocupando el segundo lugar a nivel mundial después del cáncer de pulmón, y el primer lugar de mortalidad por cáncer en el país. 2) La incertidumbre de una correcta interpretación de las mastografías digitales para un diagnóstico temprano eficiente aún se ofrece como nicho de oportunidad y 3) la escasez de radiólogos certificados en México para poder interpretar la mastografía con microcalcificaciones malignas.

Como se ha venido mencionando, a nivel mundial, el cáncer es una de las enfermedades más graves que se ha extendido en los últimos años, afectando gravemente a la población, según la OMS, ocupa el segundo lugar en causa de muerte, provocando 8,8 millones de defunciones en 2015 [19]. En México, desde 2006 el cáncer de mama es el más frecuente en el segmento de la población femenina, para ser precisos en este año se superó el cáncer de cuello uterino. Según cifras del Instituto Nacional de Estadística y Geografía (INEGI) [20], hasta el año 2012 se reportaron 26.64 casos por cada 100,000 mujeres mayores de 20 años, siendo la segunda causa de muerte por cáncer en el mismo grupo de edad, con 15.4%. Aunado a ello, de acuerdo a la Organización Panamericana de la Salud (OPS), se espera que el número de personas con la enfermedad aumente en un 46% en los próximos años.

La mastografía es una herramienta utilizada para la detección oportuna del cáncer de mama, es el único estudio que ha demostrado que su uso reduce la mortalidad por cáncer de mama hasta en un 30% [1]. Pero, es indispensable que las evaluaciones de las mastografías sean evaluadas por expertos certificados para interpretar estos estudios, los cuales, de acuerdo con [18], existen poco más de 40 en México.

Por lo anterior, se deduce que la detección temprana por mastografía puede reducir la mortalidad de esta enfermedad.

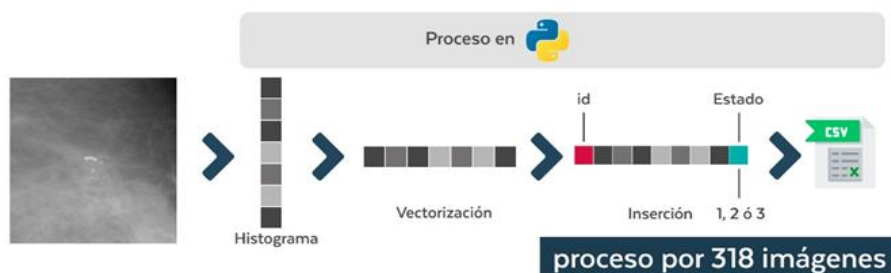


Fig. 3. Proceso de obtención del Dataset.



Fig. 2. Metodología basada en Moradkhani et al. [6].

4. Propuesta de solución

El objetivo de esta investigación es identificar microcalcificaciones que pueden existir dentro de la mastografía digital y su probable estado (Normal, Benigno o Maligno) utilizando algoritmos ensamblados para minimizar el margen de error, ya que se pueden conseguir tasas de detección competitivas y altamente reproducibles que faciliten la detección oportuna del cáncer de mama. En la sección 5 se analiza con más detalle la metodología que debe seguirse para la solución del problema.

5. Metodología

Para el desarrollo del presente trabajo se utilizan dos metodologías para llegar al análisis final de los resultados, el primero se describe como el proceso general para llegar a la obtención de los datos de interés derivados de las imágenes pertenecientes al dataset MIAS, y el segundo para poder realizar la evaluación de efectividad de acuerdo a parámetros establecidos que corresponden a las pruebas iniciales.

5.1. Descripción de la metodología del proceso general

Un modelo clasificador está asociado con el reconocimiento de patrones. Una breve descripción de la metodología se da a continuación (ver Fig. 2.):

- Database Original. Consiste en el uso de un database con 322 imágenes, con dimensiones de 1024x1024 píxeles, de las cuales 207 son mastografías normales, 60 tienen microcalcificaciones detectadas como benignas y 51 malignas que presentan una concentración atípica de células en el seno.

- Zoom 250 px. Se realiza un recorte de las imágenes seleccionadas con una altura y anchura de 250 píxeles, basado en la metodología de [2] ello debido a que la imagen 226 y 239 cuentan con más de un grupo microcalcificaciones y el tamaño del recorte abarca directamente las presentes.
- Giro de 180 grados Non. Las imágenes de numeración impar se realiza el procedimiento para que tengan la misma dirección que las imágenes pares, esto para tener una serie de imágenes con características similares, aunado a ello se descartan las imágenes 133, 134, 151 y 152 por el exceso de dimensiones, esto para tener una serie de elementos con características similares, ello con el fin de analizar y comparar con otros sistemas de recorte que puedan existir en el futuro donde se maneje un pixelaje mayor.
- Obtención del dataset. A través de un algoritmo desarrollado en el lenguaje Python, las imágenes se someten a una obtención de su histograma y posterior vectorización del mismo, dicho algoritmo agrega un identificador (id) al inicio del vector recorriendo la información del histograma, asimismo al finalizar los datos obtenidos de la imagen se agrega al final del vector tres identificadores los cuales son: 1 (normal), 2 (benigno) y 3 (maligno), esto de acuerdo a los datos brindados por la página oficial del dataset MIAS. Es decir 318 archivos de extensión .csv los cuales son tratados posteriormente.
- Selección de características. En la obtención del dataset se integra un id, 250 elementos propios del histograma, más el estado Normal, Benigno o Maligno, es decir 252 características fungiendo como principal la última.
- Creación de dataset multiclase. A partir de los 318 archivos generados en la obtención del dataset estos a través de un enclaustramiento en una carpeta y utilizando funciones nativas de Python se compilan en un solo dataset, el cual contiene 318 líneas, sin embargo, para poder utilizarlo en posterior análisis se requiere insertar las cabeceras correspondientes de texto asimismo como la conversión de los últimos valores que contiene los estados, es decir; 1 (sustituir por n), 2 (sustituir por b) y 3 (sustituir por m).
- Creación de datasets binarios. Tomando en cuenta que se tiene tres tipos de clases estos se pondrán a prueba colocando versus entre ellos, para ello se divide el dataset multiclase en subdatasets con combinaciones de; Maligno-Benigno el cual contiene 111 registros de histograma vectorizadas, Normal-Benigno, el cual contendrá 267 líneas y Normal-Maligno contando con 257 líneas.

Selección de modelo y entrenamiento. Para ello se utilizan los criterios de muestreo: 1/3-2/3, Validación cruzada con 10 iteraciones y muestra representativa. Los algoritmos mejor posicionados, es decir, los que mejores resultados generaron, fueron *LogitBoost+Dlj4Mlp*, *AttributeSelectedClassifier+Dlj4Mlp*, *FilteredClassifier+Dlj4Mlp* y *Staking+Dlj4Mlp* los cuales se describen a continuación:

LogitBoost [17]. También conocido como regresión logística aditiva, optimiza la probabilidad directamente. Desde un punto de vista práctico, LogitBoost utiliza un

esquema de regresión base. Este algoritmo puede ser visto como una optimización convexa, específicamente, dado que se busca un modelo aditivo de la forma $f = \sum_t a_t h_t$, donde el algoritmo LogitBoost minimiza las pérdidas logísticas mediante $\sum_i \log(1 + e^{-y_i f(x_i)})$.

AttributeSelectedClassifier [15]. Este algoritmo utiliza el ranking con InfoGainAttributeEval y la búsqueda de Ranker y puede eliminar atributos menos útiles. Este algoritmo debe ser usado para transformar los datos antes de pasarlos a su proceso.

FilteredClassifier [15]. Esta es una clase que ejecuta un clasificador arbitrario en datos que han pasado por un filtro. Al igual que un clasificador, la estructura del filtro se basa exclusivamente en los datos de formación y las instancias de prueba pueden ser procesadas por el filtro sin cambiar su estructura. Si existen pesos de instancia o pesos de atributo desiguales y el filtro o el clasificador no son capaces de tratarlos, las instancias y/o atributos se vuelven a muestrear con un reemplazo basado en los pesos antes de pasarlos al filtro o al clasificador (según corresponda).

Stacking [16]. Algoritmo donde existe un conjunto de n miembros. Cada uno de estos miembros está entrenado en un conjunto dado de datos de entrenamiento. Los miembros de este conjunto pueden compartir el mismo tipo de clasificador (homogéneo) o utilizar diferentes clasificadores (heterogéneos). La diversidad de datos fomenta entre los miembros para que cada miembro genere diferentes estimaciones.

El algoritmo permite configurar las siguientes capas para construir arquitecturas más sofisticadas: Capa de submuestreo, la cual subdivide grupos de unidades de la capa madre por diferentes estrategias (media, máxima, etc.); BatchNormalization, que aplica la estrategia común de normalización de lotes en las activaciones de la capa madre; OutputLayer, la cual genera salidas de clasificación / regresión, entre algunas otras para mejorar el aprendizaje. Evaluación. Habiendo ejecutado uno a uno los algoritmos de clasificación en conjunto con el algoritmo DLj4Mlp, se procedió a validar los mejores resultados mediante las siguientes métricas:

- i), matriz de confusión,
- ii) sensibilidad, que es la capacidad de prever los casos positivos cuando realmente son enfermos o con presencia de microcalcificaciones, es decir la capacidad para detectar enfermedad en mastografías con signos de micro-calcificaciones,
- iii) especificidad, la cual brindará información de casos negativos de los que son realmente sanos y la proporción de sanos correctamente identificados, es decir la capacidad de detectar la enfermedad en mastografías de características sanas.

6. Experimentos y análisis de resultados

Los experimentos fueron realizados en una computadora con las siguientes características: Windows 10 Home Single Language, Intel(R) Core i7-6500U CPU 2.50 GHz, Ram 8.00 GB, HDD Estado Sólido de 480Gb, Sistema operativo de 64bits, procesador x64, El tratamiento de imágenes fue realizado con el software XnView, los

Tabla 1. Algoritmos aplicados al dataset multiclase.

Ensamblados(Meta)	DeepLearnig	2/3 – 1/3	CV-10	MR(45)
	Dlj4Mlp	47.1698	54.0881	48.5714
LogitBoost	Dlj4Mlp	65.1509	65.0945	63.4286
Stacking	Dlj4Mlp	65.1509	65.0945	63.4286
FilteredClassifier	Dlj4Mlp	17.9245	51.2579	65.1429

Tabla 2. Algoritmos aplicados al dataset binario Normal-Maligno.

Ensamblados(Meta)	DeepLearnig	2/3 – 1/3	CV-10	MR(45)
	Dlj4Mlp	66.2791	71.5963	60
AttributeSelectedClassifier	Dlj4Mlp	88.3721	80.9339	80
LogitBoost	Dlj4Mlp	82.5581	80.5447	77.4194
FilteredClassifier	Dlj4Mlp	86.9465	75.4864	32.2581

algoritmos fueron programados en Spyder (Python 3.6) y la clasificación en Weka 3.8.2.

Para realizar las pruebas se aplicaron algoritmos ensamblados a los datasets generados, es decir Multiclase, Binario Normal-Maligno, Binario Normal-Benigno, y Binario Benigno-Maligno, tomando los criterios de 1/3 2/3, validación cruzada de 10 iteraciones y Muestra representativa.

6.1. Dataset 1: multiclase

Para esta prueba se tomó un dataset que contiene 318 elementos y tres diferentes clases, para lo cual observamos que, de acuerdo a la Tabla 1. Los mejores algoritmos son *LogitBoost+Dlj4Mlp* y *Staking+Dlj4Mlp* en los criterios 2/3 1/3 y validación cruzada de 10 iteraciones con porcentajes iguales de 65.1509% y 65.0945% respectivamente, mientras que para el criterio de muestra representativa existe un algoritmo con una tasa de efectividad del 65.1429% a pesar de no ser beneficiado en los primeros dos criterios.

6.2. Dataset 2: binario normal-maligno

Para esta prueba el dataset cuenta con 257 elementos, con dos clases Normal y Maligno. El mejor algoritmo es *AttributeSelectedClassifier+Dlj4Mlp* ofreciendo un resultado del 88.3721% de efectividad ver Tabla 2. Sin embargo, existe la presencia de los algoritmos *LogitBoost+Dlj4Mlp* y *FilteredClassifier+Dlj4Mlp* que figuran también en la Tabla 1. Ello nos puede dar una referencia para futuras pruebas con estos datasets aplicando algún preprocesamiento distinto.

Tabla 3. Algoritmos aplicados al dataset binario Normal-Benigno.

Ensamblados(Meta)	DeepLearnig	2/3 – 1/3	CV-10	MR(45)
	Dlj4Mlp	69.6629	75.2809	62.6582
AttributeSelectedClassifier	Dlj4Mlp	78.6517	79.7753	81.0127
Staking	Dlj4Mlp	78.6517	77.5281	75.9494
FilteredClassifier	Dlj4Mlp	78.6517	74.5318	75.9494

Tabla 4. Algoritmos aplicados al dataset binario Benigno-Maligno.

Ensamblados(Meta)	DeepLearnig	2/3 – 1/3	CV-10	MR(45)
	Dlj4Mlp	69.6629	75.2809	62.6582
LogitBoost	Dlj4Mlp	51.3514	54.5455	54.6512
AttributeSelectedClassifier	Dlj4Mlp	48.6486	51.8182	55.8114
FilteredClassifier	Dlj4Mlp	51.3514	50.9091	54.6512

6.3. Dataset 3: binario normal-benigno

En esta prueba el dataset cuenta 267 elementos, con las clases de Normal y Benigno, de acuerdo a la Tabla 3. , el algoritmo más competitivo es *AttributeSelectedClassifier+Dlj4Mlp* con un rendimiento de 78.6517%, y haciendo una retrospectiva a la Tabla 2., podemos definir que el algoritmo señalado es bueno para identificar y realizar una segmentación efectiva de una mastografía normal a una que puede presentar un grado de lesión o presencia de micro-calcificaciones.

Es importante señalar que los algoritmos *Staking+Dlj4Mlp* y *FilteredClassifier+Dlj4Mlp* en el criterio 2/3-1/3 muestran una efectividad competitiva igualando el resultado del primer algoritmo.

6.4. Dataset 4: binario benigno-maligno

En el análisis del dataset y contando con 110 líneas resultantes y dos estados posibles Maligno y Benigno, tenemos la Tabla 4 en la cual se aprecia nuevamente al algoritmo LogitBoost en primer lugar compartiendo efectividad con el algoritmo *FilteredClassifier* en el criterio 2/3-1/3, que en retrospectiva a los análisis anteriores podemos declarar que para datos que pueden presentar confusión o multiclase es mejor aplicar el algoritmo *LogitBoost+Dlj4Mlp*.

Comparando estos resultados con los trabajos previos, los clasificadores propuestos en el presente trabajo superan los resultados obtenidos en Arevalo et al. [9] con 71.62 % de efectividad y a Dalmi et al. [14] con una tasa de efectividad de 85.52%. Sin embargo, Arafí et al. [7] obtuvo un 94.74%, Carreras et al. [3] obtuvo 95%. Neto et al. [8] y Pedraza et al. [13] obtuvieron un porcentaje de 95.2 y 95.62% respectivamente.

	a = N	b = M	
Normal	68	3	71
Maligno	8	7	15
Total			86

Sensibilidad	0.957746479
Especificidad	0.466666667
Falso Negativo	0.042253521
Falso Positivo	0.533333333

Fig. 4. Matriz de confusión y valores de sensibilidad, especificidad, falso negativo y falso positivo del dataset binario *Normal-Maligno*.

	a = B	b = N	
Normal	68	2	70
Benigno	2	17	19
Total			89

Sensibilidad	0.971428571
Especificidad	0.894736842
Falso Negativo	0.028571429
Falso Positivo	0.105263158

Fig. 5. Matriz de confusión y valores de sensibilidad, especificidad, falso negativo y falso positivo del dataset binario *Normal-Benigno*.

Por último, cabe mencionar a Moradkhani et al. [6] que obtuvieron una efectividad del 97%.

Finalmente, se adjuntan las matrices de confusión de los experimentos con mejores resultados, es decir Binario Normal Maligno (Fig. 4) y Normal Benigno (Fig. 5) tomando el criterio 2/3-1/3 que ofreció el porcentaje más alto asimismo como la sensibilidad y especificidad en conjunto con los valores falso positivos y falso negativos.

7. Conclusión y trabajo futuro

Después de una búsqueda exhaustiva y con una tasa de clasificación del 88.37% correcto en un dataset binario, se determinaron los siguientes esfuerzos para ampliar este estudio en una segunda etapa:

- El algoritmo *LogitBoost+Dlj4Mlp* en general es bueno para clasificar datasets multiclase o que puedan presentar datos confusos en el análisis de nuestros datos derivados del histograma.
- El algoritmo *AttributeSelectedClassifier+Dlj4Mlp* demostró ser bueno en datasets cuyas características puedan separarse de manera sustancial, es decir, de acuerdo a los resultados obtenidos se pueda utilizar para definir si una mastografía presenta características normales o alguna microcalcificación.
- El algoritmo *FilteredClassifier+Dlj4Mlp* y *Staking+Dlj4Mlp* muestra tasas de efectividad buenas en algunos criterios de muestreo, los cuales se pueden utilizar como refuerzos para definir el diagnóstico final.

Cabe destacar que para mejorar los resultados obtenidos se debe realizar un proceso de umbralización o segmentación de las imágenes, ya que en este trabajo se analizaron datos brutos e imágenes sin procesamiento previo.

Se considera, además, aplicar la metodología general a las imágenes de mastografías digitales resultado de la patente con registro MX/a/2008/038357 denominada “Procesamiento e interpretación automatizada de imágenes apoyada en la segmentación y equipo para llevar a cabo este procedimiento” perteneciente a Centros Culturales de México A.C., propietaria de la Universidad Panamericana a través de la firma de convenio con el Instituto Tecnológico Superior de Misantla. También se requiere, aplicar los clasificadores *AttributeSelectedClassifier+Dlj4Mlp* y *LogitBost+Dlj4Mlp*, para su análisis y mejoramiento de la tasa obtenida con datos brutos e incluir a los algoritmos *FilteredClassifier+Dlj4Mlp* y *Staking+Dlj4Mlp* para ayudar a definir el estado final de la mastografía analizada.

Derivado de lo anterior, se propone con los anteriores ejercicios alcanzar un mínimo de 95% de tasa precisión para poder establecer un vínculo más competitivo con los trabajos relacionados en esta área.

Referencias

1. Cunha, P., Nunes, M., Patrocinio, A.: Breast density pattern characterization by histogram features and texture descriptors. *Research on Biomedical Engineering* 33(1), pp. 69–77 (2017)
2. Mellado, M., Osab, M., Murillo, A.: Influencia de la mamografía digital en la detección y manejo de microcalcificaciones. *Radiología: Publicación oficial de la Sociedad Española de Radiología Médica* 55(2), pp.142–147 (2013)
3. Carreras, M., Martínez, M., Rosas, K.: Mass segmentation in digital mammograms. *Ambient Intelligence for Health* 9456(1), pp. 110–115 (2015)
4. Camacho, S.: Método Heurístico para el Diagnóstico de Cáncer de Mama basado en Minería de Datos. *Revista PGI - Investigación, Científica y Tecnología* 1, pp. 97–101 (2014)
5. Cruz, A., Gilmore, H., Basavanthally, A.: Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent. *Scientific Reports*, pp. 97–101(2017)
6. Moradkhani, F., Sadeghi, B.: A New Image Mining Approach for Detecting Micro-Calcification in Digital Mammograms. *Applied Artificial Intelligence* 31(5), pp. 411–424 (2017)
7. Araf, A., Fajr, R., Bouroumi, A.: Breast cancer data analysis using support vector machines and particle swarm optimization. In: *Complex systems (WCCS), Second world conference*, pp. 1–6 (2014)
8. Neto, O., Carvalho, O., Sampaio, W.: Automatic segmentation of masses in digital mammograms using particle swarm optimization and graph clustering. In: *International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp.109–112 (2015)
9. Arevalo, J., González, F., Ramos, R. et al.: Representation learning for mammography mass lesion classification with convolutional neural networks. *Computer methods and programs in biomedicine* 127(1), pp. 248–257 (2016)
10. Lévy, D., González, F.: Breast mass classification from mammograms using deep convolutional neural networks. In: *CoRR* (2016)

11. Gerazov, B., Conceicao, R.: Deep learning for tumour classification in homogeneous breast tissue in medical microwave imaging. In: IEEE (EUROCON'17) - 17th International Conference on Smart Technologies, pp. 564–569 (2017)
12. Al-masni, M., Al-antari, M., Park, J. et al.: Detection and classification of the breast abnormalities in digital mammograms via regional Convolutional Neural Network. In: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1230–1233 (2017)
13. Pedraza, A., Serrano, I., Fernández, M., et al.: Diagnóstico Automático del HER2 con Deep Learning. Google Scholar (2016)
14. Dalmi, M., Gubern, A., Vreemann, S. et al.: A computer-aided diagnosis system for breast dce-mri at high spatiotemporal resolution. *Medical physics* 43(1), pp. 84–94 (2016)
15. Witten, I., Frank, E., Hall, M., et al.: *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann (2017)
16. Wolpert, D.: Stacked generalization. *Neural Networks Journal* 5, pp. 241–259 (1992)
17. Li, P.: ABC-LogitBoost for Multi-Class Classification. Department of Statistical Science, Cornell University (2012)
18. CONACYT: Desarrollan algoritmo para la detección precoz de cáncer de mama, newsnet.conacytprensa.mx/index.php/documentos/36532-desarrollan-algoritmo-para-la-deteccion-precoz-de-cancer-de-mama (2018)
19. WHO: Position paper on mammography screening, <http://www.who.int/cancer/publications> (2018)
20. INEGI: Estadísticas a propósito del Día mundial contra el cáncer, http://www.beta.inegi.org.mx/contenidos/saladeprensa/aproposito/2018/cancer2018_Nal (2018)