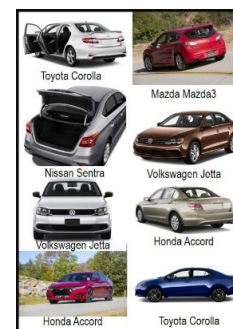


# Reconocimiento automático de automóviles mediante *machine learning*

## Automatic recognition of automobiles using machine learning



Deborah G. Martínez-Camacho<sup>1</sup>, Miguel Torres-Cisneros<sup>2</sup>, Daniel A. May-Arrijo<sup>1</sup>, Mary-Carmen Peña-Gomar<sup>3</sup> y Rafael Guzmán-Cabrera<sup>2</sup>

<sup>1</sup> Centro de Investigaciones en Optica A.C. Calle Prol. Constitución 607, Fraccionamiento Reserva Loma Bonita - 20200 Aguascalientes (México).

<sup>2</sup> Universidad de Guanajuato. Física Aplicada y Tecnologías Avanzadas. Campus Irapuato-Salamanca, 36885. Salamanca (México)

<sup>3</sup> FCFM, Universidad Michoacana de San Nicolás de Hidalgo. C. de Santiago Tapia 403, Centro - 58000 Morelia (México).

DOI: <https://doi.org/10.6036/10673> | Recibido: 26/jul/2022 • Inicio Evaluación: 08/nov/2022 • Aceptado: 21/abr/2023

To cite this article: MARTÍNEZ-CAMACHO, Deborah; TORRES-CISNEROS, Miguel; MAY-ARRIOJA, Daniel; PEÑA-GOMAR, Mary-Carmen; GUZMÁN-CABRERA, Rafael. AUTOMATIC RECOGNITION OF AUTOMOBILES USING MACHINE LEARNING. DYNA. DOI: <https://dx.doi.org/10.6036/2023>

### FINANCIACIÓN

Este trabajo ha sido apoyado parcialmente por el Consejo Nacional de Ciencia y Tecnología (CONACyT) mediante los apoyos: CB2016-01-286368, CB-2016-01-286629 y CF2019-102963. Así como por la Universidad de Guanajuato, mediante los apoyos CIIC 2022 y 2023, y por la Universidad Michoacana de San Nicolás de Hidalgo, mediante el apoyo CIC-2023; "Interferencia de radiación evanescente en el campo lejano".

### ABSTRACT

- In this work, we perform the automatic classification of 1,000 images of five different models of automobiles. To obtain the highest precision, we have used two different classification scenarios, three algorithms, and five metrics. Also, we assume that the results can be improved by extracting the image characteristics using descriptors and using them as input. Then, we used two descriptors: a histogram of oriented gradient and a convolutional neural network ResNet-50. Our results show that the descriptors improve the classification results and obtain the highest value for the accuracy metric of 88.01 % using the ResNet-50 as a descriptor, the Training and Test Set as a scenario, and Vector Support Machine as the classification algorithm.
- Keywords:** Convolutional Neural Networks, Gradient Oriented Histogram, Machine Learning, Fine Grain Classification, Car Images.

### RESUMEN

En este trabajo se realiza la clasificación automática de 1,000 imágenes que contienen cinco diferentes modelos de automóviles. Para ello, y con el objetivo de determinar el método más preciso, se propuso el uso de dos escenarios de clasificación, con tres algoritmos de clasificación y cinco métricas de evaluación. Para mejorar los resultados, decidimos utilizar dos descriptores; el histograma de gradiente orientado y la red neuronal convolucional: ResNet-50, para extraer el vector de características de las imágenes, y utilizarlo como dato de entrada en los algoritmos clasificadores. Los resultados obtenidos muestran un valor máximo de la exactitud de 88.01 % utilizando la combinación: ResNet-50 en el escenario de clasificación: conjunto de entrenamiento y prueba, y el algoritmo de máquina de vector de soporte.

**Palabras Clave:** Redes Neuronales Convolucionales, Histograma de Gradiente Orientado, Aprendizaje Automático, Clasificación de Grano Fino, Imágenes de automóviles.

### 1. INTRODUCCIÓN

En los últimos años, la clasificación de imágenes se ha convertido en una herramienta fundamental para diferenciar entre objetos similares como plantas [1], distintos tipos de células [2] o incluso detectar la presencia de diversas enfermedades [3–5]. En el presente trabajo nos vamos a referir a la clasificación de grano fino; la cual tiene como objetivo la tarea de distinguir subcategorías con pequeñas variaciones entre las diferentes clases [6]. Las imágenes de grano fino son un problema clásico en la clasificación automática realizada utilizando aprendizaje automático [7].

Con referencia a la tarea de clasificación de los vehículos, en la actualidad es considerada de alta prioridad por varias razones; por ejemplo, para identificar autos que no cumplen la ley de tránsito [8], encontrar vehículos robados o dañados [9], o incluso poder identificar el tipo de vehículo en el tráfico [10]. Por estas razones, este trabajo se enfocó en reconocer entre modelos de diferentes marcas con apariencia similar; sin importar el color, la parte del vehículo contenida en la imagen o la posición del automóvil.

Para este objetivo, uno de los principales retos computacionales es extraer la mayor cantidad de características útiles de las imágenes. En este tenor, los descriptores y redes neuronales son ampliamente utilizados para la clasificación, reconocimiento o detección de vehículos. Por ejemplo; el Histograma de Gradiente Orientado (HGO) [11], el vector de Fisher [12] y las características tipo Haar [13], son descriptores que han sido utilizados anteriormente para la extracción de características en la clasificación de vehículos. Los descriptores obtienen la información como bordes, esquinas, intensidad de los píxeles de las imágenes entre otros puntos de interés.

Con el objetivo de realizar una clasificación de grano fino en tiempo real; el presente trabajo utiliza el aprendizaje automático para clasificar cinco modelos distintos de automóviles de la base de datos; "The Car Connection Picture Dataset", la cual está disponible en: [www.kaggle.com/prondeau/the-car-connection-picture-dataset](http://www.kaggle.com/prondeau/the-car-connection-picture-dataset). De manera previa al proceso de clasificación, con el objetivo de extraer características de las imágenes y mejorar la tarea de clasificación, de acuerdo con la literatura presentada previamente, se propone el uso del histograma de gradiente orientado (HGO) y de la red neuronal ResNet-50. Los resultados de esta etapa alimentarán a 3 algoritmos de clasificación: Regresión Logística (RL), Máquina de Vector de Soporte (MVS) y Bosques Aleatorios (BA). Estos clasificadores serán implementados en dos escenarios

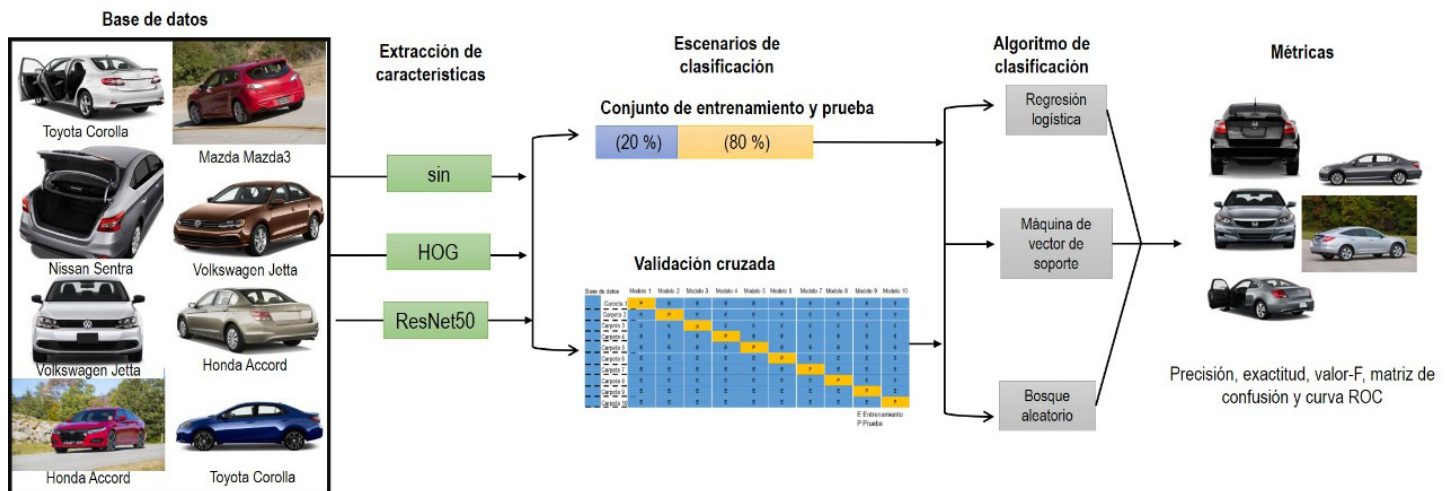


Figura 1: Etapas de la metodología implementada en el presente trabajo; bases de datos, extracción de características, escenarios de clasificación, algoritmos de clasificación y predicción.

distintos: Validación Cruzada (VC) y Conjunto de Entrenamiento y Prueba (CEyP). Finalmente, con el fin de determinar cual de todas estas alternativas arroja los mejores resultados, se utilizaron 5 métricas de evaluación distintas: Exactitud (Accuracy), Precisión, Valor-F, Curvas ROC y Matriz de Confusión.

El resto del artículo se encuentra organizado de la siguiente manera: En la sección 2 se detalla la metodología implementada, incluyendo los métodos de extracción de características, los escenarios de clasificación y métodos de aprendizaje utilizados, así como las métricas de evaluación usadas para medir el desempeño del sistema de clasificación implementado. Los resultados obtenidos son descritos mediante gráficas y tablas en la sección 3. La sección 4 presenta la discusión de los resultados obtenidos y, finalmente, en la sección 5 se presentan las conclusiones del presente trabajo.

## 2. METODOLOGÍA

La metodología implementada en el presente trabajo se divide en cinco etapas, las cuales se pueden apreciar en la Figura 1. Estas son: base de datos de vehículos, extracción de características, escenarios de clasificación, algoritmos de clasificación y por último la etapa de evaluación del desempeño. Observando la Figura 1 es fácil estimar que el número de experimentos por realizar son 18 [ $N=(3)(2)(3)$ ], y que los resultados obtenidos en cada uno de estos 18 experimentos serán evaluados por las mismas 5 métricas. A continuación, describimos los detalles de cada una de estas etapas.

La base de datos utilizada contiene más de 60,000 imágenes con vehículos de distintos tipos, como son subcompactos, compactos, sedán, utilitarios e incluso camionetas de trabajo. Las imágenes de los vehículos en esta base de datos están etiquetadas con la información de las diferentes marcas, modelos, año, precio sugerido del fabricante, rodada de las llantas, etc. Para el presente trabajo, en el que se busca poner a prueba los escenarios y algoritmos de clasificación automática de imágenes, se seleccionaron los modelos de automóviles más comunes y que consideramos presentan mayor similitud entre ellos, pues la intención es realizar una clasificación de grano fino. Por esta razón, redujimos el número de modelos de vehículos a cinco, de diferentes marcas; estos son: Accord (Honda), Corolla (Toyota), Mazda 3 (Mazda), Jetta (VW) y Sentra (Nissan). Para cada uno de estos modelos fueron seleccionadas 200 imágenes, resultando un total de 1,000 imágenes. Cabe

mencionar que las imágenes se seleccionaron de forma manual. Además, se cuidó que estas fueran lo más diversas posibles, en lo que se refiere a colores, fondos, y perspectivas en las que se pueden apreciar diferentes partes de los vehículos, esto es: frente, parte trasera, así como lateral izquierdo o derecho.

Las 1,000 imágenes seleccionadas de la base de datos se utilizarán por los 3 métodos de aprendizaje establecidos en la Figura 1. Sin embargo, en correspondencia con la literatura reportada [14], se consideró que estos algoritmos podrían arrojar mejores resultados en la clasificación si, de manera previa, se realiza la extracción de características de las imágenes utilizando diferentes métodos. Por ello se ha propuesto el uso del histograma de gradiente orientado (HGO) y de la ResNet-50. Una vez que se ha obtenido el vector de características para cada una de las imágenes y que contamos con las etiquetas de las cinco clases, se puede generar la relación que existe entre ellas.

Para realizar la tarea de clasificación, proponemos dos escenarios: la validación cruzada (VC) y el conjunto de entrenamiento y prueba (CEyP). El objetivo es comparar cuál de los dos arroja el mejor resultado de clasificación en las métricas de evaluación utilizadas. La validación cruzada (VC) ya ha sido utilizada extensamente en las tareas de clasificación de vehículos. En este escenario, de acuerdo a reportes sobre la eficiencia de trabajos previos [16], las 1,000 imágenes utilizadas, de los cinco diferentes modelos de vehículos, se dividen aleatoriamente en 10 subgrupos del mismo tamaño balanceados. Balanceado significa que cada subgrupo contiene el mismo porcentaje de las 5 clases (20 %). El primero de ellos se usa para entrenamiento y los 9 restantes para prueba. Posteriormente se utiliza el segundo conjunto para entrenar y se vuelve a probar en los 9 restantes. Esto se repite hasta que los 10 subconjuntos son utilizados como entrenamiento, uno a la vez, y el resto como prueba. El resultado reportado por las métricas de evaluación corresponde a la media de las iteraciones realizadas.

Por otro lado, para el escenario de conjunto de entrenamiento y prueba (CEyP), las 1,000 imágenes son divididas en dos grupos: 80% para entrenamiento y 20 % para prueba. Esta separación es aleatoria y balanceada.

Se propone el uso de 3 diferentes algoritmos de clasificación, los cuales ya han sido ampliamente utilizados en el estado del arte, estos son: máquina de vector de soporte (MVS, se utilizó la función polinomial de grado 3, y un parámetro de regularización  $C = 5.0$ ), regresión logística (RL, se utilizó parámetro de regularización  $C = 5.0$ ) y bosque aleatorio (BA, se utilizaron 750 árboles

en el bosque y la función de entropía para medir la calidad de la división). Cada uno de ellos es utilizado en cada uno de los dos escenarios de clasificación, y a su vez fueron empleados para cada uno de los métodos de extracción de características. Los resultados de cada algoritmo podrán ser comparados mediante el uso de métricas de evaluación similares en cada caso. Los parámetros utilizados en cada algoritmo son mostrados a continuación y fueron los que nos brindaron mejores resultados.

Posteriormente, para evaluar los resultados obtenidos se utilizaron 5 métricas: matriz de confusión, exactitud, precisión, valor-F, y la curva ROC. Las métricas de evaluación nos permiten determinar cuál de las 18 secuencias planteadas para la asignación de la clase a las imágenes bajo estudio permite obtener el mejor resultado y poder comparar los resultados de este trabajo con aquellos reportados de manera previa.

La matriz de confusión es una matriz cuadrada de orden N, en donde N representa el número de clases del experimento. La diagonal principal contiene el número de predicciones que fueron correctamente realizadas (TP o verdaderos positivos), mientras que en el resto de la fila están las predicciones de la clase erróneamente clasificadas en otras clases. Esta métrica es muy útil e intuitiva para medir el comportamiento de los clasificadores.

Todos los experimentos fueron realizados usando Python®, en una computadora Alienware® 17 R5, con procesadores Intel(R) Core(TM) i7-8750H CPU 2.20 GHz, 32 GB RAM y una tarjeta gráfica NVIDIA GeForce GTX® 1070. Además, cada uno de los 18 experimentos fue realizado en 3 ocasiones, entonces, los resultados que se muestran en la siguiente sección, son en realidad el promedio de éstas.

**3. RESULTADOS**

Siguiendo la metodología planteada, se realizaron los 18 experimentos propuestos. La Tabla 1 muestra los resultados obtenidos para las métricas de: exactitud, precisión y valor-F. En ella se pueden apreciar la media y la desviación estándar de las 30 veces que se realiza el experimento en el escenario de VC y de las 3 veces que se realiza en el caso de CEyP. Esto, para cada una de las tres métricas mencionadas en los 18 experimentos propuestos. La última columna muestra la media del tiempo de máquina que consume cada experimento, en segundos.

Con el objetivo de poder apreciar los mejores resultados, la Figura 2 presenta los resultados de las métricas de exactitud, precisión y valor-F, en forma gráfica y que corresponden al caso de referencia, esto es, en el que no hemos usado extracción de características a las imágenes de manera previa. En ella se puede apreciar que la mejor exactitud obtenida fue 77.33 % con una combinación entre CEyP y MVS. Cabe mencionar que para este caso, también las métricas de precisión y de valor-F obtuvieron sus valores máximos con 79.62 % y 77.49 %, respectivamente. Por otro lado, es notable que los valores más bajos se obtuvieron bajo el escenario de VC, usando el algoritmo de RL con 70.64 %, 71.34 % y 70.16 % para la exactitud, precisión y valor-F, respectivamente. Cabe resaltar que aún en este último caso, los resultados son superiores al azar y que la desviación estándar obtenida en las mediciones es superior en un orden de magnitud, comparado a aquellas en las que se obtuvo el valor máximo.

La matriz de confusión nos brinda información complementaria sobre la clasificación obtenida con la secuencia de aprendizaje que mayor exactitud obtiene, que es la que emplea CEyP con el

Sin descriptor (a)					
Método	Algoritmo	Exactitud (%)	Precisión (%)	valor F (%)	Tiempo total (s)
Validación Cruzada	Regresión Logística	70.64 (13.22)	71.34 (13.17)	70.16 (13.14)	1083.45
	Máquina de Vector de Soporte	74.4 (9.51)	77.85 (9.96)	74.19 (9.85)	1385.13
	Bosque aleatorio	71.03 (13.20)	72.04 (13.84)	70.39 (13.25)	<b>1051.32</b>
Conjunto de Entrenamiento y Prueba	Regresión Logística	72 (0.40)	72.39 (0.21)	72.05 (0.53)	1077.11
	Máquina de Vector de Soporte	<b>77.33 (2.32)</b>	<b>79.62 (1.88)</b>	<b>77.49 (2.15)</b>	1261.87
	Bosque Aleatorio	77.17 (2.39)	79.05 (1.55)	78.01 (1.67)	1138.96

Histograma de Gradiente Orientado (b)					
Método	Algoritmo	Exactitud (%)	Precisión (%)	valor F (%)	Tiempo total (s)
Validación Cruzada	Regresión Logística	72.65 (10.69)	73.90 (9.82)	72.44 (10.63)	74.9
	Máquina de Vector de Soporte	78.35 (9.97)	80.27 (8.41)	78.34 (9.79)	86.4
	Bosque aleatorio	74.45 (11.32)	77.01 (10.22)	74.44 (11.23)	83.3
Conjunto de Entrenamiento y Prueba	Regresión Logística	77.68 (2.72)	78.13 (2.54)	77.65 (2.72)	72.8
	Máquina de Vector de Soporte	<b>81.85 (3.32)</b>	<b>82.91 (2.70)</b>	<b>82.08 (3.13)</b>	75.2
	Bosque Aleatorio	78.01 (1.88)	79.35 (1.64)	78.27 (1.87)	80.4

ResNet-50 (c)					
Método	Algoritmo	Exactitud (%)	Precisión (%)	valor F (%)	Tiempo total (s)
Validación Cruzada	Regresión Logística	84.51 (9.81)	85.43 (9.06)	84.56 (9.93)	154.49
	Máquina de Vector de Soporte	82.47 (10.67)	83.27 (10.32)	82.38 (10.84)	156.04
	Bosque aleatorio	78.48 (14)	79.72 (13.74)	78.44 (14.00)	161.24
Conjunto de Entrenamiento y Prueba	Regresión Logística	88.00 (1.46)	88.37 (1.40)	<b>88.17 (1.48)</b>	154.43
	Máquina de Vector de Soporte	<b>88.01 (1.47)</b>	<b>88.38 (1.41)</b>	88.00 (1.47)	155.56
	Bosque Aleatorio	84.40 (2.47)	84.41 (1.83)	83.79 (2.07)	160.17

Tabla 1: Resultados del rendimiento del entrenamiento para tres métricas (exactitud, precisión y valor-F). En estos experimentos a) No se usó descriptor. b) Se usó el método del HGO para extraer características de manera previa y c) Se usó la ResNet-50 para extraer el vector característico de manera previa a la clasificación. Para estos tres casos se consideraron 2 escenarios: VC y CEyP y los 3 métodos de aprendizaje en cada escenario: RF, MVS, y BA.



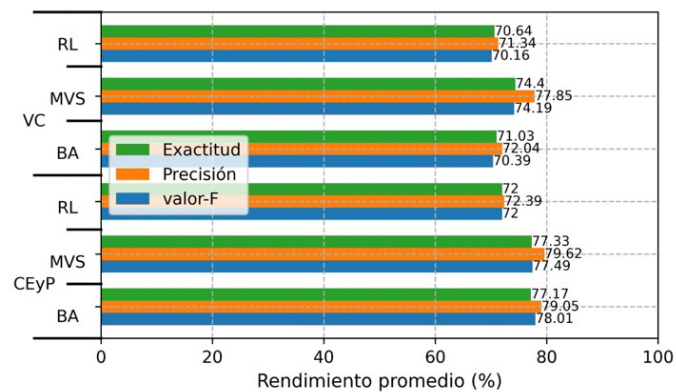


Figura 2: Seis resultados gráficos del rendimiento del entrenamiento para tres métricas (exactitud, precisión y valor-F). En estos experimentos no se usó descriptor y se consideraron los 2 escenarios planteados; Validación Cruzada (VC) y Conjunto de Entrenamiento y Prueba (CEyP) y los 3 métodos de aprendizaje en cada escenario: Regresión Logística (RL), Máquina Vector de Soporte (MVS) y Bosque Aleatorio (BA).

algoritmo MVS. Esta proporciona información directa acerca de cuál marca de vehículo tiene mayor problema para ser clasificada correctamente. Los valores obtenidos en la diagonal principal de la matriz son los verdaderos positivos (TP), es decir, los carros clasificados correctamente. En ella se puede apreciar de manera clara que los vehículos Nissan Sentra, clase 2, se clasifican casi de manera perfecta, ya que se identificaron correctamente 39 de 40. En ella también se aprecia que la clase 4 (Volkswagen Jetta) es la peor clasificada por el algoritmo de aprendizaje, en las circunstancias descritas, ya que solo clasificó correctamente 27 de las 40 imágenes que lo contenían.

La matriz de confusión nos permite seleccionar el algoritmo de aprendizaje adecuado, de acuerdo con nuestros intereses de reconocimiento. Por ejemplo, la Tabla 2 nos muestra los resultados de la matriz de confusión para la secuencia: No descriptor-CEyP-BA. La Tabla 2 y la Figura 4 muestran que esta secuencia obtuvo una precisión de 79.05 %, es decir, apenas 0.57 puntos porcentuales menor a la precisión máxima obtenida, para el caso en que no se usa descriptor. La Tabla 2 nos muestra que, en estas circunstancias, la clase 0 (Honda Accord) es la mejor clasificada con 35 de 40. Es-

MVA					
	Honda Accord	Mazda Mazda3	Nissan Sentra	Toyota Corolla	Volkswagen Jetta
Honda Accord	34	0	2	0	4
Mazda Mazda3	1	29	6	4	0
Nissan Sentra	1	0	39	0	0
Toyota Corolla	3	2	3	32	0
Volkswagen Jetta	5	2	5	1	27
BA					
	Honda Accord	Mazda Mazda3	Nissan Sentra	Toyota Corolla	Volkswagen Jetta
Honda Accord	35	1	0	1	3
Mazda Mazda3	1	25	6	5	3
Nissan Sentra	1	3	32	0	4
Toyota Corolla	1	2	2	34	1
Volkswagen Jetta	4	4	1	1	30

Tabla 2: Matriz de confusión para el experimento que no usa descriptor usando el escenario de Conjunto de Entrenamiento y Prueba (CEyP) Usando Máquina Vector de Soporte (MVS) y Bosques Aleatorios (BA).

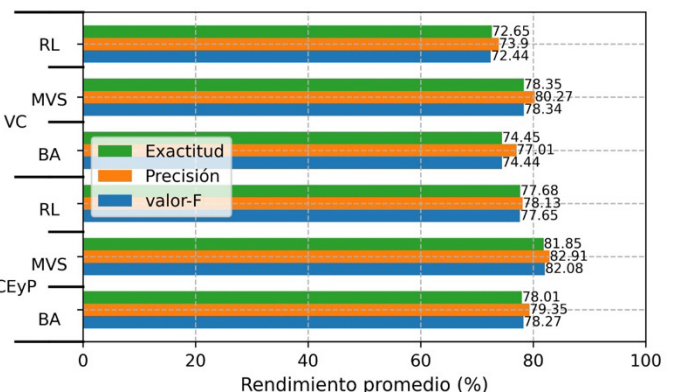


Figura 3: Gráfica del rendimiento del entrenamiento para tres métricas (exactitud, precisión y valor-F) usando HGO para extraer las características de las imágenes. Posteriormente se utilizaron los escenarios Validación Cruzada (VC) y Conjunto de Entrenamiento y Prueba (CEyP), y para cada uno de estos casos, los tres algoritmos de prueba: Regresión Logística (RL), Máquina Vector de Soporte (MVS) y Bosque Aleatorio (BA).

tas dos matrices nos pueden ayudar a decidir que clasificador usar; por ejemplo, si el modelo Mazda 3 (clase 1) no es de interés, se podría utilizar bosques aleatorios (en las condiciones descritas). Sin embargo, si el modelo Jetta es el que no interesa, el uso de la máquina de vector de soporte podría ser más recomendado.

Ahora, si usamos un descriptor para extraer características y particularmente el HGO, los resultados gráficos para las métricas de exactitud, precisión y valor-F, se muestran en la Figura 3. En ella es posible apreciar que la mejor precisión obtenida es de 82.91 %. Este resultado fue obtenido en el experimento que utilizó como escenario CEyP y MVS. En este caso debemos resaltar dos hechos; el primero es que la combinación mencionada genera los mejores resultados en las tres métricas mencionadas, lo que es idéntico al caso anterior, en el que no hay extracción de características de manera previa. El otro hecho, y que es obvio, es que el uso del HGO para extraer las características de las imágenes de manera previa, mejora en un poco más de 3 puntos porcentuales la métrica de la precisión y en más de 4.5 puntos la métrica de exactitud.

La matriz de confusión que corresponde a la secuencia HGO con CEyP y MVS, que es en donde las 3 métricas (exactitud, precisión y valor-F) tienen un máximo, es mostrado en la Tabla 2. En ella se puede apreciar de manera clara que las clases 0 (Honda Accord) y 4 (Volkswagen Jetta) son las mejores clasificadas, ya que se identificaron correctamente 36 de las 40 imágenes en el conjunto de prueba. De hecho, puede apreciarse que el resto de

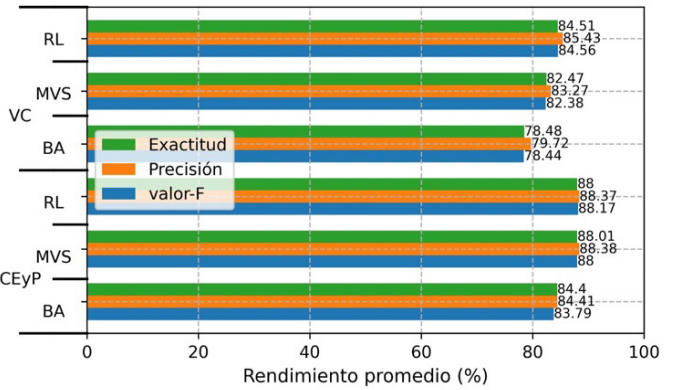


Figura 4: Gráfica del rendimiento del entrenamiento para tres métricas (exactitud, precisión y valor-F) usando ResNet50 para Validación Cruzada (VC) y Conjunto de Entrenamiento y Prueba (CEyP), y para cada uno de estos casos, los tres algoritmos de prueba: Regresión Logística (RL), Máquina Vector de Soporte (MVS) y Bosque Aleatorio (BA).

HGO					
	Honda Accord	Mazda Mazda3	Nissan Sentra	Toyota Corolla	Volkswagen Jetta
Honda Accord	39	0	0	1	0
Mazda Mazda3	2	35	0	2	1
Nissan Sentra	0	0	37	1	2
Toyota Corolla	2	3	1	33	1
Volkswagen Jetta	3	0	1	1	35
ResNet-50					
	Honda Accord	Mazda Mazda3	Nissan Sentra	Toyota Corolla	Volkswagen Jetta
Honda Accord	36	0	1	1	2
Mazda Mazda3	1	33	3	1	2
Nissan Sentra	2	3	34	0	1
Toyota Corolla	2	1	1	34	2
Volkswagen Jetta	0	2	2	0	36

Tabla 3: Matriz de confusión para el experimento que utiliza HGO y ResNet-50 como descriptor, bajo el escenario de Conjunto de Entrenamiento y Prueba (CEyP) con Máquina Vector de Soporte (MVS) como clasificador.

las clases tienen un buen resultado, ya que la más baja tiene una clasificación correcta de 33 vehículos de los 40 en total (82.5 %).

Para el tercer caso, en el que el método usado para la extracción de características fue mediante la red neuronal convolucional ResNet-50, la Figura 4 muestra los resultados gráficos correspondientes para las 3 métricas mostradas anteriormente en la Tabla 2 (exactitud, precisión y valor-F), considerando dos escenarios de clasificación, y para cada uno de ellos, los 3 algoritmos de clasificación. En ella, es posible apreciar que los valores de exactitud y precisión máximos son 88.01 % y de 88.38 %, respectivamente. Estos valores fueron obtenidos en el experimento que utiliza la secuencia CEyP y MVS. A diferencia de los 2 casos previos (sin descriptor y usando el descriptor HGO), esta secuencia no obtiene los máximos valores de la métrica de valor-F, aunque su variación es de tan solo de 17 centésimas de punto, obtenido usando el clasificador de regresión logística (RL).

Los resultados muestran que el uso de la ResNet-50 para extraer características de las imágenes de manera previa a la clasificación mejora el desempeño por casi 9 puntos porcentuales, con

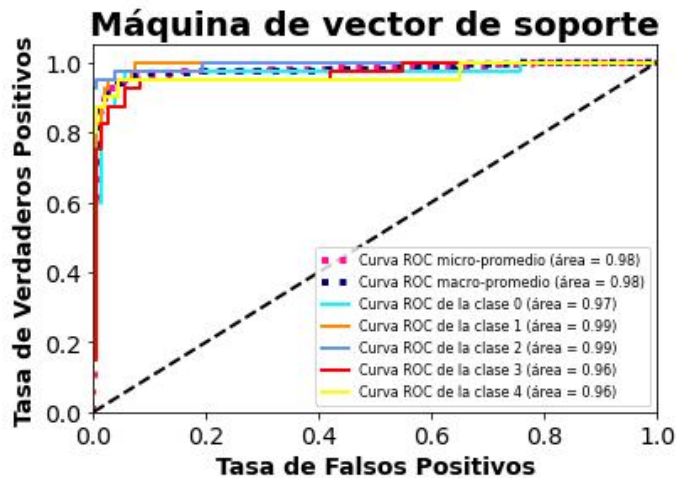


Figura 5: Curva ROC de ResNet-50 con Conjunto de Entrenamiento y Prueba (CEyP) de Máquina Vector de Soporte (MVS).

respecto a no usar ningún descriptor, y casi 6 puntos porcentuales con respecto al uso del HGO para extraer características. La matriz de confusión para este caso se muestra en la Tabla 3. En ella es posible apreciar que la Clase 0 (Honda Accord) es clasificada casi de manera perfecta, ya que 39 de las 40 imágenes que la contenían fueron reconocidas de manera adecuada. Sin embargo, los resultados de clasificación para las otras clases son también bastante buenos, ya que la clase 2 tuvo un 92.5 % positivos, mientras que las clases 1 y 4 obtuvieron un 87.5 %. La clase 3 fue la más baja con un 82.5 % de efectividad.

Comparando las matrices de confusión, mostradas en la Tabla 3, los mayores valores para la exactitud y la precisión para las secuencias que fueron propuestas en este trabajo, fueron obtenidos con la combinación de; ResNet-50 como descriptor, conjunto de entrenamiento y prueba (CEyP), como escenario y a la máquina vector de soporte (MVS) como clasificador, es decir, la que arroja la Tabla 3. Esto implica que la clase mejor reconocida (97.5 % de efectividad) por este método es la Clase 0 (Honda Accord).

Otra métrica que nos permite cuantificar los resultados de las distintas secuencias realizadas en este trabajo es la curva ROC. La Figura 5 presenta las curvas ROC obtenidas para cada clase, cuando se utilizó la secuencia que mejores resultados han tenido con las otras métricas. Esto es; la ResNet-50 con CEyP y MVS. La curva ROC, al ser un método estadístico para determinar la exactitud de las pruebas, nos permite conocer el rendimiento global de cada clase.

4. DISCUSIÓN

Con referencia a la clasificación, detección y reconocimiento de vehículos, es prácticamente imposible realizar una comparación específica con lo publicado en el estado del arte. Este hecho se debe a la gran variedad de las formas, métodos, objetos y características a clasificar e incluso, las bases de datos utilizadas. Por ejemplo, algunos autores han utilizado el aprendizaje profundo en un video de tráfico automotriz para clasificar 3 tipos de vehículos (compactos, medios y de carga), logrando obtener una exactitud de 85.56 % [17], superando el valor de exactitud de 81.42 % reportado al usar el Histograma de Gradiente Orientado (HGO) y la Máquina de vector de Soporte (MVS) para una tarea semejante [15]. Sin embargo, en aprendizaje profundo, el uso de la AlexNet permitió obtener una exactitud de 99.71 % [18].

Aunque, como se mencionó, este trabajo no puede ser comparado directamente con lo que encontramos en la literatura, existen varios puntos que pueden ser resaltados en ese contexto. Considerando las métricas de exactitud, precisión y valor-F; es posible considerar que el uso de algún descriptor mejora los resultados obtenidos, comparado con el hecho de no usarlo. El valor de la exactitud varía desde un mínimo de 70.64 %, obtenida sin el uso de un descriptor y usando la secuencia: VC-RL, hasta un máximo de 88.01 % cuando se realiza la secuencia: ResNet-50-CEyP-MVS.

Sin embargo, lejos de pensar en seleccionar una secuencia específica por el valor de una métrica; los resultados obtenidos muestran que es posible seleccionarla con base en un objetivo específico. Por ejemplo, en lugar de utilizar la secuencia ResNet-50-CEyP-MVS con la que se obtiene el máximo valor de precisión (88.38 %), es posible usar la secuencia HOG-CEyP-MVS con la que es posible obtener una precisión de 82.91 %, pero con la mitad de tiempo de máquina utilizado (75.2 s). De la misma manera, si la métrica de interés se refiere a clasificar correctamente un modelo de vehículo en particular, la matriz de confusión nos permite observar que, mediante el uso del Histograma de Gradiente Orien-

tado es posible clasificar correctamente 36 de 40 automóviles si el modelo es Honda Accord o Volkswagen Jetta. Mientras que, si usamos a la ResNet-50 como descriptor es posible una clasificación casi perfecta del automóvil modelo Accord de Honda con 39 de 40 imágenes.

## 5. CONCLUSIONES

Considerando una base de datos de mil imágenes de automóviles, que muestran diferentes perspectivas de éstos, fue posible obtener valores de exactitud superiores al 81.8 % en tan solo 75.2 segundos, utilizando al Histograma de Gradiente Orientado (HGO) como descriptor, en un escenario de Conjunto de Entrenamiento y Prueba (CEyP) y utilizando el algoritmo de Máquina de vector de Soporte (MVS). La exactitud puede mejorar un poco más de 6 puntos porcentuales si se usa a la ResNet-50 como descriptor, en el caso anterior, con un costo en el tiempo de máquina ligeramente superior al doble de la primera propuesta (155.56 s).

Otros resultados reportados en este trabajo muestran que es posible variar la metodología del modelo de acuerdo con el tipo de resultado que se pretende elegir. Por ejemplo, en la tabla 3 se muestra la matriz de confusión para la segunda propuesta planteada en el párrafo anterior (HOG-CEyP-MVS) en la que es claro que es posible clasificar un modelo de automóvil específico, el Honda Accord (Clase 0) con una precisión de 97.5 % (39 de 40 imágenes fueron clasificadas correctamente). Recordemos que todo este trabajo fue realizado usando librerías de libre acceso y computadora personal.

Por lo anterior, tomando en cuenta los resultados obtenidos en este trabajo, podría concluirse que es posible llevar a cabo la clasificación de vehículos utilizando aprendizaje automático y obtener precisiones relativamente altas con un tiempo computacional bajo; lo que permitiría su uso en tiempo real. Los resultados obtenidos tienen valiosas aplicaciones prácticas, particularmente en sistemas de maquinarias de visión. Esto es, el diseño y la fabricación de sistemas de localización y selección automática de objetos de interés. Estos sistemas incluyen, entre otras cosas; cámaras para captura de imágenes, sistemas de iluminación y algoritmos de clasificación automática, siendo este último en donde se puede incorporar el conocimiento reportado en este trabajo. Estos sistemas pueden ser utilizados para la detección de modelos particulares de automóvil que circulan en carreteras específicas, como lo mostrado aquí.

El trabajo a futuro podría incluir ampliar el tipo de objetos a clasificar, así como el uso de otros algoritmos de clasificación e incluso, el uso de redes neuronales como clasificadores. El incremento en el número de imágenes podría ser soportada con mayores recursos computacionales.

## REFERENCIAS

- [1] M. To açar, B. Ergen, and Z. Cömert. "Classification of flower species by using features extracted from the intersection of feature selection methods in convolutional neural network models". *Meas. J. Int. Meas. Confed.* Vol. 158. 2020. DOI: <https://doi.org/10.1016/j.measurement.2020.107703>
- [2] F. Özyurt. "A fused CNN model for WBC detection with MRMR feature selection and extreme learning machine". *Soft Comput.* 2020. Vol. 24-11. p. 8163-8172. DOI: <https://doi.org/10.1007/s00500-019-04383-8>
- [3] S. Lahmiri. "Hybrid deep learning convolutional neural networks and optimal nonlinear support vector machine to detect presence of hemorrhage in retina". *Biomed. Signal Process. Control.* 2020. Vol. 60. DOI: <https://doi.org/10.1016/j.bspc.2020.101978>

- [4] S. Sharma and R. Mehra. "Conventional Machine Learning and Deep Learning Approach for Multi-Classification of Breast Cancer Histopathology Images—a Comparative Insight". *J. Digit. Imaging.* 2020. Vol. 33-3. p. 632-654. DOI: <https://doi.org/10.1007/s10278-019-00307-y>
- [5] C. Sun et al. "Evaluation of a deep learning-based computer-aided diagnosis system for distinguishing benign from malignant thyroid nodules in ultrasound images". *Med. Phys.* 2020. Vol. 47-9, p. 3952-3960. DOI: <https://doi.org/10.1002/mp.14301>
- [6] T.-Y. Lin, A. RoyChowdhury, and S. Maji. "Bilinear CNN Models for Fine-grained Visual Recognition". *Proc. IEEE Int. Conf. Comput. Vis.* 2015. p. 1449-1457. DOI: <https://doi.org/10.48550/arXiv.1504.07889>
- [7] J. Ji, L. Jiang, T. Zhang, W. Zhong and H. Xiong. "Adversarial erasing attention for fine-grained image classification". *Multimed. Tools Appl.* 2020. DOI: <https://doi.org/10.1007/s11042-020-08666-3>
- [8] Y. Chen, G. Tao, H. Ren, X. Lin, and L. Zhang, "Accurate seat belt detection in road surveillance images based on CNN and SVM," *Neurocomputing*, vol. 274, pp. 80-87, 2018, DOI: <https://doi.org/10.1016/j.neucom.2016.06.098>.
- [9] K. Patil, M. Kulkarni, A. Sriraman, and S. Karande, "Deep learning based car damage classification," *Proc. - 16th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2017*, vol. 2017-Decem, no. May 2018, pp. 50-54, 2017, DOI: <https://doi.org/10.1109/ICMLA.2017.0-179>.
- [10] S. Wang, F. Liu, Z. Gan, and Z. Cui, "Vehicle type classification via adaptive feature clustering for traffic surveillance video," *2016 8th Int. Conf. Wirel. Commun. Signal Process. WCSP 2016*, no. December, 2016, DOI: <https://doi.org/10.1109/WCSP.2016.7752672>.
- [11] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *Proceeding IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 886-893, 2005 DOI: <https://doi.org/10.1109/CVPR.2005.177>.
- [12] T. M. Florent Perronnin, Jorge Sanchez, and T. Mensink "Improving the fisher kernel for large-scale image classification.," *Jln Daniilidis, K., Maragos, P., Paragios, N.*, vol. 6314, no. 4, pp. 143-156, 2010, DOI: [https://doi.org/10.1007/978-3-642-15561-1\\_11](https://doi.org/10.1007/978-3-642-15561-1_11).
- [13] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition.*, vol. 2, pp. 511-518, 2001, DOI: <https://doi.org/10.1109/CVPR.2001.990517>.
- [14] N. Blanc, B. Steux, and T. Hinz, "LaRA SideCam: A fast and robust vision-based blindspot detection system," *IEEE Intell. Veh. Symp. Proc.*, pp. 480-485, 2007, DOI: <https://doi.org/10.1109/ivs.2007.4290161>.
- [15] H. C. Karaimer, I. Cinaroglu, and Y. Bastanlar, "Combining shape-based and gradient-based classifiers for vehicle classification," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Sept 2015, pp. 800-805. DOI: <https://doi.org/10.1109/ITSC.2015.135>.
- [16] G. Singh, A. Professor, and R. K. Panda, "Daily sediment yield modeling with artificial neural network using 10-fold cross validation method: A small agricultural watershed, Kappari, India," *Proceeding Int. J. Earth Sci. Eng.*, vol. 04, no. 6, pp. 443-450, 2011.
- [17] S. Wang, F. Liu, Z. Gan, and Z. Cui, "Vehicle type classification via adaptive feature clustering for traffic surveillance video," *2016 8th Int. Conf. Wirel. Commun. Signal Process. WCSP 2016*, no. October 2016, 2016, DOI: <https://doi.org/10.1109/WCSP.2016.7752672>.
- [18] D. C. de Oliveira and M. A. Wehrmeister, "Using deep learning and low-cost rgb and thermal cameras to detect pedestrians in aerial images captured by multirotor uav," *Sensors (Switzerland)*, vol. 18, no. 7, 2018, DOI: <https://doi.org/10.3390/s18072244>.

## AGRADECIMIENTOS

Este trabajo ha sido apoyado parcialmente por el Consejo Nacional de Ciencia y Tecnología (CONACyT) mediante los apoyos: CB2016-01-286368, CB-2016-01-286629 y CF2019-102963. Así como por la Universidad de Guanajuato, mediante los apoyos CIIC 2022 y 2023, y por la Universidad Michoacana de San Nicolás de Hidalgo, mediante el apoyo CIC-2023; "Interferencia de radiación evanescente en el campo lejano".