

TRABAJO FINAL

DETECCIÓN DE DISTRACCIONES EN
CONDUCTORES MEDIANTE REDES NEURONALES
CONVOLUCIONALES

Alumno: Juan Damián Felix
Director: Ing. Ricardo Petrino



Distracciones en conductores

Cualquier evento o cosa que desvía la atención del conductor de la tarea de manejar de manera segura.



Manuales

El conductor suelta el volante o los pedales para realizar otra acción no necesaria para la conducción.



Visuales

El conductor mantiene la vista fuera de la carretera.



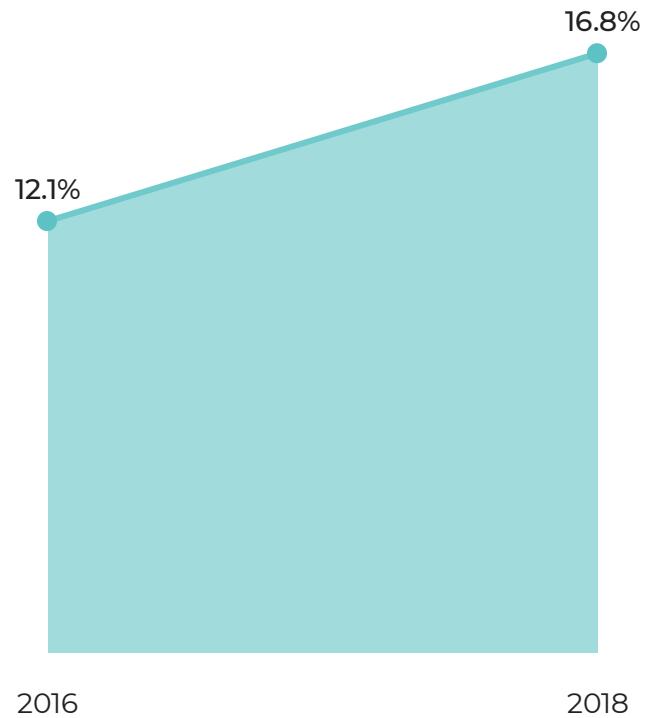
Cognitivas

El conductor se distrae pensando en situaciones que no están relacionadas con la tarea de manejar el vehículo.

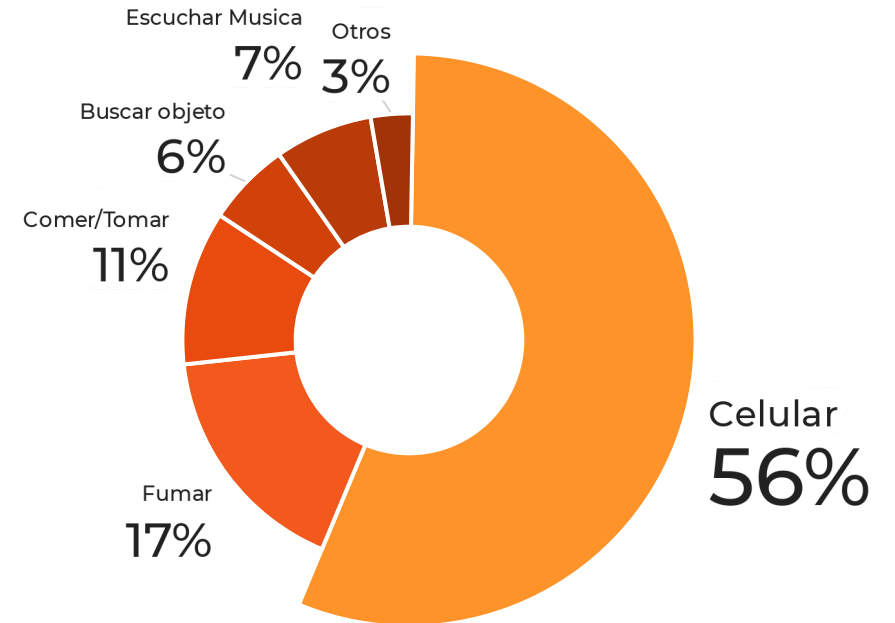
Distracciones en conductores


Datos de la Agencia Nacional de Seguridad Vial tomados en Argentina

Conductores con factores de distracción en Argentina



Principales factores de distracción





Visión artificial y redes neuronales convolucionales



Redes Neuronales Convolucionales

Son un tipo de red neuronal artificial en las cuales la entrada es una imagen y la red está diseñada para procesar datos de imágenes.



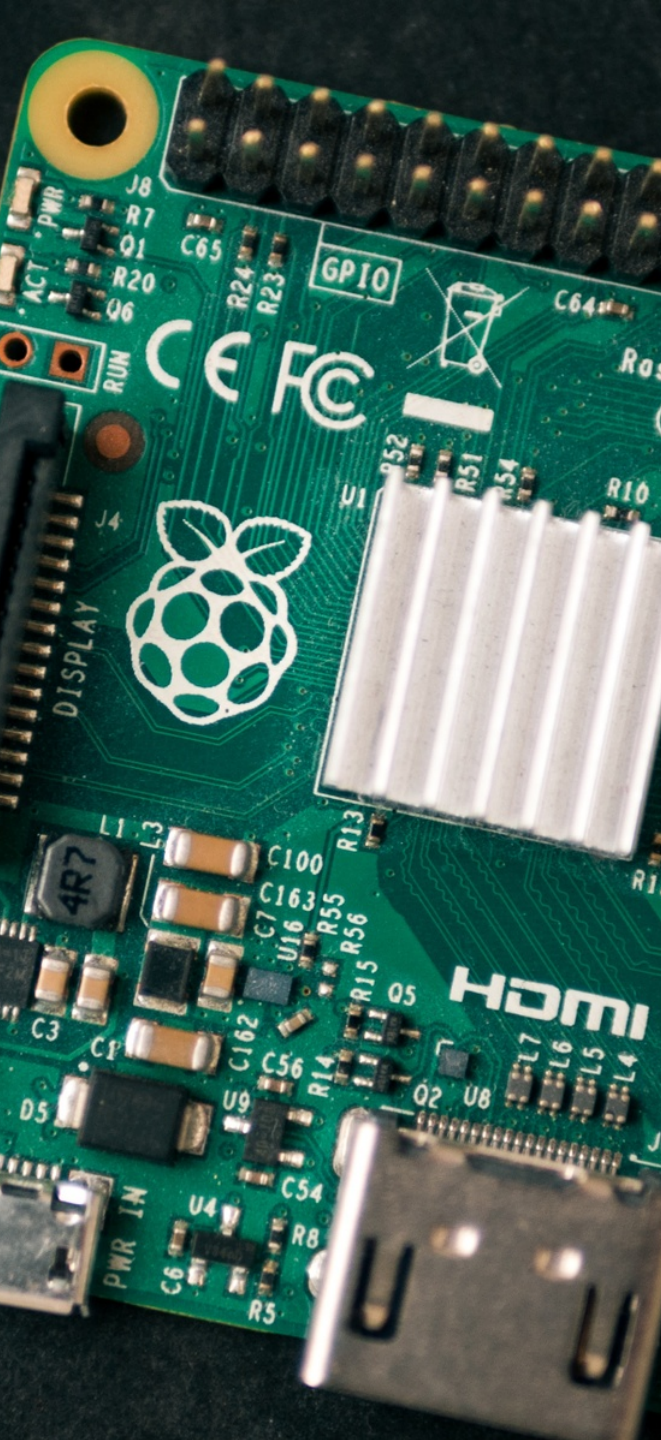
Arquitecturas

La arquitectura de la CNN se compone de capas convolucionales, de activación, pooling, seguidas de capas totalmente conectadas.



Mobilenet

Diseñada para aplicaciones con recursos limitados, como dispositivos móviles o sistemas de visión aplicados en robótica. Utiliza filtros de convolución separables en profundidad, lo que reduce la cantidad de cálculos necesarios y la cantidad de parámetros en la red, haciendo que sea más eficiente y más rápido para ejecutar en dispositivos con recursos limitados.



Que consideremos tiempo real

Capacidad de un modelo para procesar y clasificar una imagen en un corto periodo de tiempo. La duración de este periodo de tiempo depende de la aplicación.

Hardware

El principal dispositivo a utilizarse en este trabajo será una Raspberry Pi 4. Es un dispositivo de bajo costo, tamaño compacto y bajo consumo de energía.

Software

Los algoritmos deben ser lo suficientemente precisos como para garantizar una alta tasa de acierto, pero también lo suficientemente rápidos como para procesar la imagen en tiempo real.

Conjunto de datos

Entrenamiento de la CNN y ajuste fino

Clasificación en tiempo real sobre el hardware elegido

Análisis estadístico de resultados

Diseño del Estudio

Conjunto de datos

- Etiquetado en 10 categorías:

Conducción segura, mensajes de texto, hablando por teléfono, operar la radio, bebiendo, buscando un objeto, mirando en espejo de parasol, mirando al lateral.

- Datos de entrenamiento

Consta de aproximadamente 2200 imágenes brindadas por el dataset original y 327 imágenes tomadas del dataset de elaboración propia, por categoría.

- Datos de validación

Consta de un 1891 imágenes por categoría tomadas del subconjunto de validación que brinda el dataset original y un subconjunto tomado del dataset propio.



Entrenamiento de la CNN

1

Herramientas y plataformas utilizadas para el entrenamiento del modelo

Se utilizaron principalmente las librerías de Tensorflow y Keras, junto con otras librerías como OpenCV, numpy o pandas.

2

Definición de la arquitectura de CNN

Se utilizó la arquitectura MobileNet que está diseñada específicamente para dispositivos móviles con recursos limitados de cómputo y memoria.

3

Entrenamiento de la CNN

Se entrenó el modelo utilizando la técnica de búsqueda en malla para ajustar los hiperparámetros y se usaron callbacks para evitar el sobreajuste.

4

Ajuste fino del modelo

Se reentrenaron las últimas capas del modelo con un ratio de aprendizaje menor, usando imágenes del dataset propio. Esto mejoró notablemente la precisión del modelo.

Preparacion y montaje del hardware

- **Raspberry pi 4**

Buena capacidad para tareas ligeras en tiempo real.
Limitaciones en aplicaciones críticas de alta precisión.

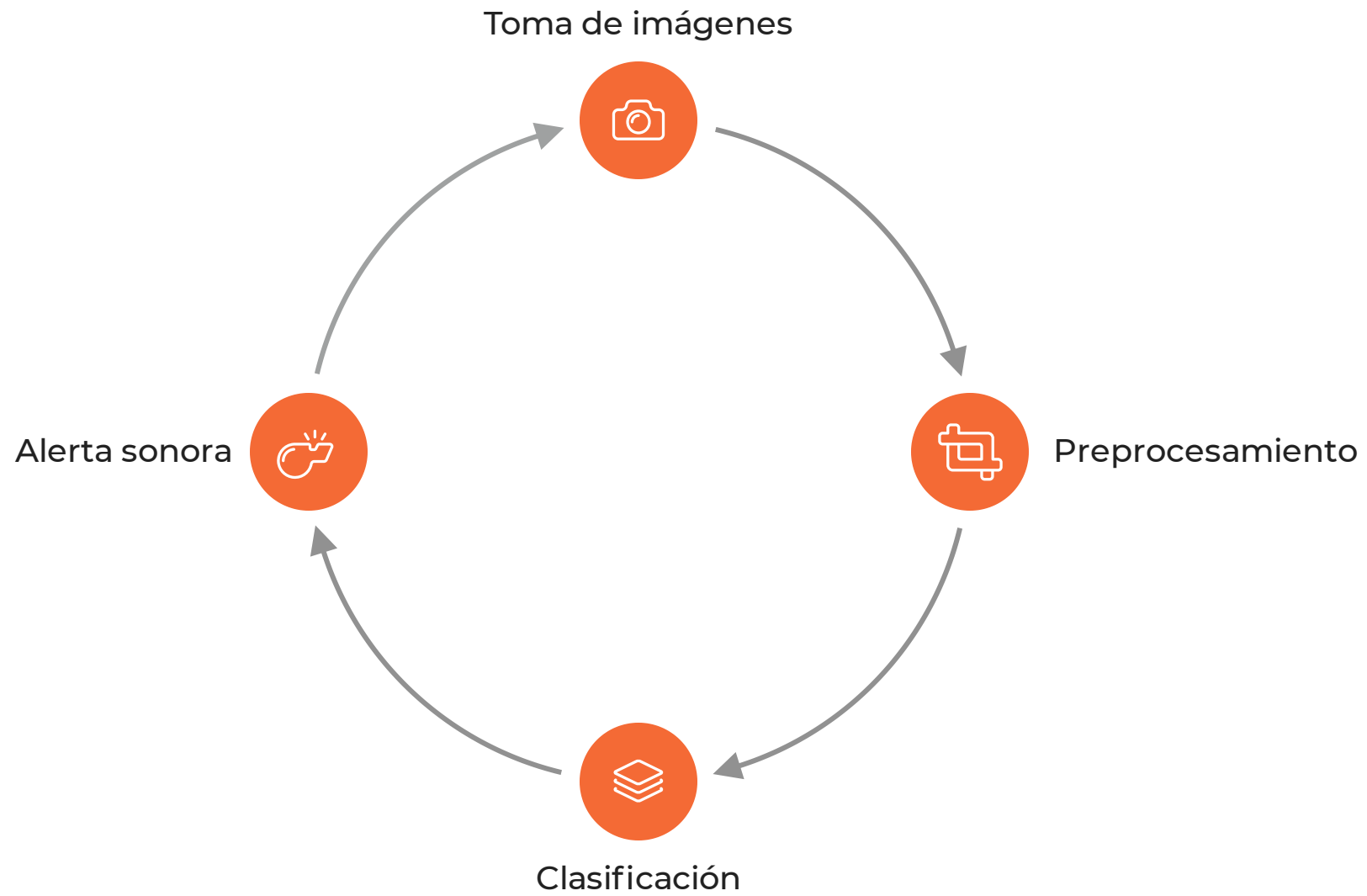
- **Raspberry Pi camera V2**

Sensor de imagen de 8 megapíxeles Sony IMX219 diseñado a medida para Raspberry Pi. Es capaz de imágenes estáticas de 3280 x 2464 píxeles, y también soporta vídeo 1080p30.

- **Preparación del entorno**

Se preparó un ambiente virtual con Tensorflow y Keras para correr el código. Además se instalaron otras librerías que luego se usarán en el código de clasificación en tiempo real, NumPy, Matplotlib, pyaudio y opencv.





Ciclo de clasificación en tiempo real

Video clasificado
por el modelo
entrenado



Conduccion segura

Análisis estadístico

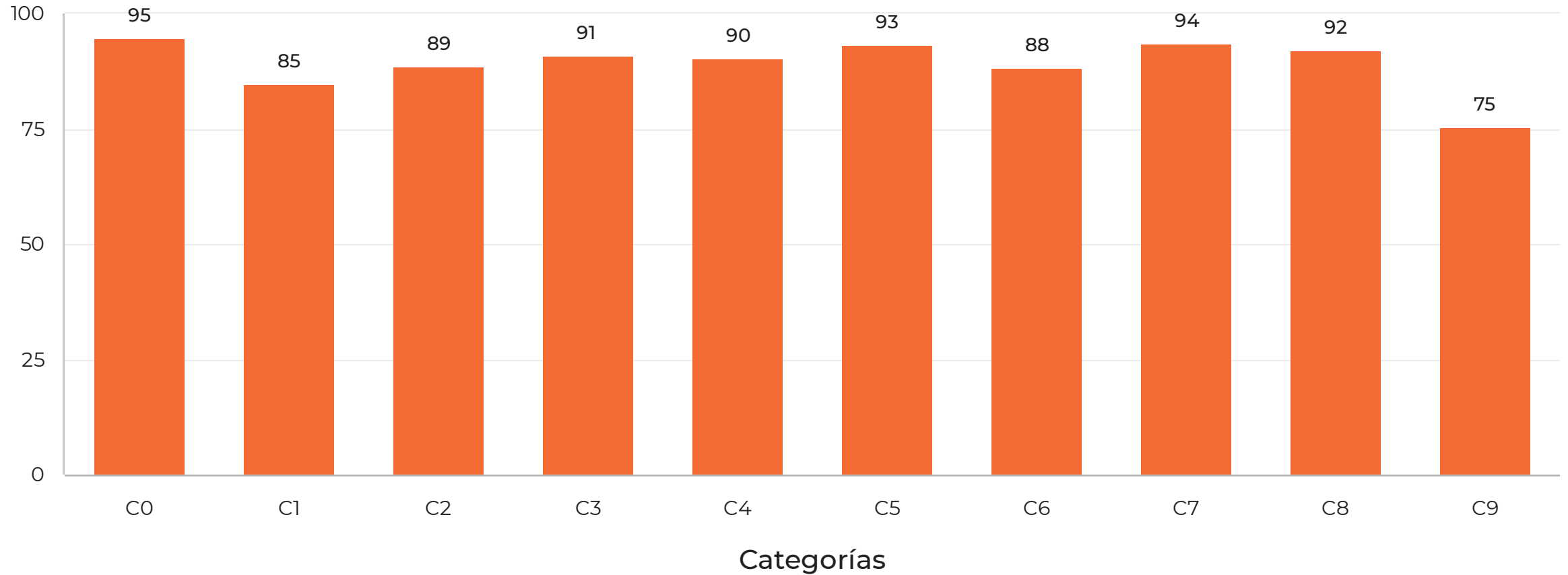
Se aplicaron técnicas estadísticas para calcular diversas métricas de evaluación del modelo, como la precisión, el recall y el valor F1.

La matriz de confusión fue una herramienta fundamental para analizar y comprender los resultados obtenidos (se elaboró con 1891 imágenes por categoría).

	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9
Conducción segura C0	1790	109	12	76	40	50	17	15	18	313
Teléfono en mano izquierda C1	6	1602	16	13	2	0	48	29	9	28
Llamada con mano izquierda C2	1	1	1674	0	1	0	13	7	6	15
Teléfono en mano derecha C3	28	16	7	1722	9	20	2	0	9	29
Llamada con mano derecha C4	9	0	21	9	1708	3	10	2	43	9
Operando la radio C5	11	4	2	6	0	1764	0	4	6	22
Bebiendo / Comiendo C6	4	74	41	8	4	4	1671	9	42	11
Buscando objeto C7	10	32	21	30	26	7	15	1772	7	19
Mirando el espejo de parasol C8	7	6	85	5	85	22	111	21	1744	19
Mirando al lateral / hablando con pasajero C9	25	47	12	22	16	21	4	32	7	1426

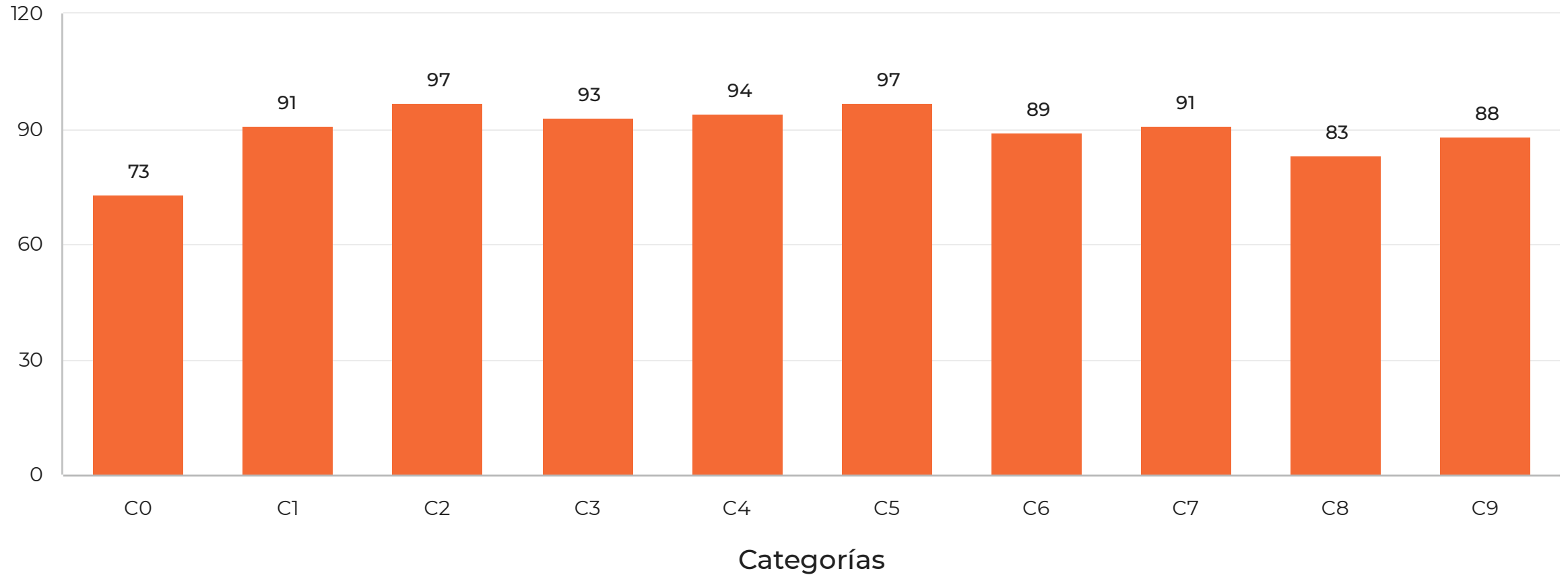
Precisión

Número de predicciones correctas (verdaderos positivos) entre el total de verdaderos positivos y falsos positivos



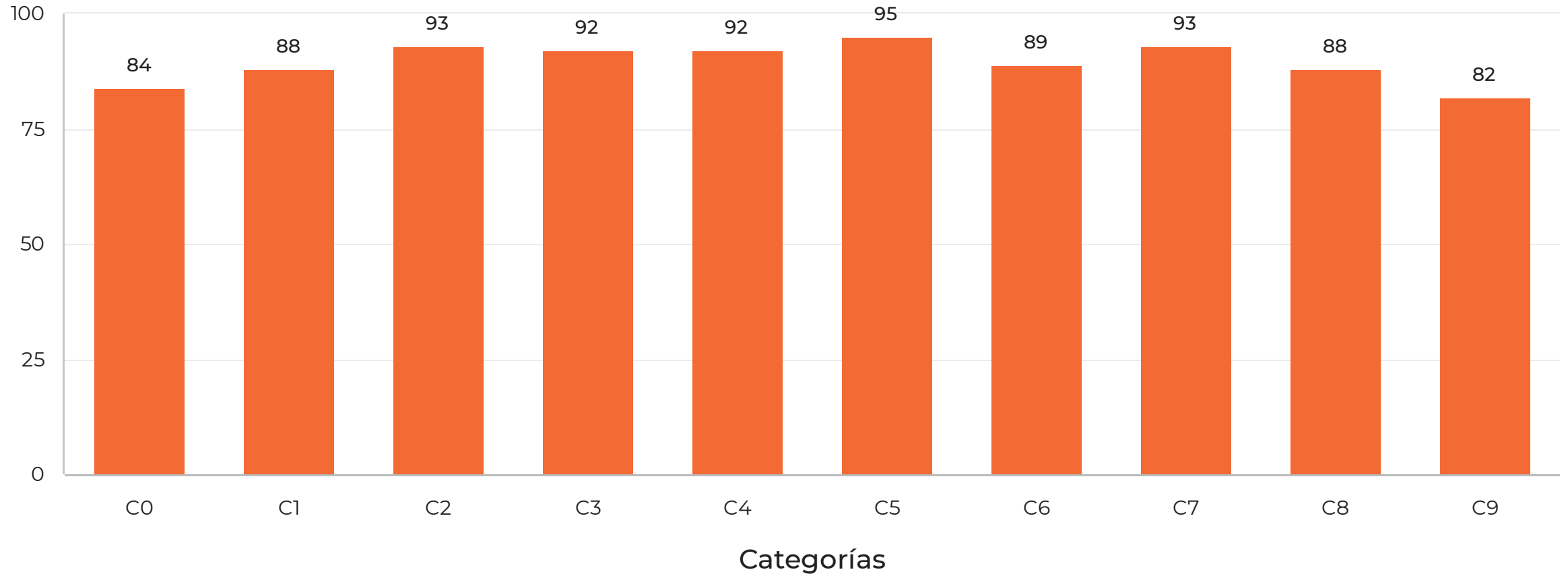
Recall

Es la relación entre el número de verdaderos positivos con la suma de los verdaderos positivos y los falsos negativos



F1 score

Es la media armónica de Recall y Precisión



Exactitud global del modelo

Se calcula como el total de predicciones correctas entre el número total de predicciones realizadas.

89,23%

Análisis cualitativo

Se representaron situaciones de conducción más complejas, que presentaban escenarios y condiciones no contemplados en el conjunto de datos original





Conclusiones

- Buen desempeño de las tecnologías usadas
- Cumplimiento de los requisitos de tiempo real con hardware elegido
- Conjunto de datos apropiado para lograr la precisión buscada
- Necesidad de incluir mas situaciones de distracción y fotografías nocturnas
- Posible uso en vehiculos comerciales de transporte de cargas o pasajeros y mejoras a futuro