

Detección de somnolencia y distracción en conductores y su implementación en dispositivos móviles

Jonathan Flores-Monroy¹ <https://orcid.org/0000-0002-2467-3600>
Mariko Nakano-Miyatake^{1*} <https://orcid.org/0000-0003-1346-7825>
Enrique Escamilla-Hernández¹ <https://orcid.org/0000-0002-4832-8270>
Héctor Pérez-Meana¹ <https://orcid.org/0000-0002-7786-2050>

¹ Escuela Superior de Ingeniería Mecánica y Eléctrica Culhuacán, Instituto Politécnico Nacional, San Francisco Culhuacán, Culhuacán CTM V, Coyoacán, México (Correo-e: jonathan123987@gmail.com, mnakano@ipn.mx, eescamillah@ipn.mx, hmperez@ipn.mx)

* Autor a quien debe ser dirigida la correspondencia.

Recibido Dic. 14, 2022; Aceptado Ene. 31, 2023; Versión final Mar. 29, 2023, Publicado Ago. 2023

Resumen

Este estudio propone un sistema que aplica técnicas de inteligencia artificial para reducir el número de accidentes vehiculares debido al cansancio o distracción del conductor. El sistema puede ser usado en cualquier marca o modelo de vehículo. Inicialmente detecta el rostro del conductor, el cual se introduce a una red neuronal profunda que lo analiza. La salida se inserta en una etapa de detección que determina si el conductor está distraído o somnoliento, activando una alarma. El sistema presenta una baja complejidad computacional, lo cual permite su realización en tiempo real en dispositivos móviles. Los resultados experimentales muestran que el sistema propuesto presenta una precisión de 95.8% al detectar somnolencia y distracción en un cómputo de alto rendimiento. También tiene un 84.5% de precisión en dispositivos móviles con capacidad limitada, manteniendo una velocidad de operación aceptable para la implementación en tiempo real. Se concluye que el sistema propuesto detecta distracción y somnolencia en el conductor con alta precisión.

Palabras clave: detección de somnolencia; detección de distracciones; red neuronal convolucional; dispositivos móviles; detección de rostros; accidentes vehiculares

Detection of driver drowsiness and distraction and its implementation on mobile devices

Abstract

This study proposes a system that applies artificial intelligence tools to reduce the number of vehicular accidents produced by driver drowsiness and distraction. The proposed system, which is suitable for any kind of vehicle, detects the driver's face, which is fed into a deep neural network that analyzes it. The network output is then fed into a detection stage which determines whether the driver is drowsy or distracted, activating an alarm. The proposed system has a low computational complexity, allowing real time implementation on mobile devices. The experimental results show that the proposed system can detect drowsiness and distraction with an accuracy of 95.8% in high performance computers. It also shows an 84.5% accuracy on mobile devices with limited capacity, keeping an acceptable operational speed for its implementation in real time. It is concluded that the proposed system detects driver drowsiness and distraction with high accuracy.

Keywords: drowsiness detection; distraction detection; convolutional neural networks; mobile devices; face detection; car accidents

INTRODUCCIÓN

Los accidentes vehiculares son un serio problema en todas partes del mundo. En México este problema es alarmante ya que lo ubica en el séptimo lugar mundial y tercer lugar en Latinoamérica en lo que se refiere a muertes causadas por accidentes vehiculares, de acuerdo con los datos del Instituto Nacional de Salud Pública (Fact and Stats, 2022). La Tabla 1 muestra las causas de los accidentes de tráfico en México de acuerdo con la información reportada por el Instituto Nacional de Estadística, Geografía e Informática (INEGI, 2022). De aquí se observa que más del 90% de los accidentes son causados negligencia o irresponsabilidad del conductor, mientras que el resto debe a descuidos de los pasajeros o peatones, problemas mecánicos, daños en los caminos y otros factores naturales.

Entre los factores que se pueden atribuir al conductor están: Manejar bajo la influencia del alcohol y drogas; imprudencia al conducir; exceso de velocidad al conducir, condición física deficiente del conductor; así como cansancio y somnolencia del conductor cuando maneja con fatiga. Varios reportes estadísticos a nivel mundial (CNS, Traffic Safty Culture Index, 2019) muestran que el 33% de los accidentes atribuibles al conductor se deben a la falta de sueño, fatiga física y mental del conductor; así como la condición monótona del camino. Otro factor importante es el uso de teléfonos celulares durante la conducción lo que causa la falta de atención del conductor en los eventos circundantes. Considerando que la mayoría de los accidentes son causados por el conductor y cerca del 33% de estos se deben a somnolencia, cansancio y distracción del conductor, durante los últimos años se han propuesto diversos esquemas para detectar somnolencia y distracción del conductor a fin de evitar un lamentable accidente (Chacon, 2015; Flores-Monroy et al. 2021; Uma y Eswari, 2021).

Tabla 1: Reporte del INEGI: Causas de accidentes de tráfico en México (INEGI, 2022)

Variable	2016	2017	2018	2019	2020
Número total de eventos	360,051	367,789	365,281	362,729	301,678
Conductor	91%	91%	92%	92%	95%
Peatones	1%	1%	1%	1%	1%
Fallas mecánicas	1%	1%	1%	1%	1%
Malas condiciones del camino	3%	3%	3%	3%	2%
Otras causas	4%	4%	3%	3%	1%

Las técnicas usadas para detectar somnolencia en los conductores se pueden dividir básicamente en tres grupos: Métodos basados en comportamiento del vehículo, métodos basados en señales biológicas de conductor y el método basado en visión por computadora. El primer grupo trata de detectar anomalía causada por somnolencia usando el comportamiento de vehículo, tales como la velocidad o aceleración, la dirección de manejo y la desviación de la trayectoria de este con respecto a las líneas del camino (Wu et al., 2013; Zhang et al., 2014). Este primer grupo requiere señales del vehículo, además, ya que el comportamiento del vehículo depende de la forma de manejo del conductor y el tipo del vehículo, la exactitud de detección es baja en general. El segundo grupo se basa en el análisis de las señales biológicas del conductor tales como: el electrooculograma (EOG), electroencefalograma (EEG), electrocardiograma (ECG) y algunos compuestos de la sangre obtenidos mediante luz infrarroja, entre otros (Kokonosi et al., 2008; Vidente et al. 2016). Este grupo requiere algún dispositivo para medir señales biológicas del conductor. Estos dispositivos son generalmente invasivos, aunque sean mínimos, que afectan negativamente la conducción de vehículo, por lo tanto, muchos conductores no aceptan técnicas de este grupo.

El último grupo está basado en señales visuales del conductor tomadas por una webcam puesta dentro del vehículo. Las imágenes del conductor se analizan usando técnicas de aprendizaje automático (Fierro et al., 2019) para determinar somnolencia y distracción en el conductor. Este último grupo no es invasivo y usando los algoritmos actuales basados en aprendizaje profundo la exactitud de detección es muy aceptable (Phan et al. 2021; Flores-Monroy et al. 2022; Magàn et al., 2022; Tamanani et al. 2021). Sin embargo, los trabajos realizados en este grupo requieren equipos de cómputo con mayor capacidad de cálculo y una gran cantidad de memoria para poder llevar a cabo la detección, lo cual dificulta una implementación compacta dentro del vehículo. El uso de cómputo en la nube es una alternativa (Adil et al. 2023; Uma y Eswari, 2021), aunque en este caso requiere un administrador de la nube y la constante comunicación entre el vehículo y la nube que en muchas regiones no se puede garantizar.

Con el fin de evitar inconvenientes los mencionados en los dos grupos relacionados con las técnicas existentes y en el uso del cómputo en la nube, este artículo propone un sistema para la detección de somnolencia y distracción en conductor basado en visión por computadora. El sistema es portátil y está optimizado para operar en tiempo real en dispositivos de cómputo con capacidad limitada tales como teléfonos celulares y otros dispositivos móviles. Adicionalmente el sistema propuesto esta desarrollado para operar en cualquier tipo de automóvil ya que no requiere ninguna instalación especial. Este no es invasivo para el automóvil y para el conductor, ya que únicamente requiere una toma de corriente donde se pueda conectar el dispositivo.

El sistema propuesto detecta tanto el cabeceo como la distracción del conductor, de manera que no solamente detecta somnolencia sino la falta de atención del conductor en el camino; previniendo accidentes por estos dos motivos mediante la emisión de alarma. Si el conductor hace caso omiso a las mismas, el sistema envía la ubicación del vehículo a un número telefónico determinado previamente. En lo que sigue se analizan publicaciones previas relacionadas con el presente trabajo, en particular, sistemas que detectan somnolencia y/o distracción en conductor usando dispositivos móviles, tal como teléfono celular, que son compactos y se pueden colocar fácilmente dentro de cualquier vehículo.

OTROS ANTECEDENTES

Jabbar et al., (2018) propusieron un sistema para detectar la somnolencia del conductor y su implementación en dispositivos móviles, en el cual las coordenadas de los puntos característicos se extraen de la imagen del rostro usando la librería Dlib de OpenCV. Seguidamente los puntos característicos son introducidos a una red neuronal artificial (ANN). Este algoritmo, el cual proporciona un reconocimiento de 81.00%, se programó en la plataforma Android. Polumpung et al. (2022) propusieron un esquema para detectar somnolencia en conductores usando algoritmos de rastreo de objetos basado en color junto con el algoritmo de Viola y Jones (2004) para rastrear el rostro y ojos del conductor. El algoritmo se implementó en la plataforma Android para poder ejecutar en teléfonos celulares. La exactitud de la detección de somnolencia es 78.81%. En (Shaily et al., 2021), se implementó un sistema de detección de somnolencia del conductor bajo esquemas de diagnóstico a bordo usando una cámara instalada dentro de vehículo. El algoritmo de detección consiste en detectar el rostro usando el algoritmo de Viola y Jones (2004) y el ratio de aspecto de los ojos obtenido usando puntos relevantes de rostro. El sistema fue construido usando una Raspberry Pi en la plataforma Android. Los autores no proporcionaron el funcionamiento de sus trabajos, tales como exactitud y la velocidad de los procesos que alcanza en operaciones en tiempo real.

Dentro de los algoritmos de detección de somnolencia basados en aprendizaje profundo, se encuentra sistema propuesto por Shakeel et al. (2019), en el cual se usa MobileNet-SSD como el localizador de rostro, ojos, nariz y boca del conductor. El sistema fue implementado en la plataforma Android versión 5.0 con una velocidad de operación máxima que se puede alcanzar con el teléfono celular de alta capacidad, tal como Google Pixel 2, es 12-16 tramas por segundo (fps). La exactitud de detección de somnolencia es 83.7%. Mientras Wijnands et al. (2020) propusieron un sistema de detección de somnolencia para plataforma móvil usando CNNs tridimensionales (3D) separables en profundidad para manejar tanto la información espacial como información temporal del video obtenido por la cámara de dispositivo móvil. La CNN propuesta fue entrenada usando la base de datos de ImageNet y Kinetics, y se realizaron ajustes usando la base de datos del conductor. La exactitud reportada en el artículo es 73.9% con una velocidad de proceso de 10 fps. Yu et al. (2019) presentaron un sistema basado aprendizaje profundo, llamado DrowsyDet que consiste en tres CNNs que se encargan de caracterizar rostro completo, ojos y boca, respectivamente, para determinar el estado de somnolencia en el conductor. La exactitud del sistema es 97.8%, sin embargo, se puede operar máximo 6 fps lo que dificulta una implementación en tiempo real del sistema.

SISTEMA PROPUESTO

El diagrama a bloques del sistema propuesto de detección de somnolencia y distracción, adecuado para su realización en un dispositivo móvil se muestra en la Fig. 1. Inicialmente se captura la señal de video de entrada la cual se segmenta en M tramas. Seguidamente, debido al orden en el cual los paquetes de software usados en el dispositivo móvil procesan las imágenes, se modificará el orden de los planos RGB, empleados en imágenes a color por un arreglo BGR. Seguidamente se modifica el tamaño de las tramas de manera que la detección del rostro sea factible para su realización en tiempo real. Una vez reducido el tamaño de las tramas se lleva a cabo la detección del rostro y la segmentación de la región respectiva a fin de obtener la región a analizar. En caso de que no se detecte la presencia del rostro del conductor, se envía la información a la etapa de activación/desactivación de la alarma donde se determina si se activa la alarma debido a la falta de atención del conductor. En caso de que se detecte la presencia del conductor, se procede a segmentar la región para continuar con el análisis, modificando el tamaño de la región de interés a fin de adecuarla a las características de la red neuronal profunda usada para el análisis. La salida de la red se introduce entonces,

a una etapa de análisis la cual determina si se activa la alarma dependiendo de los resultados obtenidos. Esto se lleva a cabo junto con el despliegue de la imagen del conductor, para lo cual se incrementa el tamaño de la imagen y se revierte el orden de los planos de color de BGR a RGB. A continuación, se describe cada una de las etapas que componen el sistema.

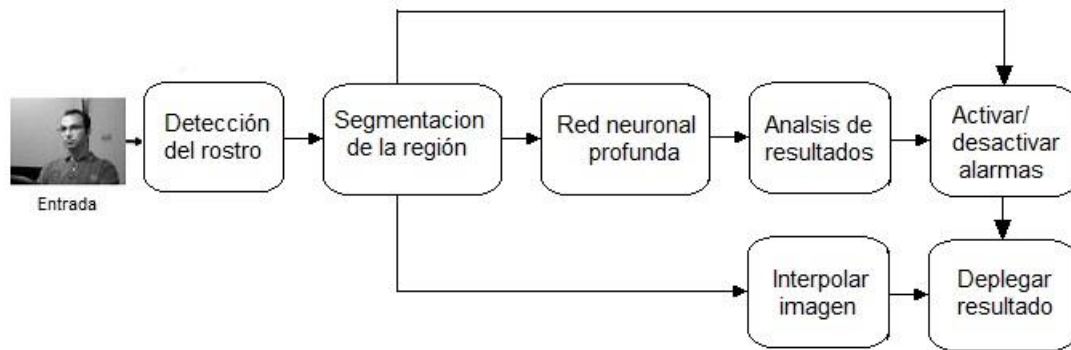


Fig. 1: Diagrama a bloques del sistema propuesto para la detección de somnolencia y distracción

Detección y segmentación de la región del rostro del conductor

Inicialmente el sistema captura la secuencia de video a partir de la cámara del dispositivo móvil o de una cámara externa, como se muestra en la Fig. 2, y procede a segmentarla en tramas. Seguidamente se procede a adecuar el tamaño de la trama de entrada y en número de estas que permitan una eficiente detección del rostro en tiempo real. Para estos módulos de gran importancia para la detección del rostro son la reducción del tamaño de la trama de entrada y la detección misma del rostro. Así, una vez que el video de entrada ha sido segmentado en tramas, el tamaño de cada una de las tramas es reducido para incrementar la cantidad de tramas por segundo que el sistema móvil es capaz de procesar de manera que sea factible su operación en tiempo real. Esto se debe a que, suponiendo que el sistema propuesto se programa en un teléfono Samsung s20fe cuya pantalla tiene una dimensión de 864×1920 píxeles, lo cual representa una cantidad importante de costo computacional, el número de tramas que se pueden procesar es bajo, dando como resultados un número muy bajo de rostros detectados por segundo, como se muestra en la Fig. 3. En esta figura se puede observar que mientras el dispositivo es capaz de desplegar 15 fps sin operación de detección del rostro, cuando se agrega la operación, este número se reduce a solamente 4 fps, lo cual es insuficiente para una adecuada operación en tiempo real, siendo necesario realizar una reducción del tamaño de la trama antes de llevar a cabo la operación de detección del rostro.

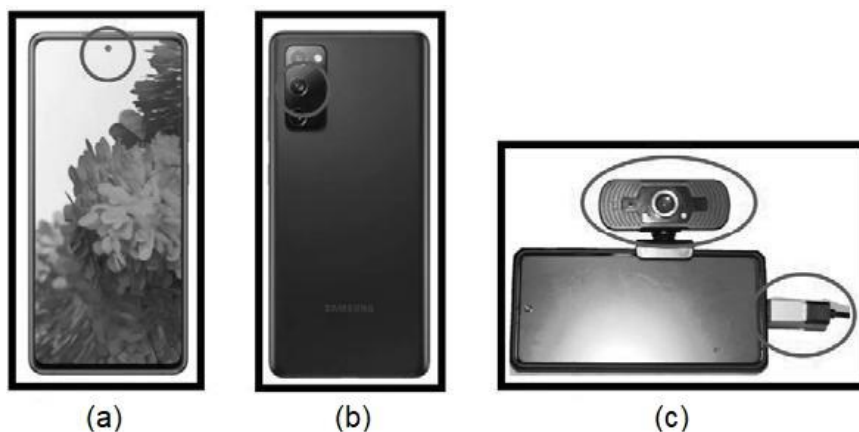


Fig. 2: Dispositivo de captura de las imágenes. (a) Cámara frontal del dispositivo, (b) cámara trasera del celular, (c) cámara externa del celular la cual puede ser infrarroja.



Fig. 3 Velocidad de procesamiento de video en fps cuando la trama es de 864×1920 píxeles. (a) sin operación de detección de rostro. b) con operación de detección de rostro.

Con la finalidad de determinar el tamaño máximo de la trama permitido para la operación del sistema en tiempo real, se tomaron en cuenta varios factores tales como: la reducción del tamaño de la trama con respecto a la original, el número de tramas por segundo que la reducción permite procesar en tiempo real y finalmente el método de reducción empleado. Después de evaluar varios esquemas se determinó emplear el método llamado inter-área (Magán et al., 2022) en el cual la imagen original se segmenta en bloques de $m \times n$ píxeles el cual se reduce a un solo píxel promediando los píxeles contenidos en el bloque, como se muestra en la Fig. 4. Este esquema muestra un adecuado compromiso entre la complejidad de procesamiento y la calidad de la imagen. Los resultados de esta evaluación se resumen en la Tabla 2. De aquí se observa que con una reducción del tamaño de la trama a $(m/3) \times (n/3)$, se obtiene una velocidad de procesamiento de 13 fps lo cual permite un procesamiento adecuado de las etapas que componen el sistema; ya que, si bien reducciones mayores permiten un ligero incremento, sin embargo, el tamaño del rostro resultante podría dificultar la operación del sistema debido al tamaño de las imágenes después de su reducción.

Tabla 2: Relación entre el tamaño de la trama y la velocidad de procesamiento

Factor de reducción	Tamaño resultante	Número de tramas por segundo (fps)	Incremento de la velocidad
1/1	864×1920	4	1
1/4	432×960	9	2.25
1/9	288×640	13	3.25
1/16	216×480	14	3.50
1/25	173×384	15	3.75
1/36	244×320	15	3.75

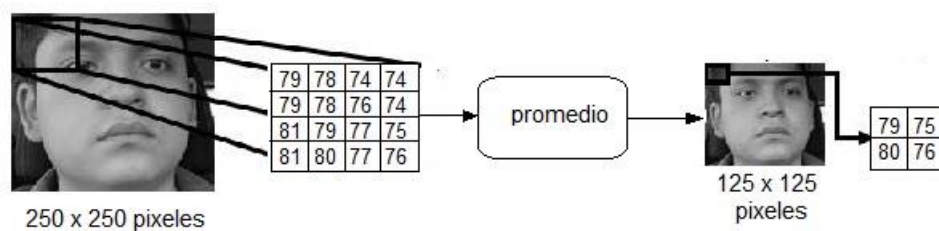


Fig. 4: Esquema empleado para reducir el tamaño de la trama

Una vez reducido el tamaño de la trama se procede a detectar el rostro del conductor empleando el esquema de Viola y Jones (2004). Si no se detecta el rostro, se envía un mensaje al bloque de análisis por medio del cual determina si se debe activar la alarma. Si se detecta el rostro del conductor en la trama bajo análisis, se procede a segmentar la región del rostro la cual se reduce a un tamaño estándar de 64×64×3, como se muestra en la Fig. 5, la cual se inserta en una red neuronal profunda de tipo CNN.



Fig. 5: Imagen de entrada a la red neuronal profunda

Red Neuronal Profunda

Una vez que se segmenta el rostro del conductor, la imagen resultante se inserta en una red neuronal profunda (Flores-Monroy et al., 2021), mostrada en la Fig. 6, la cual llevará a cabo el proceso de detección de somnolencia o distracción. Esta, como se muestra en la Fig. 6, consiste en tres capas convolucionales. Las primeras dos con 32 kernels de 3x3, mientras que la tercera consta de 64 kernels de 3x3, con una función de activación ReLU y un proceso de MaxPooling de 2x2. Aquí la primera capa crea 32 mapas de características (MCs) de 31x31. La segunda capa produce también 32 MCs de 14x14. Finalmente, la tercera capa convolucionales produce 64 MCs de 6x6. Una vez obtenida la salida de la tercera capa convolucional, las imágenes obtenidas se arreglan en forma de vector de 2304 elementos, el cual se introduce en una red neuronal completamente conectada. Seguidamente, durante el entrenamiento de las capas completamente conectadas se emplea el procedimiento conocido como Dropout (Goodfellow et al., 2016), el cual consiste en desactivar del entrenamiento, en cada iteración durante el proceso de entrenamiento, de un número dado de neuronas. Este mecanismo permite reducir el sobreajuste generado durante el proceso de entrenamiento. En este caso se aplicaron dos procesos de Dropout del 20% del total de neuronas empleadas. Finalmente, el sistema propuesto cuenta con tres capas completamente conectadas. La primera con 2304 valores de entrada y 128 de salida, la segunda, que también emplea Dropout con 128 neuronas de entrada y 128 de salida, y finalmente la tercera capa completamente conectada que cuenta con 128 neuronas de entrada y 3 neuronas de salida. Esta capa extraerá los últimos patrones de la capa anterior y los clasificará en tres estados probabilísticamente usando como la función de activación Softmax (Goodfellow et al., 2016), arrojando como resultado un vector de tres componentes. Como función de pérdida se usó entropía cruzada categórica, ya que la salida de la red es un vector de tres elementos que corresponde a la probabilidad de las tres categorías: normal, somnoliento y distraído, y el optimizador ADAM con tasa de aprendizaje 0.001 y decrecimientos exponenciales $\beta_1=0.9$ y $\beta_2=0.999$, fue seleccionado de forma empírica. El número de épocas fue 100, ya que más épocas causaba sobreajuste.

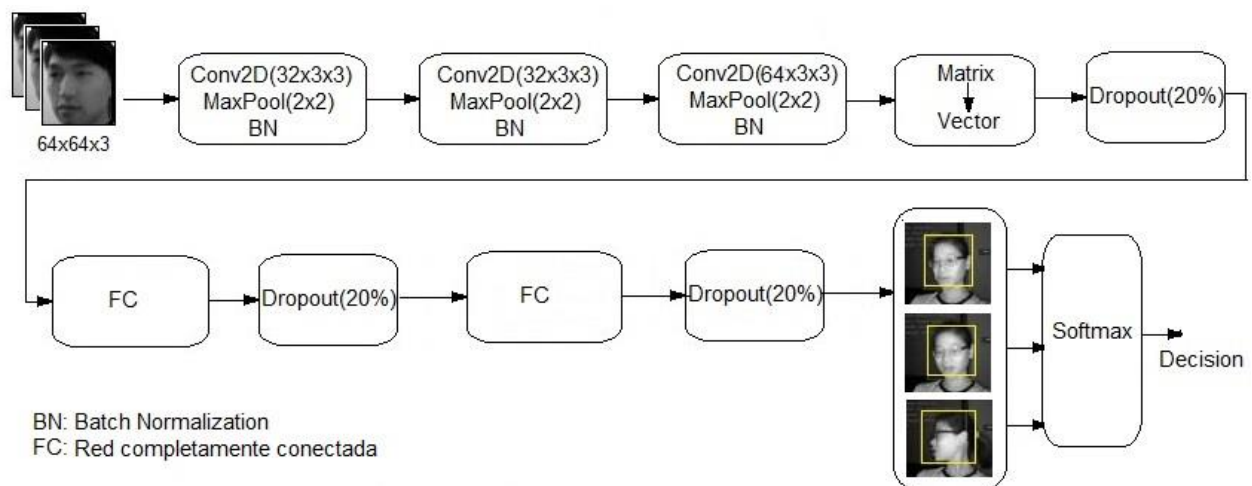


Fig. 6: Red neuronal profunda empleada.

La salida de la red, obtenida para cada una de las tramas consecutivas, se analiza para diferenciar entre el parpadeo normal y la somnolencia real, tomando en cuenta que en la somnolencia real el intervalo de tiempo durante el cual los ojos del conductor permanecen cerrados deberá exceder cierto umbral el cual, experimentalmente se determinó igual a 1.5s (Kwon et al. 2013), considerando la diferencia entre el parpadeo normal y la somnolencia real.

Datos de entrenamiento de la red

La base de datos usada para generar los conjuntos del entrenamiento y prueba del sistema propuesto es la NTHU-DDD de la Universidad Nacional Tsing Hua (Weng et al., 2017), la cual consiste en 36 sujetos de diferentes etnias bajo 5 diferentes condiciones las cuales son: a) sujetos con lentes bajo iluminación normal, b) Sujetos sin lentes bajo iluminación normal, c) sujetos con lentes de sol bajo iluminación normal, d) sujetos con lentes bajo iluminación infrarroja y e) sujetos sin lentes bajo iluminación infrarroja. De aquí se eliminan las secuencias de video en las cuales los conductores usan lentes de sol, ya que la detección de somnolencia requiere determinar si los ojos del conductor están abiertos o cerrados. La duración de cada video es de aproximadamente 1.5 min y el tamaño de cada trama original es de 640x480 píxeles.

Desde esta base de datos, 4800 tramas de video se seleccionaron para entrenamiento y prueba del sistema. Las 4800 tramas se analizaron para etiquetar equitativamente a tres estados de conductor, los cuales son "normal", "somnoliento" y "distráido", ya que en la base de datos NTHU-DDD original, las tramas están etiquetadas en dos estados: "normal" y "somnoliento", este trabajo fue necesario. La Fig. 7 muestra los criterios tomados para la asignación de tres estados del conductor a cada trama de video. De aquí se observa que cuando el conductor no tiene somnolencia y está atento al camino, como se muestra en Fig. 7(a), se puede usar la posición de los ojos, 1, las cejas, 2, la frente, 3, las líneas de expresión, 6, y la posición de la boca, 5. La Fig. 7 (b) muestra algunas características que se pueden emplear para determinar somnolencia como son la posición de las cejas, 2, la apertura de los ojos, 1 y la posición de la boca y nariz, 3, 4. Finalmente, algunas de las características que se considerarán para determinar distracción será básicamente la presencia de al menos el 75% de las características del rostro como se muestra en la Fig. 7(c). Aplicando este criterio a las tramas de video, se construyó la base de datos para nuestro sistema, siendo 1600 tramas para la clase "normal", 1600 tramas para clase "somnolencia" y el resto 1600 tramas para clase "distracción". Las tramas de cada clase se dividieron 75% para el entrenamiento, el restante 25% se usó para la validación y la prueba. Durante el entrenamiento no se ha usado el método de validación cruzada, ya que datos de validación no se usaron en el proceso de entrenamiento.

Esta cantidad de tramas para el entrenamiento no es suficiente para llevar a cabo un entrenamiento adecuado para que la red neuronal profunda tenga una capacidad de generalización. Con el fin de aliviar este problema se emplea la técnica de aumento de datos (Goodfellow et al., 2016), lo cual permite incrementar artificialmente el número de tramas de la base de datos. Se aplicaron como aumento de datos las siguientes operaciones: recortes en un rango de 0.2% al 20% de su tamaño original, escalamiento de la trama original aplicando una amplificación o reducción de un rango de 0.2% al 20%, un corrimiento de la trama a lo ancho y alto; así como aplicar una rotación o una operación de espejo. El resultado de estas modificaciones se ilustra en la Fig. 7. Los parámetros que se usaron para el aumento de datos son seleccionados considerando la situación natural de captura de rostro del conductor por cámaras de dispositivo móvil colocado en la parte frontal de vehículo.

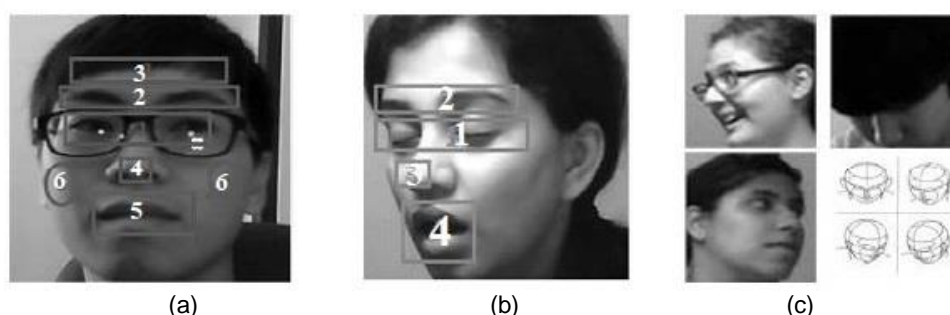


Fig. 7: Patrones usados para etiquetar las tramas contenidas en la base de datos NTHU-DDD (Weng et al., 2017). (a) rostro sin somnolencia, (b) rostro con signos de somnolencia y (c) conductor distraído.

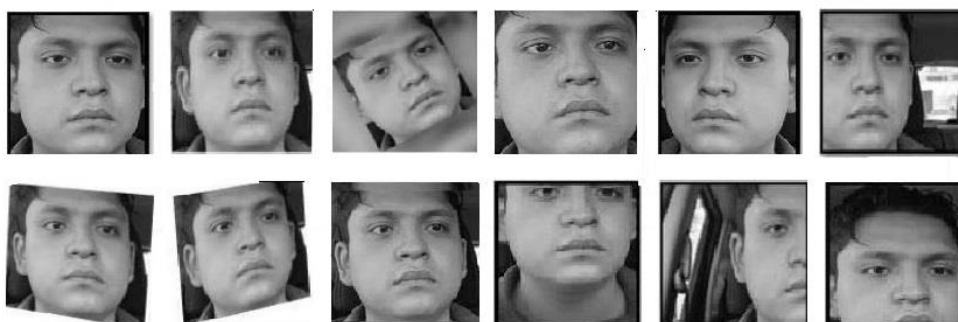


Fig. 8: Ejemplo de las modificaciones introducidas a las imágenes originales para incrementar el número de imágenes disponibles para el entrenamiento de un sistema de aprendizaje profundo.

Activación y desactivación de alarma

La detección de somnolencia y distracción se lleva a cabo mediante las etapas de resultados y análisis mostrados en la Fig. 1. Aquí una vez analizada la trama de entrada, la red neuronal detecta cualquiera de los tres estados descritos anteriormente: normal, somnolencia y distracción. La alarma se activa cuando el estado de somnolencia continua más de un valor umbral th_1 segundos o el estado de distracción del conductor

continúa más de th_2 segundos. Los valores de umbrales se determinaron tal que th_1 es igual a 1.5s y th_2 es igual a 3s, ya que la duración de ojos cerrados menor que 1.5s se considera un parpadeo normal de un adulto, según investigación reportada por Kwon et al. (2013). Así mismo si la duración de distracción rebaza 3s se considera una conducción peligrosa, tomando en cuenta que si la velocidad de vehículo es 80km por hora, el vehículo avanza más de 50 metros sin atención de conductor.

El sistema propuesto cuenta el número de tramas consecutivas clasificadas como somnolencia o distracción como salida de las redes neuronales y convierte este número de tramas en la duración de tiempo para determinar la activación de la alarma. Por ejemplo, teléfono celular "Samsung s20 fe", con 6GB de memoria RAM, se puede operar el sistema propuesto con 14 fps, entonces, cuando el estado de somnolencia en el conductor dura más de 21 (14×1.5) tramas consecutivas, la alarma se activa alertando la somnolencia en el conductor. Mientras cuando el sistema detecta 42 (14×3) tramas consecutivas, se determina el estado de distracción, la alarma se activa, alertando la distracción del conductor. Una vez que el conductor corrija su estado peligroso, la alarma se desactiva y el contador de tramas consecutivas se inicializa. El proceso de activación y desactivación de la alarma se ilustra en la Fig. 9.

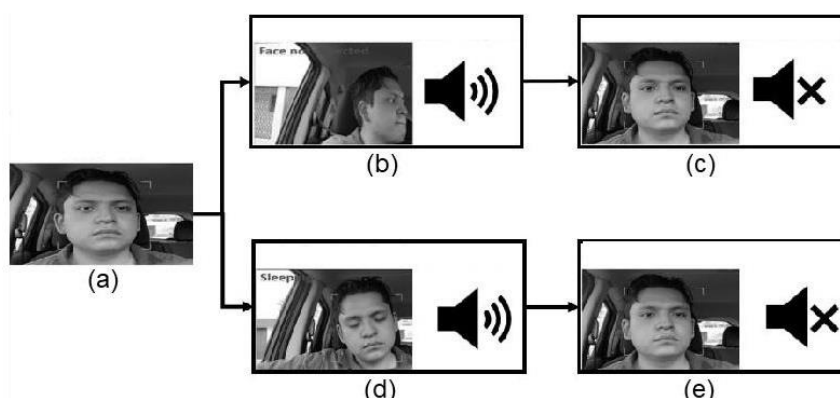


Fig. 9: Activación y desactivación de la alarma cuando conductor muestra distracción o somnolencia. (b) y (d) Activación; (c) y (e) desactivación.

Una vez que el sistema activa o desactiva la alarma, la imagen del rostro del conductor se regresa a su tamaño original usando interpolación bicúbica con el fin de mostrar la imagen del rostro del conductor en la pantalla del dispositivo móvil. El proceso completo de reducción e incremento del tamaño del rostro del conductor durante el proceso de detección de somnolencia o distracción se muestra en la Fig. 10.



Fig. 10: Proceso de incremento y decremento del tamaño de la imagen durante el proceso de reconocimiento. (a) y (b) Imagen original, (c) imagen reducida, (d) rostro detectado y clasificado en uno de tres estados, (e) imagen interpolada para obtener el tamaño original.

RESULTADOS Y DISCUSIÓN

Como se mencionó anteriormente, la base de datos que usamos para entrenar y probar el sistema propuesto es la base de datos NTHU-DDD (Weng et al., 2017). Esta base de datos contiene videos bajo diferentes condiciones de iluminación: iluminación normal que simula la conducción durante el día e iluminación infrarroja que simula la conducción durante la noche. La Fig. 11 ilustra la funcionalidad del sistema propuesto operando con iluminación normal e infrarroja usando el teléfono celular instalado dentro de vehículo. Adicionalmente, para determinar el funcionamiento del sistema en condiciones reales se llevaron a cabo varias pruebas, en las cuales el sistema se implementó en varios dispositivos móviles cuyos resultados se muestran posteriormente.

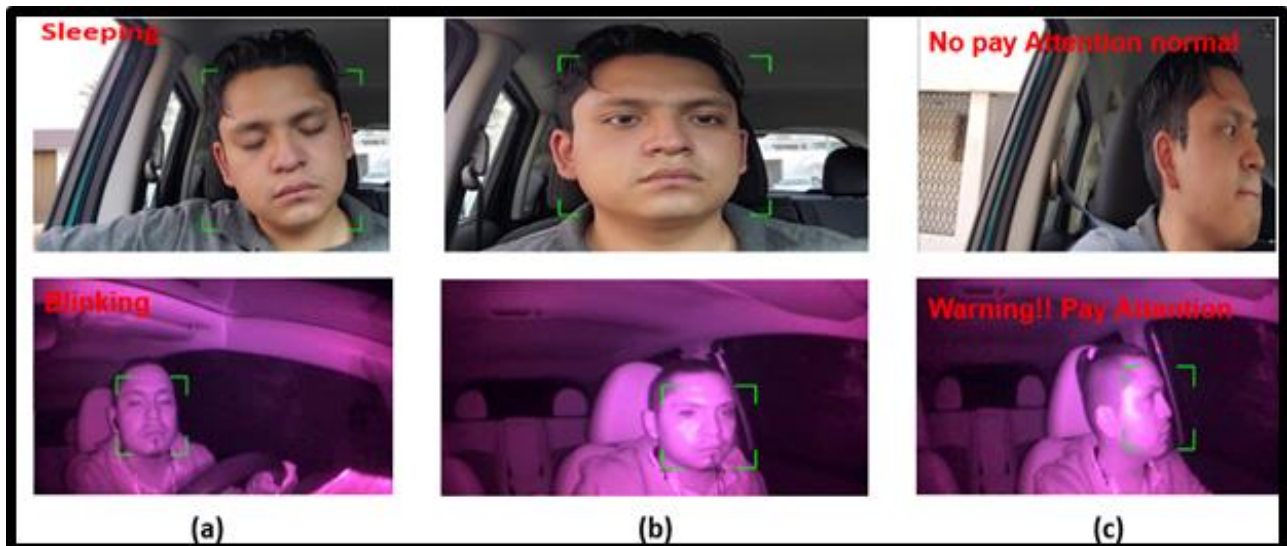


Fig. 11: Funcionalidad del sistema propuesto operando con luz visible e infrarroja. En la parte superior se muestran los tres estados analizados (a) somnolencia, (b) normal y (c) distracción. Mientras que en la parte inferior se muestran los mismos estados con iluminación infrarroja.

Funcionamiento del sistema propuesto en estación de trabajo

Para analizar el funcionamiento del sistema usando una estación de trabajo sin limitación de recursos se muestra en la Tabla 3, en la cual se muestran la exactitud, precisión, “recall” e “f1 score” proporcionadas por el sistema propuesto bajo cada una de las condiciones mencionadas anteriormente: a) sujetos con lentes bajo iluminación normal, b) Sujetos sin lentes bajo iluminación normal, c) sujetos con lentes son iluminación infrarroja y d) sujetos sin lentes bajo iluminación infrarroja.

Tabla 3: Funcionamiento del sistema propuesto

Condición	Exactitud (%)	Precisión (%)	Recall (%)	F1 score
Sin lentes bajo iluminación normal	96.67%	96.70%	97.11%	0.969
Con lentes bajo iluminación normal	95.83%	95.87%	95.83%	0.959
Sin lentes bajo iluminación infrarroja	95.51%	95.60%	95.92%	0.959
Con lentes bajo iluminación infrarroja	94.67%	94.36%	94.33%	0.943
Funcionamiento global	95.77%	95.63%	95.80%	0.957

De la Tabla 3. Se observa que el sistema propuesto proporciona un mejor funcionamiento cuando el conductor opera con iluminación sin lentes, mientras que el menor funcionamiento se obtiene cuando el conductor usa lentes y maneja de noche, esto es el sistema captura la imagen del conductor mediante una cámara infrarroja. Sin embargo, la diferencia entre los resultados obtenidos bajo las mejores condiciones y los resultados obtenidos bajo las condicione más desfavorable son de aproximadamente 2% en todas las métricas. Por su parte la exactitud, Precisión, “Recall” y “F1 score” global son de 95.77%, 95.63%, 95.80% y 0.957, respectivamente.

Uno de los principales objetivos de esta investigación es obtener un monitoreo en tiempo real con una alta precisión cuando el sistema es implementado en dispositivos móviles con limitado poder computacional. Sin embargo, aunque este proporciona un método que puede resolver los problemas de retardo que dificultan su implementación en tiempo real, tiene la desventaja de la pérdida de información introducida durante el proceso de reducción del tamaño de la trama de entrada mostrado en la Tabla 2, cuando la trama de entrada se reduce a 1/9 de su tamaño original. Aquí más del 50% de la información original es perdida. Por lo tanto, el uso de algoritmos de interpolación que contribuyan a recuperar al menos parte de la información original es esencial. Para ilustrar lo anterior, en la Table 4 se presenta una comparación entre la implementación del sistema propuesto, en dispositivos sin limitaciones computacionales, y su realización en dispositivos móviles con poder computacional limitado.

Como se muestra en la Tabla 4, cuando se usa un dispositivo móvil, hay un decrecimiento en la precisión, el número de fps que se deben procesar y el almacenamiento requerido. Por otro lado, las pruebas que se realizaron para determinar el funcionamiento del sistema propuesto, fue realizada en un equipo de alto rendimiento y un dispositivo móvil. Las características del equipo de alto rendimiento consisten en un 3070TI GPU, 8GB Memoria RAM, 6144 núcleos, frecuencia de reloj de 1,770MHz, con 32 GB de RAM. Por su parte, el equipo móvil fue un smartphone Samsung s20 fe, 32 MP, con cámara frontal y 6GB de memoria RAM. Un punto importante para mencionar es que la arquitectura usada de la CNN es la misma entre dos implementaciones, equipo de alto rendimiento y el teléfono celular. Esto significa que una vez optimizada en el proceso de adquisición y preprocesamiento de tramas para su implementación en dispositivos móviles, ejecuta la misma CNN entrenada. Aunque, como se puede observar en la Tabla 4, hay una pérdida de exactitud de menos del 12% en promedio cuando se implementa el algoritmo propuesto en dispositivos móviles, consideramos que es costo aceptable si tomamos en cuenta las mejoras que se han obtenido con el sistema propuesto las cuales se mencionan a continuación: a) El sistema propuesto es capaz de clasificar los estados del conductor en una forma muy aceptable, considerando que más del 90% de la información fue removida. b) La velocidad de reconocimiento se incrementa en cerca del 300% de velocidad inicial. c) El modelo final tiene un tamaño, en términos de memoria, poco menor a un tercio de su tamaño original.

Tabla 4: Funcionamiento del sistema propuesto en dispositivos con capacidad limitada y dispositivos móviles

Características	Sistema propuesto en equipo de alto rendimiento	Sistema propuesto para dispositivo móvil
Exactitud	95.77	84.45
Almacenamiento requerido (MB)	3.98	1.30
Velocidad de procesamiento (fps)	21	13-15

La Tabla 5 compara el sistema propuesto en este artículo con otros sistemas propuestos en la literatura (Jabbar et al., 2018; Polumpung et al., 2022; Yu et al., 2019; Shakeel et al., 2021; Wijnands et al. 2020), los cuales se tratan la implementación de detección de somnolencia en dispositivos móviles, cuyo detalle se explican en la sección de trabajos relacionados. Cabe mencionar que no se realizó la comparación con aquellos sistemas implementados en computadoras de alto rendimiento sin límite de recursos ni el computo en la nube, ya que en la nube están instalados servidores con alto rendimiento. Como se puede observar de la Tabla 5, el sistema propuesto tiene un buen balance entre la exactitud de detección y la velocidad de operación que garantiza el proceso en tiempo real. Adicionalmente, el sistema propuesto también detecta distracción de conductor, que es otra causa importante de accidente automotriz.

Tabla 5: Tabla comparativa entre el sistema propuesto y otros previamente propuestos para dispositivos móviles. La sigla '---' significa que autores no proporcionaron los datos.

Autores	Exactitud (%)	Velocidad de operación (fps)	Objetivo
Propuesto	84.44	14-15	Somnolencia / distracción
Jabbar et al. (2018)	81.00	7	Somnolencia
Polumpung et al. (2022)	78.81	---	Somnolencia
Yu et al. (2019)	97.80	6	Somnolencia
Shakeel et al. (2021)	83.70	12-16	Somnolencia
Wijnands et al. (2020)	73.9	10	Somnolencia

La Tabla 6 presenta los resultados obtenidos cuando el sistema propuesto es implementado en diversos dispositivos los cuales han estado en el mercado durante algún tiempo. Un punto importante que considerar es que la velocidad de reconocimiento no depende de la cantidad de memoria RAM de que dispone, ya que depende más de la resolución de la imagen que este maneja ya que de esto depende la velocidad de procesamiento medido en fps. Como se puede observar de la Tabla, la mayoría de los dispositivos móviles pueden procesar el sistema propuesto en tiempo real, ya que el número de tramas por segundo es mayor o cercano a 10 fps. Finalmente, el funcionamiento del sistema instalado en un teléfono celular se puede observar de la siguiente liga: https://youtu.be/3GeBGMi_pCw

Tabla 6: Dispositivos en los cuales está corriendo el sistema

Dispositivo móvil	RAM (GB)	Versión de Android	fps
S. S20 FE	6	12	14
S. A51	4	10	9
M. One	4	8.1	21
M. G20	4	11	9
X. MI 9t Pro	6	9	19
X. POCO F4	6	10	17
X. Redmi Note Pro	6	9	15
H. Nova 5T	6	9	12
H. Y9A	6	10	12
H. P30	6	6	9
Black A55	3	11	14

CONCLUSIONES

Este artículo propone un sistema para la detección de somnolencia y distracción en un conductor, junto con su implementación en un dispositivo móvil compacto que se puede colocar dentro de cualquier vehículo sin requerir algún equipo adicional. El principal objetivo del trabajo es implementar el sistema basado en CNN en dispositivos móviles con la capacidad de cómputo limitado y el algoritmo de detección opere en tiempo real para advertir al conductor justo a tiempo, cuando el conductor se encuentra dormido o distraído, para evitar un lamentable accidente.

Con la finalidad de alcanzar el objetivo, las tramas tomadas por el dispositivo móvil deben reducirse en su tamaño sin que se pierda información importante para determinar el estado del conductor. Después de usar diferentes tasas de reducción, midiendo la velocidad de operación y la exactitud de detección, se determinó la tasa de reducción de 1/9 del tamaño original de las tramas como el valor óptimo. Esta tasa de reducción ofrece un equilibrio entre la velocidad de operación y la exactitud de detección. El sistema propuesto proporciona una exactitud del 85%, con una velocidad de procesamiento de 15 fps, lo cual permite que la alarma se active con la suficiente rapidez para evitar accidentes debidos a distracción o somnolencia del conductor. El sistema propuesto se comparó con otros trabajos reportados recientemente y se muestra mejor equilibrio entre la velocidad de operación y la exactitud de detección, comparando con otros trabajos.

El funcionamiento del sistema se ha probado usando diferentes dispositivos móviles que se encuentra en mercado bajo diferentes condiciones de luminosidad: luz de día y luz de noche, mostrando un buen desempeño. El sistema propuesto detecta distracción de conductor además de somnolencia en el conductor, considerando que la distracción es otra causa importante de accidentes, el sistema ofrece mejor servicio al conductor.

REFERENCIAS

- Adil, K., Nawaz, M., y otros tres autores, IoT-based non-intrusive automated driver drowsiness monitoring framework for logistics and public transport applications to enhance road safety, <https://doi.org/10.1109/ACCESS.2023.3244008>, IEEE Access, 11, 14385-14397 (2023)
- Chacon-Murguía, M., y Prieto-Resendiz, C., Detecting driver drowsiness: a survey of system designs and technology, <https://doi.org/10.1109/MCE.2015.2463373>, IEEE Consum. Electron. Mag., 4(4), 107-119 (2015)
- Comité Nacional para Seguridad (CNS), Accidentes y sus causas, http://www.cns.gob.mx/portalWebApp/appmanager/portal/desk?_nfpb=true&_pageLabel=portals_portal_page_m2p1p2&content_id=830068&folderNode=830052&folderNode1=810277 (2019).
- Facts and Stats, National Sleep Foundation, <http://drowsydriving.org> (2022)
- Fierro, A.N., Nakano, M., Yanai, K., y Perez, H.M., Redes Convolucionales Siamesas y Tripletas para la Recuperación de Imágenes Similares en Contenido, <http://dx.doi.org/10.4067/S0718-07642019000600243>, Información Tecnológica, 30(6), 243-254 (2019)

- Flores-Monroy, J., Nakano-Miyatake, M., Perez-Meana, H., y Sanchez-Perez, G., Visual-based real time driver drowsiness detection system using CNN, <https://doi.10.1109/CCE53527.2021.9633082>, Proceedings of International Conference on Electrical Engineering, Computing Science and Automatic Control, IEEE, 1-5, Mexico City, Mexico (2021)
- Goodfellow, I., Bengio, Y., y Courville, A., Deep learning, MIT Press, Cambridge Mass, Estados Unidos de America (2016)
- INEGI, Causes of traffic accidents in Mexico (Spanish), by the National Institute of Statistics, Geographics and Informatics in Mexico, <https://www.inegi.org.mx> (2022)
- Jabbar, R., Al-Khalifa, K., y otros cuatro autores, Real-time Driver Drowsiness Detection for Android Application Using Deep Neural Networks Techniques, <https://doi.10.1016/j.procs.2018.04.060>, Procedia Comput. Sci., 400–407 (2018)
- Kokonozi, A., Michail, E., Chouvarda, I., y Maglaveras, N., A study of heart rate and brain system complexity and their interaction in sleep-deprived subjects, <https://doi.10.1109/CIC.2008.4749205>, Proceeding on Computational Cardiology, 969–971 (2008)
- Kwon, K.A., Shipley, R., y otros cinco autores, High-speed camera characterization of voluntary eye blinking kinematics, <https://doi.10.1098/rsif.2013.0227>, J. of the Royal Society Interface, 10(85), 1-6 (2013)
- Magán, E., Paz-Sesmero, M., Alonso-Weber, J.M., y Sanchis, A., Driver Drowsiness Detection by Applying Deep Learning Techniques to Sequences of Images, <https://doi.10.3390/app12031145>, Applied Sciences, Switzerland, 1-25 (2022)
- Phan, A., Nguyen, N., Trieu, T., y Phan, T., An efficient approach for detecting driver drowsiness based on deep learning, <https://doi.10.1109/ICCV.2007.4409068>, Applied Science, 11, 8441 (2001)
- Polumpung, A., Angeline, L., y otros cinco autores, Computing non-contactable drowsiness monitoring system with mobile machine vision, <https://doi.10.1109/IICAET55139.2022.9936760>, Proc. IICAIE, 1-6 (2022)
- Shaily, S., Krishnan, S., Natarajan, S., y Periyasamy, S., Smart driver monitoring system, <https://doi.10.1007/s11042-021-10877-1>, Multimedia Tools and Applications, 80, 25633-25648 (2021)
- Shakeel, M.F., Bajwa, N.A., y otros cuatro autores, Detecting driver drowsiness in real time through deep learning-based object detection, En Rojas, I., Joya, G., Catala, A. (eds). Advances in Computational Intelligence. IWANN 2019, https://doi.10.1007/978-3-030-20521-8_24, Lecture Notes in Computer Science, 11506, 283-296, Springer, Cham (2019)
- Tamanani, R., Muresan, R., y Al-Dweik, A., Estimation of driver vigilance status using real-time facial expression and deep learning, <https://doi.10.1109/CVPR.2018.00474>, IEEE Sensors Letters, 5(5), (2021)
- Uma, S., y Eswari, R., Accident prevention and safety assistance using IOT and machine learning, <https://doi.10.1007/s40860-021-00136-3>, Journal of Reliable Intelligent Environments, 1-25 (2021)
- Vidente, V., Laguna, P., Bartra, A., y Bailón, R., Drowsiness detection using heart rate variability, <https://doi.10.1007/s11517-015-1448-7>, Med. Biol. Eng. Comput., 54(6), 827-937 (2016)
- Viola, P., y Jones, M., Robust real-time face detection, <https://doi.10.1023/B:VISI.0000013087.49260.fb>, Int. Journal of Computing Vision, 57(2), 137-154 (2004)
- Weng, C.H., Lai, Y.H., y Lai, S.H., Driver drowsiness detection via a hierarchical temporal deep belief network, https://doi.10.1007/978-3-319-54526-4_9, Proc Asian Conference on Computer Vision, 117-133 (2017)
- Wijnands, J.S., Thompson, J., y Nice, K.A., Real-time monitoring of driver drowsiness on mobile platforms using 3D neural networks, <https://doi.org/10.1007/s00521-019-04506-0>, Neural Comput & Applic., 32, 9731–9743 (2020)
- Wu, B., Chen, Y., Yeh, C., y Li, Y., Reasoning-based framework for driving safety monitoring using driving event recognition, <https://doi.10.1109/TITS.2013.2257759>, IEEE Trans. Intell. Transp. Syst., 14(3), 1231–1241 (2013)
- Yu, C., Qin X., y otros tres autores, DrowsyDet: A mobile application for real-time driver drowsiness detection, <https://doi.10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00116>, Proc. IEEE SmartWorld, 425-432 (2019)
- Zhang, C., Wang, H., y Fu, R., Automated detection of driver fatigue based on entropy and complexity measures, <https://doi.10.1109/TITS.2013.2275192>, IEEE Trans. Intell. Transp. Syst., 15(1), 168–177 (2014)