Clasificación de imágenes de plataformas multimedia con algoritmos de aprendizaje automático y redes neuronales artificiales.

Image classification from multimedia platforms with machine learning algorithms and artificial neural networks.

Esperanza Guerra Rosas* (1).
Universidad Autonóma de Baja California, Campus Ensenada.
esperanza.guerra@uabc.edu.mx.

Josué Álvarez Borrego (2). Centro de Investigación Cientifica y de Educación Superior de Ensenada,
josue@cicese.mx.

Mauricio Sauzameda González (3). Centro de Investigación Cientifica y de Educación Superior de Ensenada,

Lázaro Bárbaro Fonte Valdés (4). Centro de Investigación Cientifica y de Educación Superior de Ensenada,

*corresponding author.

Artículo recibido en septiembre 22, 2025; aceptado en octubre 01, 2025.

Resumen.

Los algoritmos de aprendizaje de máquina y redes neuronales artificiales permiten reconocer y clasificar imágenes por medio de patrones característicos de imágenes. Existen diferentes métodos de extracción de características, algunos se basan en la textura de imágenes. En este trabajo se presenta una metodología para clasificar imágenes de plataformas multimedia, se obtuvieron características de textura, como Patrón Binario Local (LBP), matrices de coocurrencia de niveles de grises (GLCM) y el método de Wavelet image scattering (WIS). Se utilizaron algoritmos de aprendizaje automático y redes neuronales artificiales. Se utilizó un conjunto de datos de imágenes de mosaicos y un conjunto de descriptores, cada uno con imágenes de diferentes clases. Se aplicaron métricas para evaluar el desempeño de los algoritmos de aprendizaje y las redes neuronales. Finalmente se compararon los resultados obtenidos para la clasificación tanto de mosaicos como de descriptores. Como resultado se obtuvieron valores significanticos, que pueden ayudar a usuarios de plataformas multimedia a seleccionar contenido y a compañías de multimedia a sugerir contenido visual de series o películas.

Palabras claves: Algoritmos de aprendizaje automático, descriptores, métodos de textura, mosaicos, plataformas multimedia, redes neuronales artificiales.

Abstract.

Machine learning algorithms and artificial neural networks allow image recognition and classification based on characteristic image patterns. There are various feature extraction methods, some of which are based on image texture. This work presents a methodology for classifying images from multimedia platforms. Texture features such as Local Binary Pattern (LBP), Gray Level Co-occurrence Matrices (GLCM), and the Wavelet Image Scattering (WIS) method were obtained. Machine learning algorithms and artificial neural networks were used. A dataset of mosaic images and a set of descriptors, each with images from different classes, were used. Metrics were applied to evaluate the

performance of the learning algorithms and neural networks. Finally, the results obtained for classifying both mosaics and descriptors were compared. The results obtained yielded significant values, which can help users of multimedia platforms select content and multimedia companies suggest visual content for series or movies.

Keywords: Artificial neural networks, descriptors, machine learning algorithms, mosaics, multimedia platforms, texture methods,

1. Descripción de la conferencia.

Desde la década de 1950 la Inteligencia Artificial (IA) ha presentado diversos avances. Actualmente la IA continúa avanzando con desarrollos novedosos que transforman la forma de vivir del ser humano y el entorno de trabajo. El mercado de plataformas multimedia ofrece diversas opciones como Netflix, Apple TV, Disney, HBO Max, entre muchas otras (Jenner, 2023). El crecimiento de ofertas de contenido en estas plataformas presenta desafíos a los usuarios en términos de navegación y descubrimiento de nuevos títulos de interés. El uso de sistemas de recomendación y clasificación facilita la visibilidad de contenido, selección y personalización del usuario. Por otra parte, también resulta de gran utilidad para usuarios empresariales para el análisis de datos, administración de publicidad, moderación de contenido, habito de consumo y preferencias de su audiencia (Jannach et al., 2021).

El reconocimiento y procesamiento de imágenes es un campo extenso que incluye técnicas de Inteligencia Artificial de forma automática para reconocimiento facial, detección de objetos, análisis de imágenes médicas, seguridad, control de calidad, agricultura, análisis satelital para el monitoreo de cultivos, automatización industrial, entre otras aplicaciones en distintas áreas.

Debido a la existencia de distintas plataformas multimedia es necesario utilizar herramientas para procesar y analizar grandes cantidades de datos que permitan identificar y detectar características visuales como objetos, personas o lugares con alta precisión, lo que mejora la interacción entre las plataformas y los usuarios (Badr et al., 2024).

Las plataformas de multimedia generan grandes volúmenes de imágenes, esta cantidad de información puede analizarse y etiquetarse en categorías o clases. Técnicas de Inteligencia Artificial y algoritmos de Redes Neuronales Artificiales procesan grandes cantidades de datos y la clasificación depende de las características de entrada. La extracción de características es una etapa fundamental en el procesamiento de datos para los modelos de IA, algunos métodos para extraer características representativas se basan en el análisis de texturas en las imágenes.

En este trabajo se presenta una metodología para analizar y clasificar imágenes de acuerdo al contenido visual de los usuarios de distintas plataformas multimedia, en tiempo real. El banco de imágenes contiene una base de datos de mosaicos y una de descriptores, cada una con imágenes de distintas plataformas multimedia que se obtuvieron del dispositivo Core Meter. Una imagen de tipo mosaico contiene secciones en miniatura de series y películas disponibles para las plataformas multimedia, que permiten al usuario ver las diferentes opciones que tiene para elegir una serie o película. Una imagen de tipo descriptor consiste en la descripción de la serie o película elegida de un mosaico.

Se analizaron tres técnicas para extraer vectores de características basados en la textura de imágenes. La técnica Patrones Binarios Locales (LBP) permite extraer información de la textura presente en las imágenes, se basa en los contornos, esquinas, puntos y regiones planas, es un operador simple y muy eficiente, es robusto a cambios de iluminación y rotación (Guerra-Rosas et al., 2023). La matriz de coocurrencia de escala de grises (GLCM) extrae parámetros de textura basados en ocurrencias de los valores de intensidad de pixeles en dirección horizontal y vertical, se calcula en cualquier ángulo y distancia (Aggarwal, 2022). El método de Wavelet image scatering (WIS) calcula descriptores basados en matrices de co-ocurrencia (Bruna et. al., 2013). Puede compararse con las convoluciones de una red neuronal profunda, se puede considerar como un filtro pasa banda basado en altas frecuencias o detalles de una imagen, también como un filtro de escalamiento pasa bajas que permite recuperar las bajas frecuencias. Después de obtener los vectores de características mediante los métodos extracción de textura, se utilizaron algoritmos basados en modelos de Inteligencia Artificial y Redes Neuronales Artificiales para su respectiva clasificación. Finalmente se evaluaron los modelos de clasificación y redes neuronales. Se comparó el desempeño de los modelos con los vectores

LBP, GLCM y WIS y se obtuvieron las métricas de exactitud, precisión, sensibilidad, especificidad y la medida F1. Se clasificaron imágenes de plataformas multimedia en tiempo real, de esta manera distintas compañías pueden contar con sistemas de clasificación que faciliten la navegación y recomendación de contenido para los usuarios.

2. Métodos.

En esta sección se muestran conceptos y técnicas llevadas a cabo para la clasificación de imágenes multimedia de mosaicos y descriptores.

2.1 Mosaicos.

La base de datos de mosaicos contiene 157,448 imágenes a color de 1280x720 pixeles, se tomaron 13 plataformas multimedia, además de una clase adicional para imágenes falsas que no corresponden a alguna categoría de plataformas. De cada clase de plataformas se tomaron 4,000 imágenes para entrenar los modelos con un total de 56,000, las imágenes restantes de cada clase se utilizaron para validar los modelos. En la **Fig. 1** se presenta un ejemplo de mosaicos para cada clase, incluyendo la clase falsa.

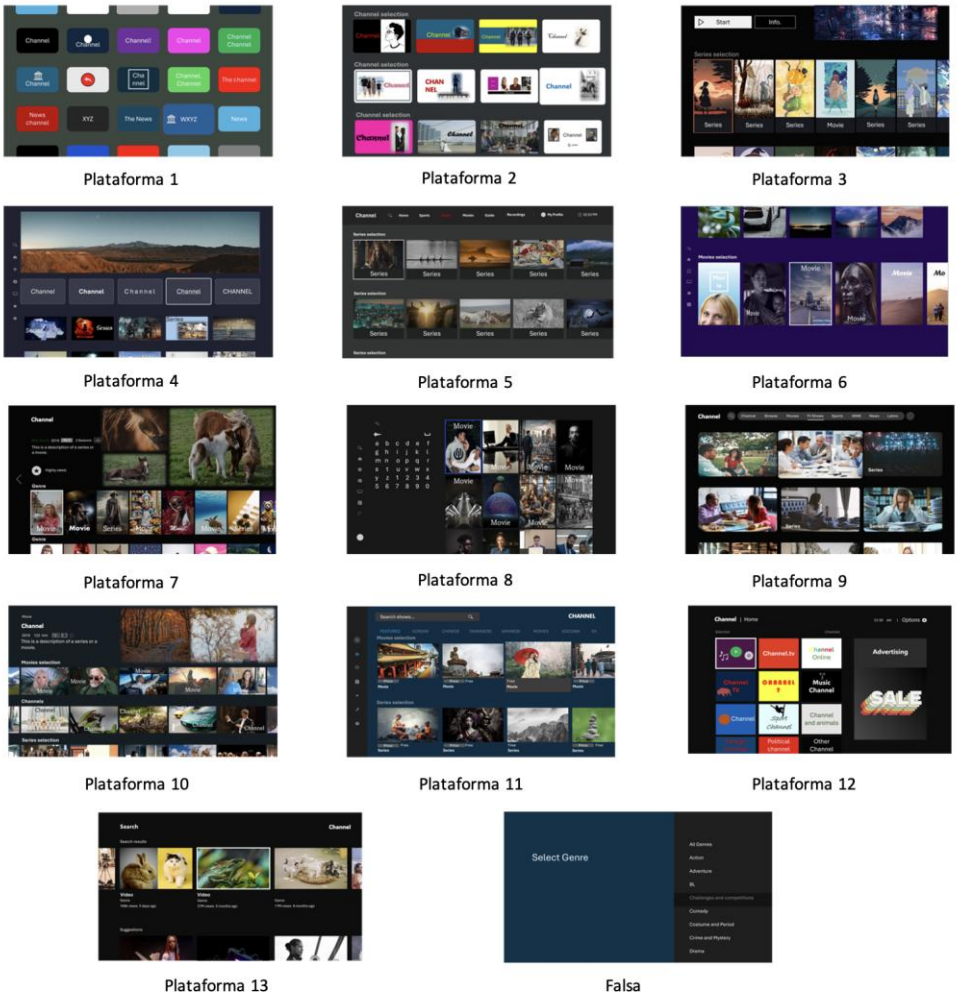

Figura 1. Muestra de imágenes de mosaicos de las 14 clases de plataformas utilizadas.

La distribución las clases para mosaicos se muestra en la **Tabla 1**.

Tabla 1. Distribución por clase de la base de datos de imágenes de mosaicos.

Clases	Entrenamiento	Validación	Total
Plataforma 1	4,000	7,064	11,064
Plataforma 2	4,000	6,540	10,540
Plataforma 3	4,000	7,300	11,300
Plataforma 4	4,000	6,478	10,478
Plataforma 5	4,000	1,994	5,994
Plataforma 6	4,000	6,002	10,002
Plataforma 7	4,000	7,516	11,516
Plataforma 8	4,000	6,524	10,524
Plataforma 9	4,000	6,000	10,000
Plataforma 10	4,000	24,884	28,884
Plataforma 11	4,000	6,000	10,000
Plataforma 12	4,000	5,962	9,962
Plataforma 13	4,000	7,906	11,906
Falso	4,000	1,278	5,278
Total	**56,000**	**101,448**	**157,448**

2.2 Descriptores.

La base de datos de descriptores contiene 38,214 imágenes a color de 1280x720 pixeles, se tomaron 10 plataformas multimedia, además de una clase adicional para imágenes falsas que no corresponden a alguna categoría de plataformas. En la **Fig. 2** se muestra un ejemplo de cada una de las clases, el área enmarcada en color rojo corresponde a una región que no presenta variaciones en cada clase. En estas imágenes de descriptores se tomó esa área como característica de cada clase para obtener los patrones de textura.
Para el entrenamiento del modelo se utilizaron 1,000 imágenes de cada clase, tomando en total 11,000 imágenes, las restantes de cada clase fueron utilizadas para validar los modelos.

2.3 Métodos de análisis de texturas.

La textura representa la distribución espacial del nivel de gris de los pixeles en una región de una imagen digital. El análisis de textura tiene un gran desempeño en muchas áreas, una de las funciones principales es obtener características para clasificar y segmentar. Se puede clasificar en métodos estadísticos, métodos estructurales y modelos basados en transformaciones (Armi et. al., 2019). En este trabajo se consideraron métodos estadísticos que se basan en la distribución de intensidad de los pixeles y métodos basados en transformaciones, donde la imagen es transformada a otro espacio mediante una función.

- ***Patrón binario local.***

La técnica de análisis estadístico y estructural de Patrón binario local (LBP), obtiene información de las texturas de la imagen. Este método compara cada pixel de una imagen con los pixeles vecinos, tomado una vecindad de 3x3 en toda la matriz de datos de la imagen, comenzando por la esquina superior izquierda y codifica la información de textura local en patrones binarios, de esta manera se obtiene el vector de características para la imagen y se utiliza directamente para la clasificación. Para este trabajo se utilizaron vectores uniformes de 59 elementos para cada imagen.

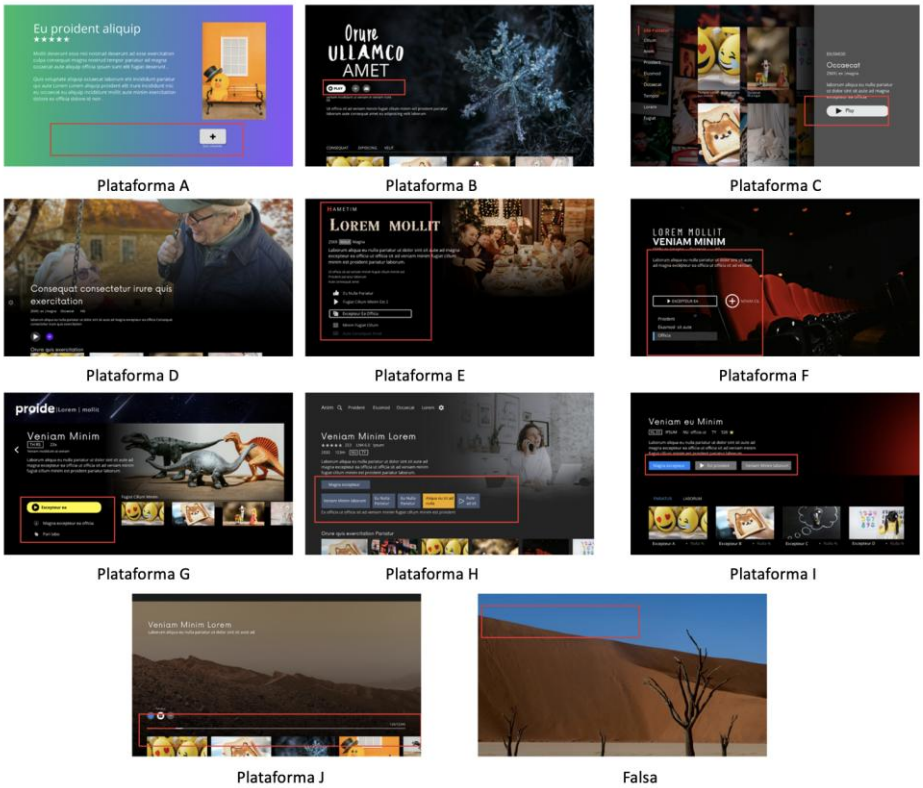

Figura 2. Muestra de imágenes de descriptores de las 11 clases de plataformas utilizadas.

La distribución de las imágenes se muestra en la **Tabla 2**.

Tabla 2. Distribución por clase de la base de datos de imágenes de descriptores.

Clases	Entrenamiento	Validación	Total
Plataforma A	1,000	307	1,307
Plataforma B	1,000	943	1,943
Plataforma C	1,000	633	1,633
Plataforma D	1,000	4,378	5,378
Plataforma E	1,000	4,657	5,657
Plataforma F	1,000	423	1,423
Plataforma G	1,000	4,060	5,060
Plataforma H	1,000	2,130	3,130
Plataforma I	1,000	939	1,939
Plataforma J	1,000	4,744	5,744
Falso	1,000	4,000	5,000
Total	**11,000**	**27,214**	**38,214**

- ***Matriz de coocurrencia de niveles de grises.***

Las características texturales de una matriz de coocurrencia (GLCM) se enfocan en el uso de estadísticas de segundo orden de los histogramas de imágenes en escala de grises. Este método considera parejas de pixeles vecinos que se encuentran separados por una distancia y un determinado ángulo. Estas matrices muestran propiedades de la distribución espacial de la escala de grises de la textura de una imagen. Las medidas de textura GLCM pueden calcularse a partir de la matriz de coocurrencia, en este trabajo se consideran ocho medidas: homogeneidad, contraste, disimilaridad, media, varianza, entropía, segundo momento angular y correlación.

- ***Wavelet image scatterinng.***

El método Wavelet image scattering (WIS) se basa en una transformación integral que analiza y extrae información de una señal o de una imagen, considera la distribución espacial y de frecuencias. (Grossmann & Morlet 1984). El proceso WIS puede considerarse como un filtrado sucesivo de altas frecuencias o detalles y con la función de escalado, que funciona como un filtro pasa bajas.

2.4 Métodos de clasificación.

Los algoritmos de clasificación de imágenes permiten analizar el contenido de información de imágenes o vectores de características. La tarea principal es clasificar de forma automática nuevas imágenes. Existen diferentes modelos de clasificación, en este trabajo se utilizaron modelos clásicos de aprendizaje supervisado como k-vecino más cercano, arboles de decisión, arboles de decisión, máquinas de soporte vectorial y Redes Neuronales Artificiales. Estos modelos pueden reconocer y clasificar imágenes de distintas plataforma de manera precisa (EAPA Batista et. al., 2009), (Du et al., 2002), (Song et. al., 2015), (Suthaharan, 2016), (Sastry, 2003), (Principe et al., 2000).

2.5 Métricas de desempeño.

Para conocer el desempeño de los métodos de clasificación, se pueden aplicar diferentes métricas de medición, en este caso se utilizaron las métricas de precisión, exactitud, sensibilidad, especificidad y la medida F1:

$$Exactitud = \frac{VP + VN}{VP + VN + FP + FN}$$

$$Precisión = \frac{VP}{VP + FP}$$

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

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

$$Medida\ F1 = \frac{2 * Sensibilidad * Precisión}{Sensibilidad + Precisión}$$

VP (Verdadero Positivo): Corresponde a las instancias en las que el modelo predijo correctamente la clase positiva o afirmativa.

FP (Falso Positivo): Se refiere a aquellos casos en los que el modelo predijo erróneamente la clase positiva, cuando en realidad la instancia pertenecía a la clase negativa.

FN (Falso Negativo): Representa las instancias en las que el modelo falló al predecir la clase positiva, clasificándolas incorrectamente como negativas.

VN (Verdadero Negativo): Engloba las instancias pertenecientes a la clase negativa que fueron correctamente predichas como tales por el modelo.

3. Desarrollo.

Se desarrolló un sistema de clasificación de imágenes en tiempo real, utilizando patrones de textura para identificar la visualización de plataformas multimedia por los usuarios. El Core Meter es un sistema que captura imágenes cada cinco segundos. Para cada imagen se extraen características de textura mediante los diferentes métodos. Los vectores que contienen la textura de imágenes se utilizaron para entrenar los siguientes algoritmos de clasificación k-vecino más cercano, arboles de decisión, máquinas de soporte vectorial y Redes Neuronales Artificiales. El resultado de la clasificación puede ayudar al proveedor para identificar las vistas de los usuarios de alguna plataforma multimedia. En **Fig. 3** se muestra un diagrama general del sistema para clasificar imágenes de plataformas multimedia para mosaicos y descriptores en tiempo real y generación de recomendaciones personalizadas.

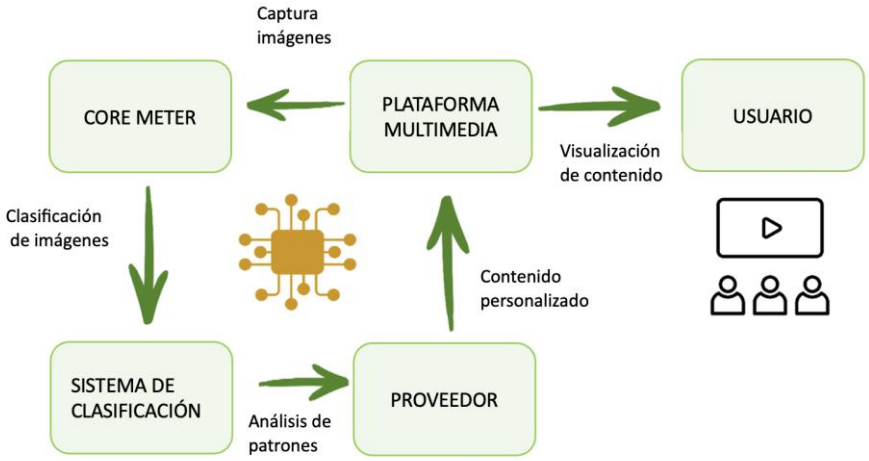

Figura 3. Diagrama general del sistema de clasificación de imágenes de plataformas en tiempo real.

3.1 Modelos clasificación.

Para la clasificación de mosaicos y descriptores se utilizaron los modelos indicados en la **Tabla 3**. Los algoritmos de aprendizaje de máquina solo utilizaron vectores LBP, para el caso de redes neuronales artificiales se utilizaron los vectores de textura LBP, GLCM y WIS.

4. Resultados.

Se obtuvieron resultados del desempeño de los 10 modelos de aprendizaje de máquina usando el vector caracteristico LBP y de redes neuronales artificiales usando los vectores de características LBP, GLCM y WIS. Las métricas empleadas incluyen la precisión, la especificidad, la sensibilidad, la exactitud y la medida F1, tanto para mosaicos como para descriptores.

Tabla 3. Vectores característicos de textura utilizados para los modelos de clasificación automatica y redes neuronales artificiales.

Modelos de clasificación	LBP	LBP	GLCM
SVM linear	•		
SVM cuadrático	•		
SVM cúbico	•		
SVM gaussiana fina	•		
SVM gaussiana media	•		
SVM gaussiana gruesa	•		
Ensemble boosted trees	•		
Ensemble bagged trees	•		
Ensemble subspace discriminant	•		
Ensemble subspace KNN	•		
Redes Neuronales Artificiales	•	•	•

A) Mosaicos.

En la **Tabla 4** se presenta la exactitud y tiempo promedio de entrenamiento de los modelos de aprendizaje máquina con vectores de entrada LBP que contienen 59 valores característicos, para clasificar plataformas multimedia de mosaicos. En esta estapa las opciones mas relevantes fueron las máquinas de soporte vectorial con kernel cúbico y cuadrático con tiempos óptimos de computo. El modelo de máquina de soporte vectorial con kernel cúbico obtuvo el mejor desempeño con precisión media de 98.81% especificidad de 99.94%, sensibilidad 99.25%%, exactitud de 99.89% y medida F1 de 99.02%.

Tabla 4. Exactitud y tiempo promedio de entrenamiento de los algoritmos de clasificación para mosaicos.

Modelos	Exactitud	Tiempo (segundos)
SVM linear	0.968	271.04
SVM cuadrático	0.989	456.09
SVM cúbico	0.989	425.25
SVM gaussiana fina	0.951	792.15
SVM gaussiana media	0.973	512.19
SVM gaussiana gruesa	0.918	724.60
Ensemble boosted trees	0.691	848.08
Ensemble bagged trees	0.960	591.24
Ensemble subspace discriminant	0.892	614.58
Ensemble subspace KNN	0.961	2740.80

En la **Tabla 5** se presenta la exactitud y tiempo promedio de entrenamiento de los modelos de redes neuronales artificiales con vectores de entrada LBP, GLCM y WIS, para clasificar plataformas multimedia de mosaicos. El mejor desempeño lo obtuvo la red neuronal artificial con el métodos de wavelet image scattering con exactitud de 99.97%, especificidad de 99.99%, sensibilidad de 99.84%, precisión de 99.59%, medida F1 de 99.71.

Tabla 5. Exactitud y tiempo promedio de entrenamiento de los algoritmos de clasificación para mosaicos.

Método de extracción de texturas	Características	Exactitud	Tiempo (segundos)
Patrón binario local uniforme	59	0.987	527.99
Matriz de coocurrencia de escala de grises	8	0.937	547.36
Wavelet image scattering	391	0.998	327.54

B) Descriptores.

En la **Tabla 6** se presenta la exactitud y tiempo promedio de entrenamiento de los modelos de aprendizaje de máquina con vectores de entrada LBP que contienen 59 valores característicos, para clasificar plataformas multimedia de descriptores. Los resultados más significativos se obtuvieron con máquinas de soporte vectorial con kernels cuadráticos y cúbicos obteniendo promedios superiores al 98 por ciento en todas sus métricas.

Tabla 6. Exactitud y tiempo promedio de entrenamiento de los algoritmos de clasificación para descriptores.

Modelos	Exactitud	Tiempo (segundos)
SVM linear	0.988	18.676
SVM cuadrático	0.991	18.358
SVM cúbico	0.993	18.482
SVM gaussiana fina	0.934	54.981
SVM gaussiana media	0.987	20.687
SVM gaussiana gruesa	0.927	33.285
Ensemble boosted trees	0.900	45.507
Ensemble bagged trees	0.988	11.309
Ensemble subspace discriminant	0.917	11.094
Ensemble subspace KNN	0.982	132.810

En la **Tabla 7** se presenta la exactitud y tiempo promedio de entrenamiento de los modelos de redes neuronales artificiales con vectores de entrada LBP, GLCM y WIS, para clasificar plataformas multimedia de descriptores. Para este caso el resultado sobresaliente desempeño sobresaliente se obtuevo con el método de wavelet image scattering con exactitud de 99.90%, especificidad de 99.94%, sensibilidad de 99.58%, precisión de 98.63%, medida F1 de 99.09%.

Tabla 7. Exactitud y tiempo promedio de entrenamiento de redes neuronales artificiales para descriptores utilizando los tres metodos de extraccíon de textura.

Método de extracción de texturas	Características	Exactitud	Tiempo (segundos)
Patrón binario local uniforme	59	0.989	58.43
Matriz de coocurrencia de escala de grises	8	0.979	97.97
Wavelet image scattering	391	0.999	372.42

Conclusiones.

La Inteligencia Artificial tiene un impacto significativo en la visualización de contenido multimedia para usuarios y empresas que se destacan en plataformas multimedia, mediante sistemas de clasificación de imágenes. Se obtuvo un desempeño con resultados favorables que permiten personalizar las recomendaciones de contenido para el usuario. La obtención de características mediante los métodos LBP, GLCM y WIS mostraron su capacidad para el entrenamiento de los modelos de clasificación automática y redes neuronales artificiales.

Referencias bibliográficas.

Aggarwal, A. K. (2022). Learning texture features from GLCM for classification of brain tumor MRI images using random forest classifier. Wseas transactions on signal processing

Armi, L., & Ershad, S. F. (2019). Texture image analysis and texture classification methods - a review. CoRR, abs/1904.06554. http://arxiv.org/abs/1904.06554

Badr, N., Sharaf, S., & Mahrous, A. A. (2024). Streaming wars: an analysis of the growth and competitive landscape of the subscription video-on-demand services market. EuroMed Journal of Management, 6(1), 23–41. https://doi.org/10.1504/EMJM.2024.135992

Bruna, J., & Mallat, S. G. (2013). Invariant scattering convolution networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1872–1886. https://doi.org/10.1109/TPAMI.2012.230

Du, W., Zhan, Z., & Du Zhijun Zhan, W. (2002). Building Decision Tree Classifier on Private Data. https://surface.syr.edu/eecshttps://surface.syr.edu/eecs/8

EAPA Batista, G., & Furtado Silva, D. (2009). How k-Nearest Neighbor Parameters Affect its Performance. http://www.icmc.usp.br/

Guerra-Rosas, E., López-Ávila, L. F., Garza-Flores, E., Vidales-Basurto, C. A., & Álvarez-Borrego, J. (2023). Classification of skin lesion images using artificial intelligence methodologies through radial Fourier–Mellin and Hilbert transform signatures. Applied Sciences, 13(20). https://doi.org/10.3390/app132011425

Jannach, D., Manzoor, A., Cai, W., & Chen, L. (2021). A Survey on Conversational Recommender Systems. En ACM Computing Surveys, 54(5). Association for Computing Machinery. https://doi.org/10.1145/3453154

Principe, J. C., Euliano, N. R., & Lefebvre, W. C. (2000). Neural and adaptive systems: Fundamentals through simulations. John Wiley & Sons.

Sastry, P. S. (2003). An Introduction to Support Vector Machines. En Computing and Information Sciences: Recent Trends.

Song, Y. Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. Shanghai Archives of Psychiatry, 27(2), 130–135. https://doi.org/10.11919/j.issn.1002-0829.215044

Suthaharan, S. (2016). Machine Learning Models and Algorithms for Big Data Classifi cation Thinking with Examples for Eff ective Learning, 36. http://www.springer.com/series/6157

Biografía de los investigadores.

Esperanza Guerra Rosas, Obtuvo su licenciatura en Ingeniería en Sistemas Computacionales y su maestría en Ciencias en Ingeniería Electrónica en el Instituto Tecnológico de Durango, y su doctorado en Ciencias (Física) en el Departamento de Investigación en Física en la Universidad de Sonora en. Sus áreas de interes incluyen el procesamiento de imágenes, el reconocimiento de patrones y los algoritmos de Inteligencia Artificial. Es miembro de la Academia Mexicana de Óptica y del Sistema Nacional de Investigadoras e Investigadores (SNII).

Josué Álvarez Borrego, obtuvo su licenciatura en oceanografía física en la Facultad de Ciencias Marinas, UABC, Ensenada, México, en 1981, y su maestría y doctorado en Óptica en el Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), México, en 1983 y 1993, respectivamente. Actualmente es Profesor Investigador de la División de Física Aplicada del Departamento de Óptica del CICESE. Sus líneas de investigación incluyen el área de procesamiento de imágenes, reconocimiento de patrones y algoritmos de Inteligencia Artificial. Es miembro de la Academia Mexicana de Óptica y de la Academia Mexicana de Ciencias y de OPTICA (anteriormente OSA), pertenece al Sistema Nacional de Investigadoras e Investigadores (SNII nivel III).

Mauricio. Sauzameda González obtuvo su grado de maestría en Óptica en CICESE en 2024.

Lázaro Bárbaro Fonte Valdés obtuvo su grado de maestría en Óptica en CICESE en 2024.