REVISTA CUBANA
DE TRANSFORMACIÓN DIGITAL

RNPS 2487 • ISSN 2708-3411
Vol. 2 • Nro. 1 • enero-marzo 2021 • págs. 46-65

COMUNICACIONES BREVES

Exploración de Redes Neuronales Holográficas con Cuantificación Difusa para la Monitoreo de Conductores en Conducción Autónoma Condicional

Exploring Fuzzy-Quantized Holographic Neural Networks for Driver Monitoring in Conditional Driving Automation

Luis Ariel Diago Marquez
luis_diago@meiji.ac.jp • http://orcid.org/0000-0002-8799-5834

Hiroe Abe
h_abe@meiji.ac.jp

Kana Adachi
kanaadachi91@gmail.com

Ichiro Hagiwara
ihagi@meiji.ac.jp
MEIJI UNIVERSITY, JAPÓN

Recibido: 2020-12-01 • Aceptado: 2021-01-20

RESUMEN

La Sociedad de Ingenieros Automotrices (SAE, siglas en inglés) define los sistemas de conducción automatizados (ADS, siglas en inglés) para vehículos de carretera como aquellos que pueden realizar toda la tarea de conducción dinámica sin un conductor humano en el lazo de control. Bajo la conducción autónoma condicional (SAE Nivel 3), cuando la conducción automatizada falla, se espera que los conductores reanuden la conducción manual. Para que esta transición se produzca de forma segura es imperativo que los conductores reaccionen de forma adecuada y oportuna, lo que es difícil que suceda una vez que el conductor ha sido sometido a largas distancias de conducción autónoma. Las técnicas de inteligencia artificial (IA) podrían utilizarse para garantizar la seguridad de los sistemas adaptativos de seguridad crítica. No solo al observar el entorno exterior del vehículo, sino también al monitorear el estado de la comunicación conductor-vehículo. Además, en este contexto el concepto de IA explicable tiene potencial para proporcionar evidencia de que un ADS podría respaldar la

garantía de seguridad y el cumplimiento normativo. Este trabajo presenta un método neuro-difuso que funciona como un enfoque de aprendizaje automático explicable adecuado para dominios en los que se requiere la validación de los modelos de predicción subyacentes. Los resultados de la comparación entre el modelo propuesto y otros modelos de la literatura muestran que el modelo propuesto podría proporcionar explicaciones sobre sus predicciones en tiempo real para garantizar transiciones fluidas en el nivel 3 de SAE.

Palabras clave: conducción autónoma condicional; inteligencia artificial explicable; monitoreo del conductor; redes neuronales holográficas con cuantificación difusa.

ABSTRACT

The Society of Automotive Engineers (SAE) defines Automated Driving Systems (ADS) for road vehicles as being that can perform the entire dynamic driving task without a human driver in the loop. Under conditional driving automation (SAE Level 3), when automated driving fails the drivers are expected to resume manual driving. For this transition to occur safely, it is imperative that drivers react in an appropriate and timely manner, which is difficult to happen once the driver has been subjected to long distances of autonomous driving. Artificial Intelligence (AI) techniques could be used for safety assurance of adaptive safety-critical systems. Not only sensing the external environment of the vehicle, but also monitoring the state of the driver-vehicle communication. Further, the concept of explainable AI was highlighted as having potential to provide evidence from ADS that could support safety assurance and regulatory compliance. This work presents a neuro-fuzzy method working as an explainable machine learning approach suitable for domains where validation of the underlying non-linear prediction models is required. The results of comparison between proposed model and other models from the literature show that the proposed model could provide explanations about its predictions in real time to ensure smooth transitions in SAE Level 3.

Keywords: *Conditional driving automation; Explainable artificial intelligence; Driver monitoring; Fuzzy-Quantized Holographic Neural Networks.*

INTRODUCCIÓN

El objetivo general de la investigación sobre vehículos autónomos es el desarrollo de sistemas totalmente automatizados capaces de conducir en cualquier escenario de tráfico. Los ocupan-

tes de tal vehículo serían entonces meros pasajeros, sin acceso a los controles. Sin embargo, para desarrollar de manera segura la tecnología para lograr este objetivo, es necesario que haya un control compartido entre el vehículo y un conductor humano. El rol de cada uno se define dentro de cada uno de los 5 niveles de automatización de vehículos definidos por la Sociedad de Ingenieros Automotrices (SAE, 2018). En los niveles 1 y 2, el vehículo cuenta con algún sistema de automatización de la conducción, ya sea para el control del movimiento longitudinal y/o lateral, pero la figura del conductor humano sigue presente para realizar las tareas de conducción dinámica. Los vehículos condicionalmente autónomos (nivel 3) pueden operar de manera autónoma en escenarios de tráfico específicos, pero se espera que un ocupante humano, detrás del volante, controle el sistema automatizado y esté preparado para las solicitudes de toma de control. Finalmente, para los niveles 4 y 5 desaparece la figura del conductor y el propio sistema de automatización de la conducción cuenta con un sistema de respaldo para actuar en caso de fallo del sistema principal y poder conducir hasta una situación de riesgo mínimo. Solo en el nivel 3 el control debe transferirse del vehículo al ser humano durante los modos de falla del sistema. Por lo tanto, la estimación continua de la disponibilidad de este ocupante para hacerse cargo es fundamental para una transferencia de control segura y oportuna. En el resto de este documento, se usa el término "conductor" en el contexto de vehículos condicionalmente autónomos para referir al ocupante responsable de tomar el control del vehículo.

Una revisión de la literatura muestra que ya se ha abordado el problema estrechamente relacionado de estimar la distracción del conductor en condiciones de conducción manual (Khan y Lee, 2019). La distracción del conductor se ha definido como el desvío de la atención del conductor de las actividades críticas para la conducción segura hacia una segunda actividad, que puede resultar en una atención insuficiente o nula a las actividades críticas para una conducción segura (Ballingall, Sarvi, y Sweatman, 2020). En el nivel 3 aumenta la posibilidad de que los conductores participen en actividades secundarias no realizadas antes, durante la conducción manual, así como la posibilidad de participar más libremente en actividades secundarias (en inglés, *non-driving related tasks* o NDRT) como atender el teléfono o leer un libro. Si bien se han propuesto sofisticados algoritmos de visión por computadora para el análisis de la actividad del conductor (Khan y Lee, 2019), relativamente pocos trabajos (Deo y Trivedi 2019; Braunagel, Rosenstiel y Kasneci 2017) han abordado el problema de mapear la actividad del conductor con la preparación para la toma de control. Esto podría atribuirse a dos desafíos principales. Primero, hay una falta de conjuntos de datos de conducción naturalistas que observen la actividad del conductor en vehículos condicionalmente autónomos. En segundo lugar, definir la verdad absoluta (en inglés, *ground truth)* acerca de la preparación para tomar el control es una tarea desafiante. Los enfoques basados en datos dependen de la disponibilidad de datos reales de lo que se estima. Un conjunto completo de datos de conducción naturalista que capture una amplia gama de comportamientos de los conductores permitiría enfoques basados en datos para mapear la actividad del conductor con su preparación para tomar el control. Sin embargo, ante la dificultad para obtener datos de conducción con-

dicional naturalista y por la naturaleza de la tarea, Braunagel y colaboradores (2017) definen la preparación para tomar el control a partir del tiempo de toma de control y la calidad de la toma de control en ensayos experimentales con solicitudes de toma de control emitidas a conductores que realizan actividades secundarias en entornos de simulación.

En este artículo, se propone un enfoque basado en datos para estimar la preparación de los conductores en vehículos condicionalmente autónomos para tomar el control, basado puramente en las salidas de sensores no invasivos (p.ej. sensores de visión) orientados hacia el conductor en un entorno de simulación donde es posible medir variables fisiológicas del conductor durante la marcha del vehículo. Para encontrar una relación entre la información que se obtiene de los sensores de visión y la condición interna del conductor es necesario estudiar la condición del conductor a partir de sus variables fisiológicas. Como no se encuentra claro cuál de todas las variables fisiológicas tiene una mayor correlación con los datos obtenidos de los sensores de visión, es necesario estudiarlas todas para establecer la verdad absoluta.

Trabajos relacionados

En trabajos recientes (Deo y Trivedi, 2019) se comenta, que si bien los sensores de electroencefalograma (EEG) permiten la representación más fiel de la actividad cerebral del conductor, (Khan y Lee, 2019) son demasiado invasivos para ser viables en vehículos comerciales. Es por esto que los autores utilizan evaluaciones subjetivas en una escala de 1 (nivel bajo) a 5 (nivel alto) de 7 evaluadores que al observar videoclips de los conductores durante la marcha de un vehículo autónomo califican el nivel de disposición del conductor para asumir el control. Finalmente, obtienen un índice subjetivo para evaluar el nivel de disposición del conductor para asumir el control a partir del promedio de las evaluaciones normalizadas de 260 videoclips de 30s cada uno dividiéndolos en segmentos de 2s para su evaluación. Desafortunadamente, las personas no siempre pueden dar calificaciones subjetivas de manera consistente. Como muestra el índice de correlación (en inglés, ICC: *Interclass Correlation Coefficient)* empleado en el trabajo de Deo y Trivedi (2019), el nivel de coincidencia entre los evaluadores puede ser moderado (0,5 < ICC< 0,75) y en ocasiones pobre (< 0,5). En este trabajo se usa una representación más fiel de la actividad cerebral del conductor a partir de los valores de EEG. En el trabajo anterior (Deo y Trivedi, 2019), también se empleó un modelo de red neuronal recurrente del tipo LSTM (en inglés, *Long Short-Term Memory)* para obtener la dependencia temporal de las representaciones por tramas y predecir continuamente el nivel de disposición del conductor a partir del índice subjetivo creado. En lugar de describir el estado del conductor con un índice continuo, en este trabajo se divide el estado de atención del conductor en clases (por ejemplo, 1-baja, 2-media y 3-alta) con el objetivo de extraer reglas que permitan argumentar el estado asignado al conductor por modelos computacionales de tipo neuro-difuso (Diago, Kitaoka, Hagiwara y Kambayashi, 2011a). Estos modelos funcionan con un enfoque de aprendizaje automático explicable adecuado para dominios donde se requiere la validación de los modelos de predicción subyacentes.

Para desarrollar los vehículos comerciales pudiera pensarse en crear un modelo predictivo independiente del sujeto entrenado con una base de datos que incluya muchos individuos

en diferentes condiciones como se ha realizado en estudios anteriores (Deo y Trivedi, 2019; Braunagel, *et al.,* 2017). Sin embargo, estos modelos tienen un comportamiento inferior al esperado cuando se llevan a la práctica, debido a que la cantidad de muestras para el aprendizaje es muy pequeña. En Braunagel y colaboradores (2017), al igual que en este trabajo, los experimentos duran alrededor de 30 min (5 min de conducción manual, 3 min de conducción automática sin intervención, 5 min de una prueba condicional en conducción recta y dos experimentos de toma de control en dos escenarios diferentes de 8 min cada uno). Los autores mencionan que para entrenar el clasificador tuvieron que equilibrar las clases reduciendo aún más la ya baja cantidad de datos de aprendizaje. Por lo tanto, el conjunto de entrenamiento aplicado es equilibrado e independiente del sujeto, pero contiene relativamente pocas situaciones de toma de control. Estudios recientes con conductores de diferentes edades (Wu, *et al.,* 2020) muestran que para conductores más jóvenes, realizar NDRT puede contrarrestar la somnolencia del conductor y no siempre puede tener efectos negativos en el rendimiento después de recibir la señal de RtI. Sin embargo, debido a que los conductores mayores (posiblemente incluidos los de mediana edad) ya son menos propensos a la somnolencia y son más vulnerables a la carga de trabajo mental inducida por las NDRT, parece inapropiado animar a los conductores de edad avanzada a participar en las NDRT cuando se necesita que respondan a una señal de RtI. Roche, Somieski, y Brandenburg (2019) demuestran que los conductores cambian su comportamiento ante peticiones repetidas de toma de control. Es por eso que la tendencia actual es a desarrollar sistemas de aprendizaje en línea que se puedan adaptar a diferentes tipos de conductores (Wu, *et al.,* 2020; Roche, *et al.,* 2019).

El objetivo del presente estudio es explorar el uso de los modelos neuro-difusos anteriores para el monitoreo del conductor en la conducción condicional y mostrar no solo que se obtiene un modelo explicable, sino que el modelo propuesto es más rápido que el modelo LSTM manteniendo similar exactitud en la predicción del estado del conductor. A continuación se muestra el enfoque propuesto dentro de la metodología y se discuten los principales resultados. Finalmente se muestran las conclusiones preliminares y se presentan algunas líneas de investigación en las que se trabaja actualmente.

METODOLOGÍA

El enfoque propuesto es similar a los presentados en Deo y Trivedi (2019) y Braunagel y colaboradores (2017), en el que se emplean técnicas de aprendizaje para predecir el nivel de preparación para la toma de control de un conductor a partir de sensores de visión. Deo y Trivedi (2019) modelan el problema como un problema de regresión y Braunagel y otros (2017) presentan el problema como un problema de clasificación. En este trabajo se sigue la propuesta de Braunagel, otros (2017), en el que se entrena un clasificador para reconocer la preparación del conductor para la toma de control a partir de la calidad de sus intervenciones (baja, media o alta) después de recibir la señal de RtI. La figura 1 resume el enfoque propuesto cuyas principales contribuciones son las siguientes:

1. Registros de datos de múltiples variables fisiológicas de conductores de vehículos condicionalmente autónomos en un simulador: Se recopila un conjunto de datos de 10-20 min de conductores al volante de un vehículo autónomo simulado. Esto se captura mediante el empleo de una cámara que observa al conductor y sensores de múltiples variables fisiológicas (electrocardiograma o ECG, electroencefalograma o EEG, actividad electrodérmica o EDA, pulso sanguíneo y frecuencia respiratoria) colocados al conductor durante la simulación de la marcha en un vehículo autónomo. Estos datos son usados para entrenar y evaluar los modelos de aprendizaje. Hasta donde se conoce, este es el primer estudio que evalúa la preparación de los conductores para la toma de control que usan un conjunto de datos con estas cinco variables fisiológicas. En este trabajo se reportan los resultados preliminares obtenidos a partir de las señales de EEG.
2. Anotaciones automáticas sobre la preparación para la toma de control: el objetivo de este trabajo es estimar continuamente la preparación de los conductores para la toma de control a partir de sensores de visión. Para probar la viabilidad de este enfoque, se utilizaron los valores de *e-Sense* (atención del conductor) obtenidos a partir de la señal de EEG de dispositivos de la familia *NeuroSky MindWave* (disponibles en http://neurosky.com) como valores enteros de 0 a 100 y se asignaron los niveles de atención (1-baja, 2-media, 3-alta) a partir de los histogramas de los valores de atención durante una ventana de tiempo T de 5 a 20s.
3. Modelo neuro-difuso para estimar la preparación para la toma de control: Se procesaron los flujos de la cámara fotograma a fotograma para extraer los rasgos característicos de las expresiones faciales del conductor a partir de 68 puntos ubicados en las fronteras de la cara (hasta la ceja), la boca, la nariz y los ojos (ver figura 1). Se propone una red neuronal holográfica para modelar la dependencia temporal de las representaciones por tramas y una cuantificación difusa de los niveles de atención que se adapta a las características de cada conductor. Esta combinación permite crear funciones de cuantificación difusa para explicar las causas de las predicciones hechas por el modelo para ese nivel de atención del conductor a partir de los rasgos extraídos de sus expresiones faciales. El modelo emite continuamente el nivel de atención del conductor para tomar el control en función de los T= 10s de actividad anterior.

El estado del arte para la clasificación de estados afectivos y cognitivos a partir de señales de EEG (Appriou, Cichocki, y Lotte, 2020), muestra que aun cuando los estados de los sujetos pueden ser modelados en una escala continua (por ejemplo: modelo circumplejo de las emociones de Russell) habitualmente se trabaja como un problema de clasificación y se usan clasificadores dependientes del sujeto debido a la gran variabilidad entre los mismos. En este trabajo se realiza un estudio comparativo de dos modelos de predicción del estado del conductor: las redes neuronales recurrentes del tipo LSTM y las redes neuronales holográficas con cuantificación difusa (FQHNN por sus siglas en inglés) (Diago, *et al.,* 2011) a partir de datos recolectados en un simulador.

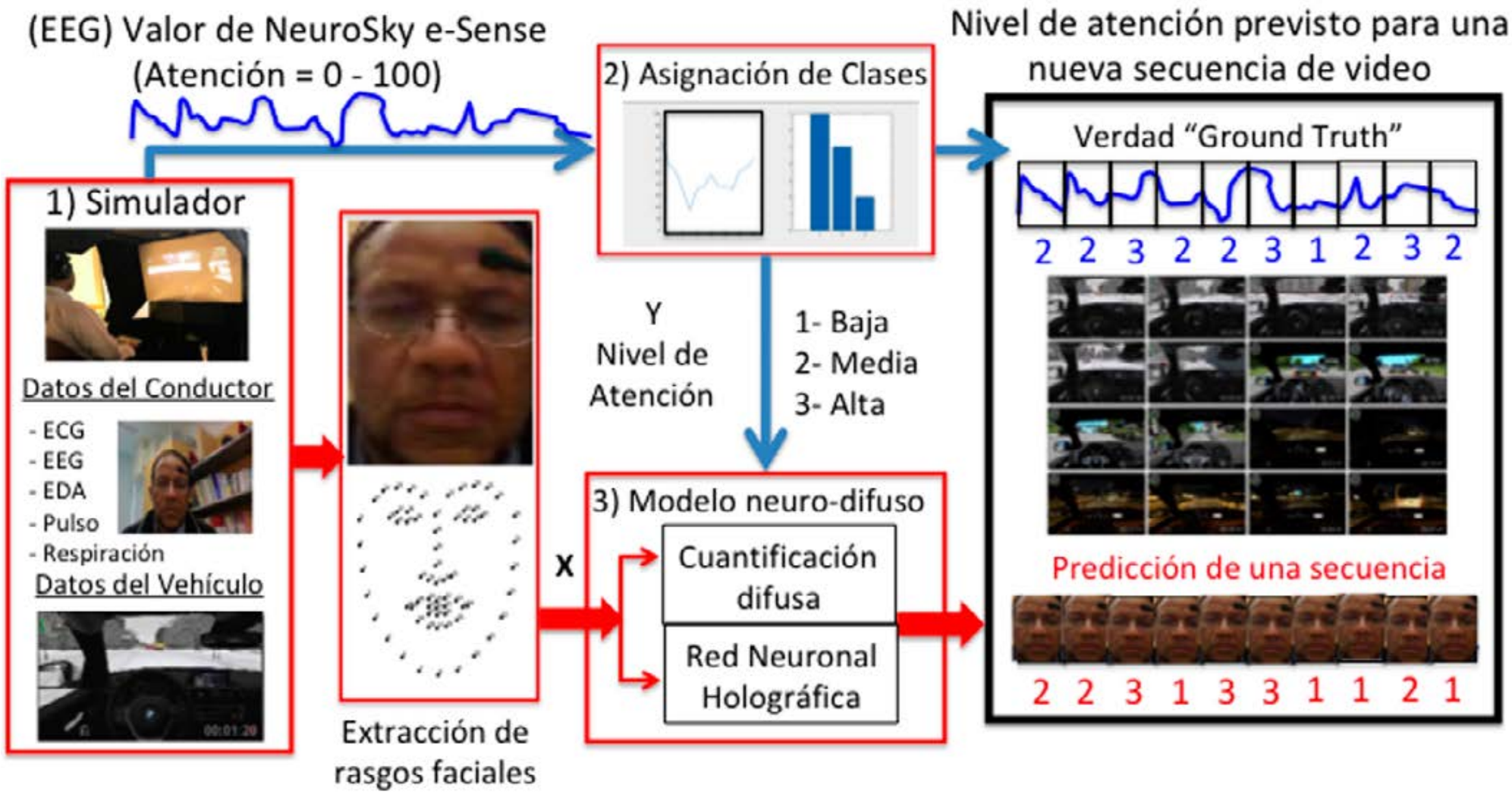


Figura 1. Resumen del enfoque propuesto: 1) A partir de los registros de datos de múltiples variables fisiológicas (electrocardiograma o ECG, electroencefalograma o EEG, actividad electrodérmica o EDA, pulso sanguíneo y frecuencia respiratoria), el video del conductor y los datos del vehículo en distintos escenarios, 2) se asignan automáticamente los niveles de atención del conductor (clases) y 3) se propone un modelo neuro-difuso para predecir el nivel de atención previsto para una nueva secuencia de video.

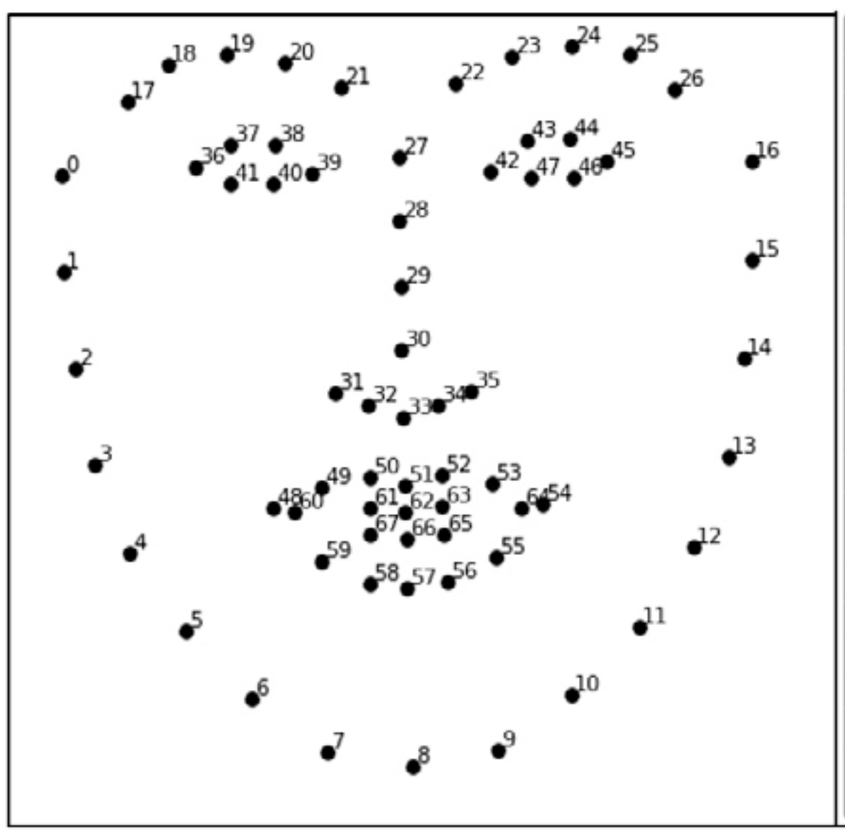
Recolección de datos

Para la recolección de datos se utilizó el marco llamado *"NeuroFaceLab"* que fue introducido en investigaciones anteriores para analizar los estados emocionales de los pasajeros de un vehículo autónomo (Diago, Yang, Abe y Hagiwara, 2018). *NeurofaceLab* ha sido desarrollado como una red *piconet* que permite que un dispositivo maestro se interconecte con hasta siete dispositivos esclavos activos utilizando protocolos de tecnología *Bluetooth.* La aplicación en el dispositivo maestro (la computadora) ha sido desarrollada empleando la biblioteca *DirectShow* para renderizar un video y guardar el video de entrada proveniente de la cámara de video en combinación con las herramientas del *software NeuroSky Inc.* para adquisición de datos en tiempo real de las ondas cerebrales. Toda la información se sincroniza en la aplicación y utiliza el filtro del administrador de gráficos en *DirectShow*. *NeurofaceLab* también incluye la posibilidad de reproducción de video durante la adquisición de datos. Se mostraron videos de entre 10 y 20 min de duración con una resolución de 960 x 540 pixeles a 18 sujetos que participaron en los experimentos. Los videos incluyen tres tipos de escenas de conducción: nevando por la mañana (SM), tarde soleada (SA) y noche oscura (DN). Durante la muestra de los videos se detectaron las expresiones faciales de los conductores y se analizan en este trabajo su relación con los valores de *NeuroSky eSense.*

Expresiones faciales

Existen principalmente dos tipos de enfoques para la extracción de rasgos faciales (Diago, *et al.,* 2011a): métodos basados en rasgos geométricos y métodos basados en apariencia. Los rasgos fa-

ciales geométricos presentan la forma y ubicación de los componentes faciales (incluye boca, ojos, cejas, nariz, entre otros) mientras que en los métodos basados en la apariencia, se aplican filtros de imagen a todo el rostro o región específica en una imagen de rostro para extraer un vector de características. ¿Cuándo utilizar uno u otro enfoque? Es una cuestión que permanece en discusión entre los investigadores. Como la aplicación propuesta se centra en la comprensibilidad de clasificadores no lineales, las expresiones faciales son representadas por vectores de características compuestos por veinte parámetros en la figura 2b que incluyen áreas (p1,..., p4) y distancias (p5,. ., p20) calculados a partir de 68 puntos característicos de la cara numerados en la figura 2a.



AREA (A=Area_del_poligono):

$p_1(Ojo\ derecho) = A(36-41)$
$p_2(Ojo\ izquierdo) = A(42-47)$
$p_3(Boca) = A(48-59)$
$p_4(Cara) = A(17-26) + A(0,17,26,16) - A(0-16)$

DISTANCIA (D=Distancia_euclideana):

$p_5(Cara\ vertical) = \frac{1}{2}D(21,8) + \frac{1}{2}D(22,8)$
$p_6(Cara\ horizontal) = D(16,0)$
$p_7(Ojos) = D(39,42)$
$p_8(Ojo-ceja\ derecha) = \frac{1}{4}D(17,36) + \frac{1}{4}D(18,37) + \frac{1}{4}D(20,38) + \frac{1}{4}D(21,39)$
$p_9(Ojo-ceja\ izquierda) = \frac{1}{4}D(22,42) + \frac{1}{4}D(23,43) + \frac{1}{4}D(25,44) + \frac{1}{4}D(26,45)$
$p_{10}(Ojos-Boca) = Altura\_del\_triangulo(36,45,57)$
$p_{11}(Boca\ vertical) = D(51,57)$
$p_{12}(Boca\ horizontal) = D(48,54)$
$p_{13}(Ojo-nariz\ derecha) = D(36,31)$
$p_{14}(Ojo-nariz\ izquierda) = D(45,35)$
$p_{15}(Izquierda) = D(31,48)$
$p_{16}(Derecha) = D(35,54)$
$p_{17}(Izquierda) = \frac{1}{3}D(1,31) + \frac{1}{3}D(2,31) + \frac{1}{3}D(3,31)$
$p_{18}(Derecha) = \frac{1}{3}D(13,35) + \frac{1}{3}D(14,35) + \frac{1}{3}D(15,35)$
$p_{19}(Arriba) = \frac{1}{2}D(21,33) + \frac{1}{2}D(22,33)$
$p_{20}(Abajo) = \frac{1}{5}D(33,6) + \frac{1}{5}D(33,7) + \frac{1}{5}D(33,8) + \frac{1}{5}D(33,9) + \frac{1}{5}D(33,10)$

Figura 2. Extracción de rasgos característicos de la cara: a) 68 puntos característicos, b) parámetros utilizados.

Valores de *e-Sense*

eSense es un algoritmo patentado de *NeuroSky* para caracterizar estados mentales. Para calcular *eSense,* la tecnología *NeuroSky thinkGear* amplifica la señal de ondas cerebrales sin procesar, y elimina el ruido ambiental y el movimiento muscular. Luego, el algoritmo eSense se aplica a la señal restante y da como resultado los valores interpretados del medidor *eSense,* llamados Atención y Meditación (ATN y MED). ATN y MED de los sujetos en estado de concentración, relajación, fatiga y sueño se han analizado en investigaciones anteriores (He, Liu, Wan y Hu, 2014). Los valores de ATN se dividieron en 3 segmentos: bajos (0-40), medios (41-60) y altos (61-100). A partir de los valores registrados durante una ventana de tiempo T entre 5 y 20s (figura 3a), se asigna el nivel de atención del conductor a partir del rango más probable obtenido a partir del histograma mostrado en la figura 3b.

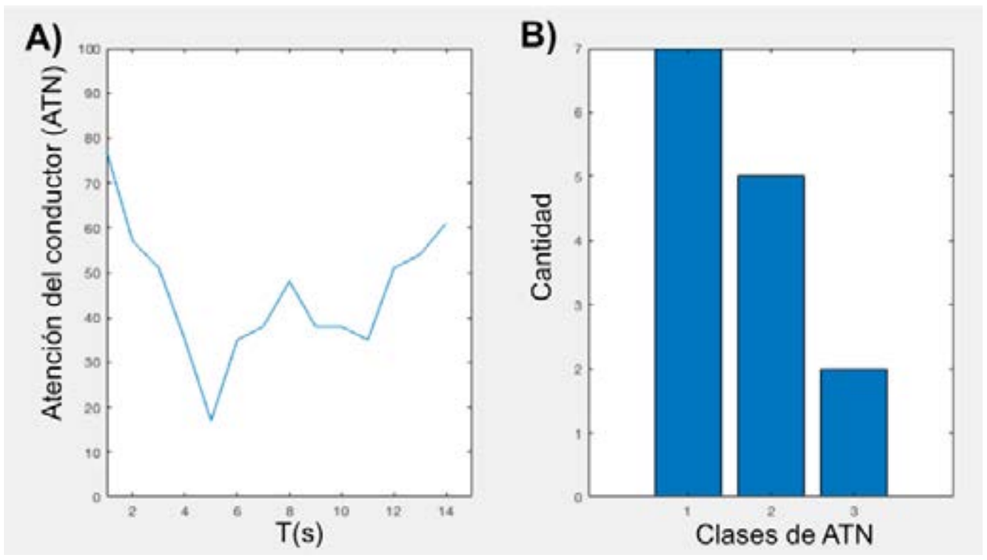


Figura 3. Asignación de clases para los niveles de atención del conductor (ATN) en una ventana de tiempo T: A) Valores de ATN B) Histograma de ATN divididos en 3 clases 1-bajos (0-40), 2-medios (41-60) y 3-altos (61-100)

Redes neuronales holográficas con cuantificación difusa (FQHNN)

Como las FQHNN han evolucionado a partir de las redes neuronales holográficas (HNN) propuestas por (Sutherland, 1990), primeramente se introduce la teoría básica y luego se describe el método propuesto.

Teoría Básica (Sutherland, 1990)

Si se asumen dos vectores reales representados por el vector de entrada $\mathbf{x} = \{x_1, x_2, ..., x_k\}^T$ y el vector de salida $\mathbf{Y} = \{y_1, y_2, ..., y_m\}^T$, en datos de muestra pueden ser representados por las matrices de entrada y salida **X** e **Y** de la ecuación (1).

$$\mathbf{X} = \begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \quad \mathbf{Y} = \begin{pmatrix} Y_1^T \\ \vdots \\ Y_n^T \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{pmatrix} \tag{1}$$

Los elementos de las matrices **X** e **Y** se pueden convertir a los ángulos θ_{ti} (t = 1, ···, n; i = 1, ···, k) y ϕ_{tj} (t = 1, ···, n; j = 1···, m) por las funciones de mapeo f_x y f_y de la ecuación (2) y luego cada ángulo se mapea al plano complejo mediante la función exponencial de (3).

$$\theta_{ti} = f_x(x_{ti}) \qquad \phi_{tj} = f_y(y_{tj}) \tag{2}$$

$$s_{ti} = \lambda_{ti} e^{\hat{\imath}\theta_{ti}} \qquad r_{tj} = \gamma_{tj} e^{\hat{\imath}\phi_{tj}} \tag{3}$$

En la ecuación (2), las funciones f_x y f_y pueden ser lineales, sigmoides o tangentes exponenciales inversas. En la ecuación (3), λ es una unidad imaginaria y mediante las operaciones de las ecuaciones (2) y (3), la entrada **X** y la salida **Y** se representan en el plano complejo como el estímulo **S** y la respuesta **R,** respectivamente.

$$\mathbf{S} = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1k} \\ s_{21} & s_{22} & \cdots & s_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nk} \end{pmatrix} \qquad \mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{pmatrix} \tag{4}$$

La función de transferencia $\mathbf{H} = [\mathbf{h}_1, ..., \mathbf{h}_m]$ de las HNN se obtiene a partir de minimizar la diferencia entre los datos de entrenamiento en **R** y el producto **S** · **H** según la ecuación (5)

$$\mathbf{H} = (\mathbf{S}^* \cdot \mathbf{S})^{-1} \cdot \mathbf{S} \cdot \mathbf{R} \tag{5}$$

Aquí, el símbolo * representa la conjugada compleja de la matriz. La salida **V** para la nueva entrada **U** se predice mediante **V** = U · **H** utilizando la matriz **H.** Además, Sutherland obtiene la siguiente ecuación (6) para trabajar con series de tiempo (Diago, *et al.,* 2011a).

$$h_{ij} = \frac{1}{c} \int_{t_0}^{T} M(\lambda_{ti}, \gamma_{tj}, t) e^{i(\phi_{tj} - \theta_{ti})} e^{(t-T)/\sigma\sigma\sigma\sigma\sigma\sigma\sigma\sigma\sigma\sigma} dt \tag{6}$$

donde la Matriz **M** se calcula a partir de la matriz inversa de **S*** **S** en la ecuación (5) para una ventana de tiempo **T** y los valores de λ_{ti} y γ_t en la ecuación (3). El valor de **σ** permite definir perfiles de memoria de corto o largo plazo.

Cuantificación difusa (Diago, *et al.*, 2011a)

En la teoría básica propuesta por Sutherland se incluye la posibilidad de expandir el número de funciones bases de la ecuación (3) a través del uso de términos estadísticos de orden superior según la expresión $\prod_i^K \lambda_i e^{i\theta i}$ para mejorar la exactitud en la predicción de las HNN. Sin embargo, la forma de determinar el número de funciones bases requeridas (K) se continúa haciendo de forma empírica para cada aplicación. Diago y colaboradores (2011a) proponen aumentar la precisión de las HNN y a la vez determinar el número de funciones bases necesarias para cada aplicación de forma automática mediante el empleo de la teoría de cuantificación difusa tipo II [6, 11]. Para ello, cada variable de entrada x_i(i = 1, …, k) se expresa como $f_1(x)$, se considera que cada una tiene una varianza arbitraria $-\infty<x_i<\infty$. Se divide el dominio externo de la variable x_i en l_i categorías C_{il} (l = 1, ..., l_i), se calculan las nuevas funciones de distribución para cada categoría y se determina las fronteras $\tau_1, \tau_2,...,\tau_{l_i-1}$) de cada una por la ecuación (7).

$$\int_{-\infty}^{\tau_1} f_1(x)dx = \int_{\tau_1}^{\tau_2} f_1(x)dx = \ldots = \int_{\tau_{l_i-1}}^{\infty} f_1(x)dx \quad (7)$$

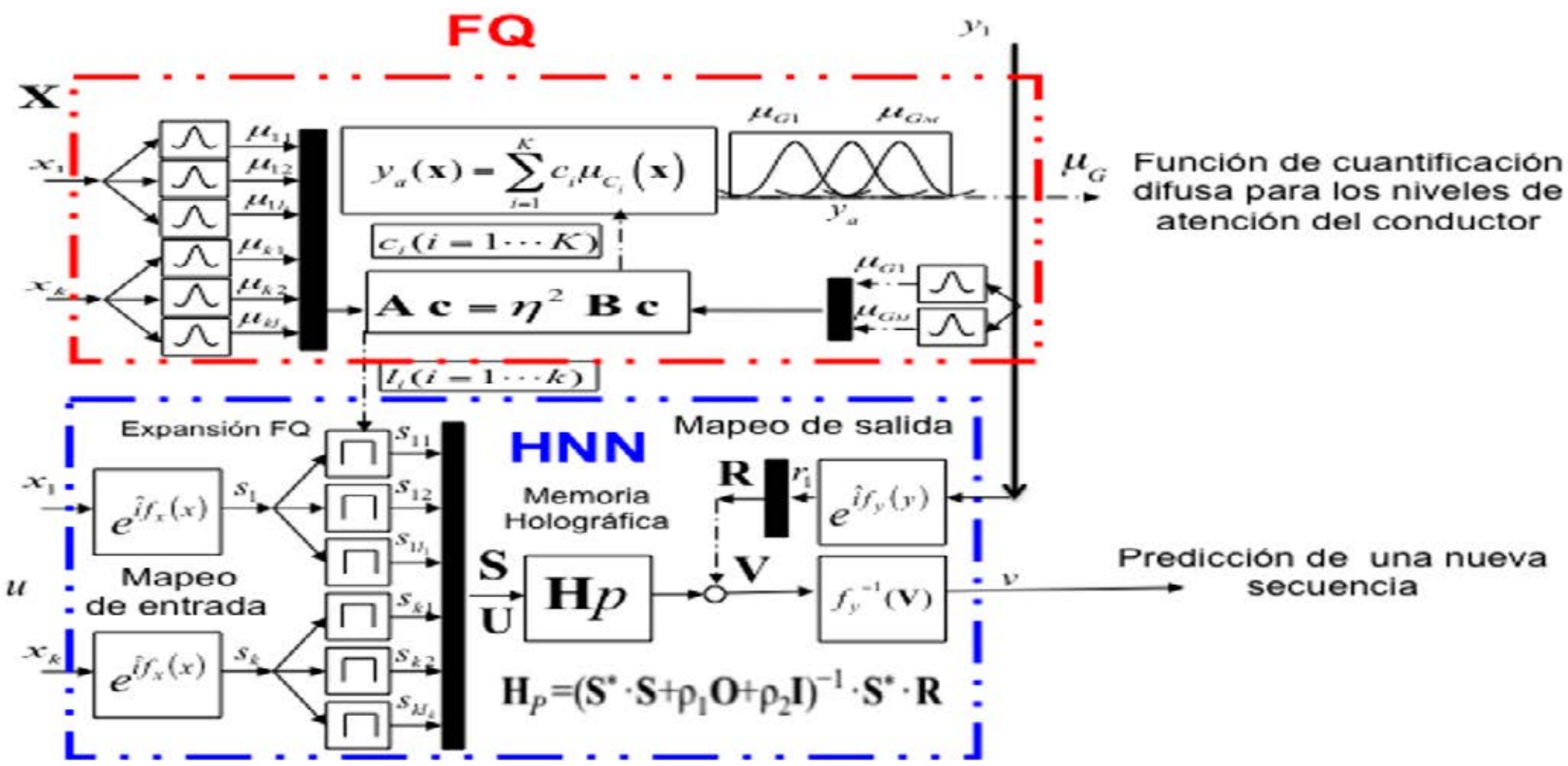
De esta manera se pueden obtener reglas del tipo IF-THEN por cada categoría, según se expresa en la ecuación (8):

$$\begin{aligned} &IF \quad -\infty < x_i \le \tau_1 \quad THEN \quad x_i \in C_{i1} \\ &IF \quad \tau_1 < x_i \le \tau_2 \quad THEN \quad x_i \in C_{i2} \\ &\quad\quad . \quad . \quad . \\ &IF \quad \tau_{il_i} < x_i < \infty \quad THEN \quad x_i \in C_{il_i} \end{aligned} \quad (8)$$

El número de categorías l_i en la ecuación (8) coincide con el número de funciones bases de la ecuación (3) y se determina automáticamente al resolver un problema de auto-valores propios generalizado $Ac=\eta^2 Bc$ donde las matrices **A** y **B** se calculan a partir de las funciones de pertenencia $\mu C_{il}(x_t)$ y $\mu G_j(y_a)$ de los grupos difusos de las variables de entrada y salida respectivamente (Diago, Kitaoka, Hagiwara e Ishiguro, 2011b). El valor propio máximo η^2 y su vector propio correspondiente (*c)* proporcionan el grado máximo de separación de los grupos difusos. Finalmente, se puede obtener una representación simple de la información contenida en **H** al dibujar la estructura de los grupos difusos de salida en el eje de números reales a partir de la ecuación (9):

$$y_a(x_t) = \sum_{i=1}^{K} c_i \, \mu_{C_i}(x_t) \quad (9)$$

Los valores de x_t (t=1…n) incluyen una ventana de tiempo representada en la ecuación (1) y se utiliza para encontrar la función de cuantificación difusa correspondiente a los niveles del conductor (*off-line)* antes de comenzar la predicción de dichos niveles a partir de las secuencias de video. La figura 4 muestra la arquitectura general de la red neuronal con cuantificación difusa propuesta. En este trabajo se emplean por primera vez las FQHNN para obtener una representación adaptada a las características de cada conductor y estimar el estado de los conductores en el nivel de conducción autónoma condicional a partir de sus expresiones faciales.



Al igual que para HNN, el aumento en el número de términos de expansión en el caso de FQHNN permite reducir el error durante el aprendizaje pero la matriz resultante en la ecuación (5) puede no ser invertible. Fukushima, Kamada y Hagiwara (2004) propusieron el uso de los parámetros de penalización p_1 y p_2 en la ecuación (10) para aumentar el rendimiento de generalización de HNN y evitar el cálculo de la inversa en la ecuación (5).

$$\mathbf{H}_P = (\mathbf{S}^* \cdot \mathbf{S} + p_1\mathbf{O} + p_2\mathbf{I})^{-1} \cdot \mathbf{S}^* \cdot \mathbf{R} \qquad (10)$$

Aquí, **O** es una matriz cuadrada con todos sus elementos iguales a uno e **I** es la identidad. Si p_2 no es cero, como el rango de la matriz ($\mathbf{S}^* \cdot \mathbf{S} + p_1\mathbf{O} + p_2\mathbf{I}$) concuerda con el tamaño de la matriz, la matriz inversa siempre existe. Cuando los parámetros p_1 y p_2 están cerca de cero, la primera parte de la ecuación (10) tiene mayor peso y la matriz se acerca a la matriz de Moore-Penrose (Fukushima *et al.,* 2004) que permite resolver tales sistemas, incluso con deficiencia de rango, y proporciona vectores de norma mínima del error de aprendizaje. Al igual que en los trabajos anteriores [6, 9] este trabajo emplea los valores de $p_1 = 1$ y $p_2 = 1$ en la ecuación (10) para obtener mejores resultados de generalización. En los experimentos con FQHNN se usaron las funciones $f_x{:}\theta_{ti} = \pi/2 + 2\pi\,(l\text{-}1) / l_i \;\forall x_{ti} \in C_{il}\,(l=1...l_i)$ y, $f_y{:}\,\phi_{tj} = \pi/2 + 2\pi y_t / m$ en la ecuación (2) donde $l_i = 3$, m=3, y los valores de $\lambda_{ti}=1$, $\gamma_t=1$ y $\tau=1$ en las ecuaciones (3) y (6).

Comparación con las redes neuronales de memoria a largo plazo (LSTM)

En la sección experimental se comparan las redes FQHNN con las redes de memoria a largo plazo (LSTM) que son ampliamente usadas en la literatura para la clasificación de secuencias de tiempo. Las LSTM son un tipo especial de redes neuronales recurrentes (RNN), capaces de aprender las dependencias a largo plazo (Hochreiter y Schmidhuber, 1997). Como el problema presentado tiene un alto desequilibrio en las clases se usa la exactitud (Acc, en inglés *Accuracy)* de la predicción (10) y los valores del coeficiente *Kappa* en (11) como métricas para la comparación calculados a partir de la matriz de confusión ***MC*** según:

$$Acc = \frac{(\sum_{i=1}^{m} MC_{ii})}{N} \quad (10)$$

$$Kappa = \frac{(Pr_{(a)} - Pr_{(e)})}{(1 - Pr_{(e)})} \quad (11)$$

donde ***MC*** es una matriz cuadrada de dimensión (*m* x *m*) donde *m* denota el número de clases del problema, los valores de la diagonal de la matriz MC_{ii} tienen la cantidad de muestras correctamente clasificadas para clase *i* y N es el número total de muestras. $Pr_{(a)}$ es el acuerdo observado relativo entre los evaluadores y Pr_e es la probabilidad hipotética de acuerdo por azar. *Kappa* mide el nivel de acuerdo en un rango de valores entre un valor mínimo igual a 0 y un máximo de 1 que indica acuerdo absoluto. Valores < 0 indican (Acuerdo menos que casual), 0-0,2 (Acuerdo leve), 0,2-0,4 (Acuerdo justo), 0,4-0,6 (Acuerdo moderado), 0,6-0,8 (Acuerdo sustancial) y 0,8-1,0 (Concordancia casi perfecta). En este caso un evaluador sería el valor de ATN obtenido a partir de asignar clases a los valores de *NeuroSky* (*"Ground Truth"*) y el otro sería la evaluación del modelo en cuestión (FQHNN o LSTM). En este trabajo se usa la implementación de *Matlab* 2018b del LSTM con los siguientes parámetros (*MaxEpochs*= 10, *MiniBatchSize*= 150, *InitialLearnRate*= 0,01, *SequenceLength*= 1000, *GradientThreshold*= 1). La arquitectura del LSTM se probó con una configuracion 20-100-3, donde los 20 parámetros de la figura 2 se utilizan como entrada, el número de unidades ocultas es 100 y el número de salidas es 3. Se utilizó el optimizador con estimación de momento adaptativo (ADAM en inglés).

Para evaluar ambos modelos de aprendizaje, el trabajo sigue el ejemplo de *Matlab* que muestra cómo clasificar datos de secuencias utilizando una red LSTM. El algoritmo se basa en el método desarrollado por Kudo, Toyama y Shimbo (1999). Durante el entrenamiento, de forma predeterminada, el *software* divide los datos de entrenamiento en mini lotes y rellena las secuencias para que tengan la misma longitud. Para evitar que el proceso de entrenamiento agregue demasiado relleno, los datos de entrenamiento se ordenan por longitud de secuencia y se elige un tamaño de mini lote para que las secuencias de un mini lote tengan una longitud similar. En los experimentos se dividen los datos en secuencias impares para entrenamiento y secuencias pares para prueba.

RESULTADOS Y DISCUSIÓN

En la figura 5 se muestran 3 entornos en los que se ha utilizado el método propuesto. En el primer entorno (figura 5A) se muestran los resultados de la detección y el seguimiento de los 68 puntos característicos de los rostros de 18 sujetos que participaron en los experimentos en un entorno no controlado usando *NeuroSky Mindwave Mobile.* Varios de los problemas encontrados en el coche autónomo se presentan en este entorno de simulación. Como se muestra en la figura, la cara no se detectó en algunos sujetos (por ejemplo, fila 3, columnas 1 y 3) debido a los movimientos de los sujetos fuera del ángulo visual de la cámara. En otros casos, incluso si se detecta la cara, las señales recibidas de los sensores no tienen un nivel suficiente (señal deficiente) para que los niveles de ATN recibidos sean fiables, por lo que en muchos casos esos datos se pierden.

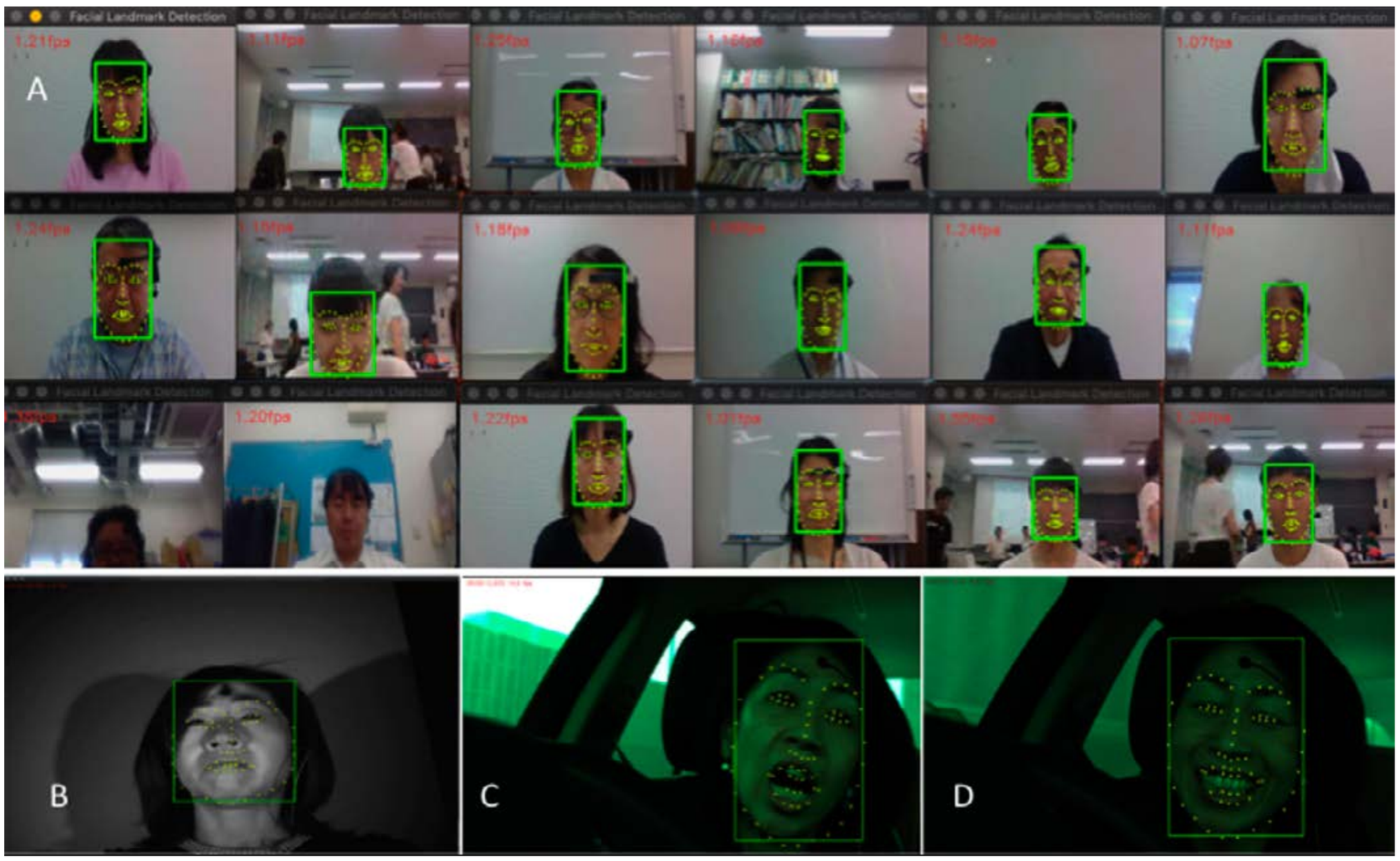


Figura 5. Ejemplos de puntos característicos detectados en la cara de los conductores en distintos entornos de simulación: A) 18 sujetos en un entorno no controlado que usa NeuroSky Mindwave Mobile B) Sujeto No.9 en un entorno controlado con medición de 5 variables fisiológicas C) Sujeto No.9 en el Simulador de conducción condicional en el momento de recibir la señal "*Request-to-Intervent* (RtI)" D) Sujeto No.9 después de conmutar al modo de conducción manual.

Ocho de los 18 sujetos fueron invitados a participar en un segundo experimento donde se muestran videos de estímulo y se miden 5 variables fisiológicas (electrocardiograma o ECG, electroencefalograma o EEG, actividad electrodérmica o EDA, pulso sanguíneo y frecuencia respiratoria) en un entorno controlado. En la figura 5B se muestra un cuadro del video obtenido por una cámara frontal con los 68 puntos característicos detectados en la cara del sujeto No. 9 (fila 2 columna 3) de la figura 5A, y en la tabla 1 se muestra un segmento de los datos obtenidos

para el mismo sujeto durante el experimento. Además, de las 5 variables fisiológicas, la base de datos almacena las señales de reloj (*Clock)* que indica el momento en que se realizó el experimento y una marca (*Mark)* que permiten sincronizar las señales de video con el estímulo mostrado al sujeto. De los 8 sujetos del segundo experimento se seleccionaron 3, para los cuales las variables fisiológicas registradas mostraron una alta correlación con los parámetros mostrados en la figura 2 para un tercer experimento en un simulador de conducción condicional. Las figuras 5C y 5D muestran a uno de los 3 sujetos en el momento de recibir la señal "*Request-to-Intervent* (RtI)" (5C) y después de conmutar al modo de conducción manual (5D). Como se muestra en la figura, los 68 puntos característicos de la cara del sujeto también fueron adquiridos incluso cuando el simulador muestra vehículo con variaciones de iluminación durante la marcha. En este trabajo solo reportamos los resultados del primer experimento de la figura 5A, cuyos datos se encuentran disponibles para investigación bajo solicitud a los autores.

Tabla 1. Segmento de los datos correspondientes a las 5 variables fisiológicas (electrocardiograma o ECG, actividad electrodérmica o EDA, electroencefalograma o EEG, pulso sanguíneo y frecuencia respiratoria) registradas para el sujeto de la figura 5B con el tiempo (Clock) del experimento y las marcas (Mark) para sincronización.

CLOCK	ECG	EEG	Resp	Pulse	EDA	Mark
11:57:52.040	0.92	-6.34	-14.95	-175.16	-7.93	4.01
11:57:52.050	6.64	-18.32	-52.04	-611	-39.64	4.01
11:57:52.060	16.17	-27.18	-68.67	-819.5	-75.84	4.01
11:57:52.070	35.02	-26.18	-63.18	-764.19	-94.11	4.01
11:57:52.080	43.11	-20.38	-61.88	-688.97	-100.36	4.01
11:57:52.090	48.45	-14.81	-62.03	-622.6	-100.01	4.01
11:57:52.100	47.61	-16.95	-62.8	-533.64	-98.27	4.01
11:57:52.110	42.27	-19.24	-61.8	-397.54	-96.81	4.01
11:57:52.120	46.01	-16.34	-61.12	-288.68	-96.09	4.01
11:57:52.130	51.73	-9.16	-60.81	-126.18	-95.84	4.01
11:57:52.140	62.1	-7.33	-60.58	2.82	-95.79	4.01
11:57:52.150	79.42	-10.53	-61.88	131.83	-95.77	4.01
11:57:52.160	87.21	-10.69	-62.95	276.93	-95.73	4.01
11:57:52.170	78.43	-7.63	-61.65	374.5	-95.66	4.01
11:57:52.180	62.03	-5.57	-59.59	506.18	-95.58	4.01
11:57:52.190	43.56	-12.14	-56.16	595.21	-95.48	4.01
11:57:52.200	28.31	-17.02	-51.35	682.33	-95.39	4.01
11:57:52.210	12.82	-11.83	-48.6	782.8	-95.31	4.01
11:57:52.220	-2.14	-6.11	-49.52	832.47	-95.22	4.01
11:57:52.230	-11.14	1.15	-53.87	917.15	-95.13	4.01
11:57:52.240	-11.98	7.18	-60.13	957.43	-95.03	4.01
11:57:52.250	-10.07	6.64	-66.76	991.76	-94.94	4.01
11:57:52.260	-8.16	-1.37	-73.1	1036.54	-94.84	4.01
11:57:52.270	0.31	-10.15	-79.12	1025.86	-94.74	4.01
11:57:52.280	15.11	-14.73	-83.78	1045.93	-94.65	4.01
11:57:52.290	23.88	-19.16	-85.76	1013.58	-94.56	4.01

La tabla 2 muestra la distribución de la atención del conductor representada en 3 clases (baja, media, alta) calculadas a partir de ventanas de tiempo T de diferentes tamaños para 17 sujetos cuyos datos fueron válidos en el experimento 1 y la exactitud (*Acc)* en la predicción con el mo-

delo FQHNN para los mejores valores de *Kappa* y su correspondiente ventana de tiempo **T**. Los valores de Kappa y Acc muestran los resultados en el conjunto de prueba. Los valores promedio de Kappa y Acc para los conjuntos de entrenamiento y prueba se muestran en la tabla 2. La tabla 2 muestra que cada sujeto presenta un problema de clasificación con clases altamente no-balanceadas y que el número de muestras para entrenamiento y prueba disminuye con el aumento de la ventana de tiempo T.

Tabla 2. Distribución de la atención del conductor representada en 3 clases (Baja, Media, Alta) calculadas a partir de ventanas de tiempo T de diferentes tamaños para 17 de los 18 sujetos que participaron en el experimento 1 y la exactitud (Acc) en la predicción en el conjunto de prueba con el modelo FQHNN para los mejores valores de Kappa y su correspondiente ventana de tiempo T.

	T=5			T=10			T=15			T=20			FQHNN (Best Kappa)		
Sujeto	Baja	Media	Alta	Baja	Media	Alta	Baja	Media	Alta	Baja	Media	Alta	Acc	Kappa	T(s)
1	8	46	27	3	27	15	0	22	9	1	15	8	0.733333333	0.4	15
2	42	18	3	25	11	2	19	7	1	13	8	0	0.692307692	0.307692308	15
3	32	35	14	18	20	9	14	15	3	13	9	3	0.347826087	0.318518519	10
4	19	45	26	8	30	11	7	15	12	4	16	6	0.176470588	0.46031746	15
5	21	11	29	13	6	18	8	4	15	7	1	12	0.5	0.111111111	5
6	7	12	13	4	8	11	4	7	8	4	5	6	**0**	**0.555555556**	**20**
7	63	3	4	35	3	3	27	1	2	20	1	2	0.857142857	0.7	5
8	76	3	4	45	1	2	30	1	2	21	1	3	**1**	**1**	**20**
9	90	0	0	49	0	0	34	0	0	26	0	0	**0.955555556**	**0.9**	**5**
10	0	32	53	0	17	30	0	13	19	0	8	17	0.25	0.407407407	20
11	46	31	13	29	14	6	18	12	4	15	8	3	0.511111111	0.090909091	5
12	90	0	0	49	0	0	34	0	0	26	0	0	**0.955555556**	**0.9**	**5**
14	12	39	34	5	21	20	2	15	15	1	12	11	**0.833333333**	**0.625**	**20**
15	33	17	40	19	9	21	14	4	16	8	5	13	0.384615385	0.277777778	20
16	34	17	33	18	6	22	14	5	13	9	2	14	0.25	0.407407407	20
17	14	41	29	6	22	18	4	18	10	2	16	6	0.428571429	0.222222222	5
18	22	35	31	10	18	21	6	15	13	4	13	9	0.318181818	0.348148148	5

Por ejemplo, si se toma una ventana de tiempo T de 5s, varios de los sujetos están casi todo el tiempo con niveles de atención baja (sujeto 9 y 12) o media-alta (sujeto 10). En estos casos se utiliza una estrategia de aprendizaje incremental parecida a la deriva virtual mencionada por Gama, Žliobaitė, Bifet, Pechenizkiy y Bouchachia (2014) que permite acomodar nuevas clases e introducir nuevos datos para el concepto de atención del conductor. Como en este caso se conoce el número total de clases posibles, se generan aleatoriamente secuencias virtuales de los parámetros de los rasgos faciales de los sujetos y se asignan clases a dichas secuencias para que el número de muestras por clase (es decir, distribución de probabilidad de clase) sea un número distinto de cero. Para T= 5s el número de casos con clases de probabilidad 0 es 3 (sujetos 9, 10 y 12), pero este número aumenta con el aumento del tamaño de la ventana (4 para **T**= 10s, 6 para **T**= 15s y 9 para **T**= 20s). En estos casos también se incluyen las clases con una sola secuencia pues tampoco se permitiría probar el modelo desarrollado. Las últimas tres columnas muestran la exactitud en la predicción con el conjunto de prueba usando el modelo FQHNN para los mejores valores de Kappa y su correspondiente ventana de tiempo T. Los resultados de la tabla muestran que hay sujetos que se pueden predecir con exactitud (85 < Acc < 100 %) a partir de sus expresiones faciales y se logra un acuerdo sustancial o casi perfecto con los niveles

de atención asignados por *NeuroSky* (*Kappa*>0.6, sujetos 7, 8, *9*, 12 y 13). Sin embargo, hay otros que aun cuando su acuerdo con el dispositivo de NeuroSky es moderado (0.4< *Kappa*< 0.6, sujetos 1, 4, 6, 10 y 15) la exactitud de la predicción está dispersa; 0 % (para sujeto 6), muy baja (< 25 % para sujetos 4, 10 y 15) o muy alta (73 % para el sujeto 1). La arquitectura de FQHNN propuesta permite obtener funciones de cuantificación difusa para los niveles de atención de cada conductor que permite aclarar las causas de los resultados de las predicciones obtenidas por el método. Después de resolver el problema de optimización según (Diago, *et al.,* 2011b), se obtiene el máximo autovalor propio y su autovector correspondiente para construir las funciones de cuantificación difusa de salida para cada nivel de atención a la marcha del vehículo con la ecuación (9). Por ejemplo, en la figura 6 se muestran las funciones de membresía para dos sujetos (9 y 4) para los cuales los modelos exhiben comportamientos diferentes. Cada curva de la gráfica muestra los resultados de la interpolación de los puntos (*Ya,Br),* donde *Ya* se calcula según la ecuación (9) y *Br* según las funciones de membresía de cada grupo de salida del sujeto: baja (línea roja), medio (línea verde) y alta (línea azul). Estas funciones están en correspondencia con los valores indicados en la tabla 2 para ambos sujetos. La gráfica muestra que el sujeto 9 mantiene casi todo el tiempo niveles de atención baja (línea roja) y presenta pocas zonas de solapamiento de las funciones de membresía. Esto hace que el porcentaje de predicción sea de 95 %. Mientras que el sujeto 4, presenta muchas zonas de solapamiento entre las funciones que hacen más difícil predecir su estado de atención (solo 17 %) aunque la mayor parte del tiempo se mantiene con un nivel medio (línea verde) de atención. De esta manera, por medio de las funciones de cuantificación difusa para los niveles de atención del conductor mostradas en la figura 6 se pueden analizar las causas de los bajos porcentajes de predicción del modelo FQHNN.

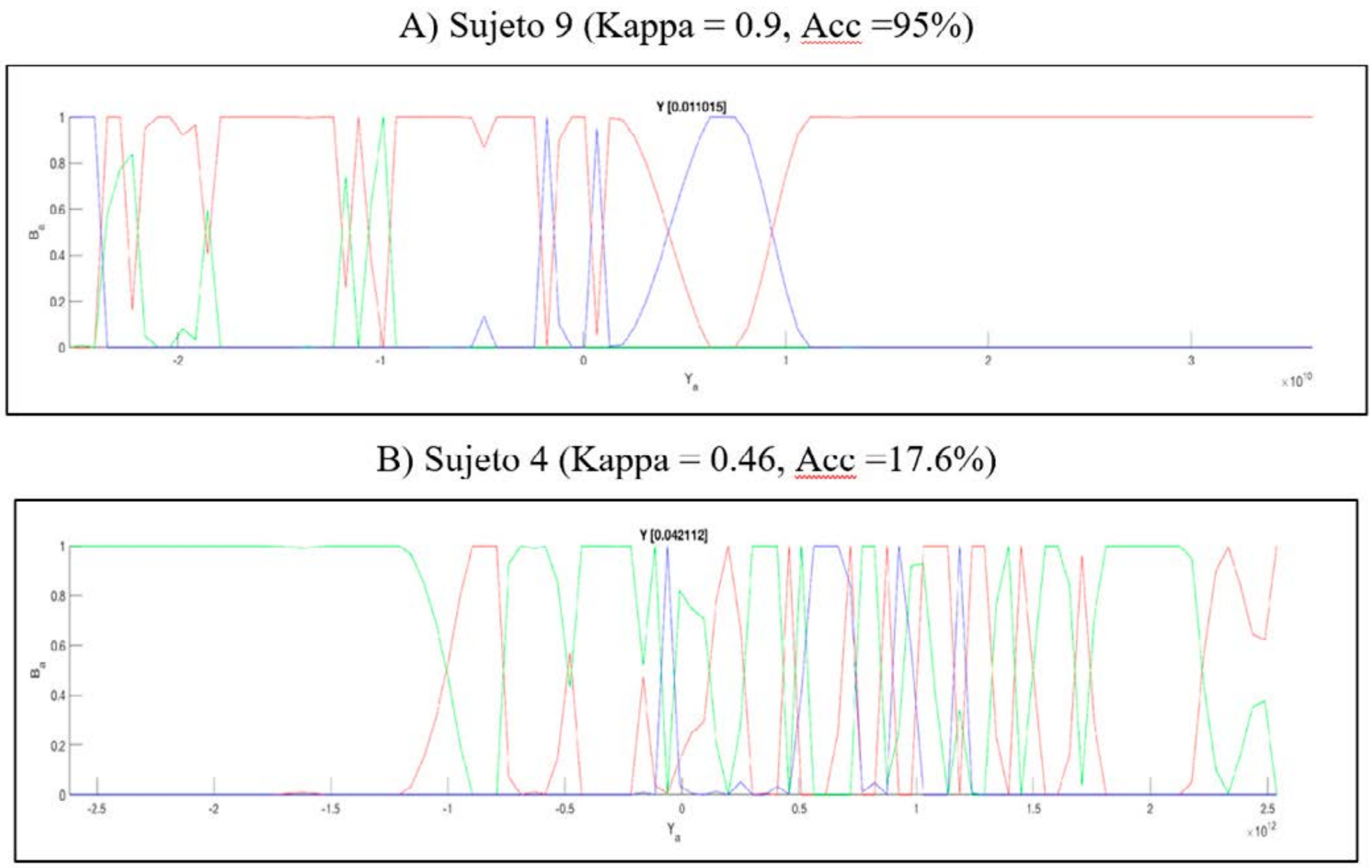


Figura 6. Funciones de cuantificación difusa para los niveles de atención de dos sujetos.

Por último, la tabla 3 muestra una comparación de los tiempos y la exactitud del entrenamiento y la prueba de los modelos LSTM y FQHNN para diferentes ventanas de tiempo **T**. Los experimentos se desarrollaron en una computadora *MacBook Air* (13-pulgadas, de mediados de 2013), con un procesador de 3 GHz *Intel Core* i5 y 4GB 1600 MHz DDR3. Todos los programas para la creación y análisis de los modelos se desarrollaron en *Matlab* Version: 9.6.0.1174912 (R2019a) actualización 5. Para ambos casos, con el aumento del tamaño de la ventana disminuye el tiempo de procesamiento tanto para el entrenamiento como para la prueba. Sin embargo, el tiempo de entrenamiento para la red FQHNN es mucho menor que el tiempo de entrenamiento para LSTM, lo que hace que en la mayoría de las aplicaciones actuales las redes LSTM sean entrenadas con base de datos fuera de línea y utilizadas para predecir solamente en etapa de pruebas del modelo donde los tiempos para ventanas de tiempo **T**= 20s son menores que de 1s (0,85 como promedio). Como los videos que se usaron en el experimento fueron solo de 10 min (600s), para una ventana de tiempo T= 20s, solo se cuenta con 30 secuencias como máximo para entrenamiento y prueba si no se pierden fotogramas por los movimientos del sujeto y errores de la comunicación Bluetooth. Para ninguno de los sujetos del experimento se logró contar con 30 secuencias válidas (ver T= 20s en la tabla 1). Sin embargo, para el caso de trabajar con pocas muestras el modelo FQHNN puede ser entrenado con un 97,1 % de exactitud (*Kappa*= 0,936), mientras que el modelo LSTM solo llega a un 54 % de exactitud (*Kappa*= 0,428). Para el modelo FQHNN no existen diferencias significativas en los tiempos de entrenamiento y prueba (0,03 y 0,02 segundos respectivamente). Aunque los valores de *Kappa* para ambos modelos exhiben una coincidencia justa (*Kappa*= 0,3) con las clases asignadas a partir de los niveles de atención del dispositivo NeuroSky MindWave, la exactitud del modelo FQHNN es más de un 7 % superior al modelo LSTM como promedio.

Tabla 3. Comparación de los tiempos y la exactitud del entrenamiento y la predicción de los modelos LSTM y FQHNN para diferentes ventanas de tiempo T.

	T(s)	Entrenamiento			Prueba		
		Tiempo (s)	Exactitud	Kappa	Tiempo (s)	Exactitud	Kappa
LSTM	5	259.6637726	0.432774432	0.36322364	6.65843087	0.385796055	0.398082896
		148.0760425	0.251985815	0.24606494	3.803059647	0.260100843	0.22487315
	10	143.1787068	0.491512445	0.397376448	2.420290973	0.497234119	0.399165647
		108.255066	0.258251426	0.252303166	2.294702962	0.265866409	0.250277418
	15	78.75893723	0.470176306	0.324880891	1.178297481	0.394258126	0.316752025
		83.71322241	0.258598633	0.216646366	1.38268358	0.216126268	0.184514028
	20	73.2305452	0.540206362	0.428842971	0.856751678	0.452146873	0.377412212
		69.70108797	0.277731119	0.248432975	0.834031237	0.275661727	0.199153852
FQHNN	5	0.045498967	0.891328997	0.755490242	0.033099027	0.549811604	0.331587513
		0.03464906	0.097752458	0.21994303	0.020262523	0.212072309	0.271242405
	10	0.039554114	0.937907076	0.860290921	0.030112153	0.56562203	0.310490304
		0.031906417	0.070858237	0.159431033	0.018540665	0.201287881	0.271681343
	15	0.038649143	0.932525952	0.848183391	0.029847085	0.571840712	0.360711268
		0.026721636	0.071241297	0.160292917	0.018052212	0.247630128	0.271025991
	20	0.033309267	0.971719457	0.936368778	0.024062084	0.561332785	0.344279141
		0.023798123	0.056298094	0.126670711	0.012910136	0.254800032	0.277773408

CONCLUSIONES

En este estudio, se propone un enfoque para caracterizar la disponibilidad observable de toma de control de los conductores en vehículos autónomos y un modelo de aprendizaje automático para estimarlo. Se recopila un conjunto de datos de múltiples variables fisiológicas de conductores de vehículos condicionalmente autónomos en un simulador que constituye una base de datos útil para evaluar diferentes modelos de aprendizaje automático durante la conducción autónoma condicional. A partir de una de las variables fisiológicas recopiladas se desarrolló un método para la anotación automática sobre la preparación para la toma de control de los conductores que permitieron entrenar modelos neuro-difusos con un enfoque de aprendizaje automático explicable. El modelo se adapta a las características de cada conductor y brinda una función de cuantificación difusa que permite explicar los resultados de las predicciones obtenidas por el modelo. El modelo ha sido probado en 2 entornos de simulación con entorno controlado, no controlado y en un simulador de conducción condicional. En este trabajo se reportan los resultados de la comparación del modelo propuesto con los modelos de redes de memoria de largo plazo ampliamente usados en la literatura para predecir secuencias de tiempo. El modelo propuesto alcanza un porcentaje de predicción promedio de 57 % (7 % superior a los modelos LSTM), a la vez que pueden ser utilizados para el aprendizaje en línea debido a que su aprendizaje es 2000 veces más rápido que los modelos de memoria a largo plazo actuales.

Aunque los resultados reportados en el trabajo muestran superioridad de los modelos basados en FQHNN en comparación con los modelos LSTM, todavía hay posibilidades de mejora. Los porcentajes de predicción obtenidos por ambos modelos están en concordancia con los resultados reportados por Appriou, Cichocki y Lotte (2020), donde la mayoría de los estudios demostraron que la clasificación de los estados afectivos a partir del EEG siguen siendo un gran desafío, ya que los resultados apenas superan la exactitud de un clasificador aleatorio (50-52 %) cuando se trabaja con pocas muestras. Según los resultados reportados por Appriou (2020), obtener un modelo independiente del sujeto es mucho más desafiante (se obtienen porcentajes menores que para los modelos dependientes del sujeto), pero si tiene éxito, permitiría obtener una verdad absoluta para el monitoreo que no requiere ninguna calibración para nuevos sujetos. Para la asignación de las clases se pudieran utilizar otras de las variables fisiológicas que se registraron en un entorno controlado en busca de una la representación más fiel de la actividad del conductor y su correlación con sus expresiones faciales. En el trabajo se dividieron los valores de ATN en 3 segmentos: bajos (0-40), medios (41-60) y altos (61-100) para la asignación de las clases. Sin embargo, estas fronteras tienden a ser difusas y variables para cada sujeto. Por ejemplo se pudieran usar unidades de acción facial correspondientes al Sistema de Codificación de Acciones Faciales (en inglés, FACS: *Facial Action Coding System)* para explorar otras técnicas de aprendizaje supervisado débil *(weak supervised learning).*

REFERENCIAS

Appriou, A., Cichocki, A., & Lotte, F. (2020). Modern Machine-Learning Algorithms: For Classifying Cognitive and Affective States From Electroencephalography Signals. *IEEE Systems, Man, and Cybernetics Magazine, 6*(3), 29-38.

Ballingall, S., Sarvi, M. & Sweatman, P. (2020). Safety Assurance Concepts for Automated Driving Systems. *SAE Technical Paper Series, 2*(3), pp. 1528-1537. doi:10.4271/2020-01-0727.

Braunagel, C., Rosenstiel, W., & Kasneci, E. (2017). Ready for take-over? A new driver assistance system for an automated classification of driver take-over readiness. *IEEE Intelligent Transportation Systems Magazine, 9*(4), 10-22.

Deo, N., & Trivedi, M. M. (2019). Looking at the driver/rider in autonomous vehicles to predict take-over readiness. *IEEE Transactions on Intelligent Vehicles, 5*(1), 41-52.

Diago, L., Kitaoka, T., Hagiwara, I., & Kambayashi, T. (2011a). Neuro-fuzzy quantification of personal perceptions of facial images based on a limited data set. *IEEE Transactions on Neural Networks, 22*(12), 2422-2434.

Diago, L., Kitaoka, T., Hagiwara, I., & Ishiguro, S. (2011b). Analyzing facial expressions with fuzzy quantification theory II: indefinite generalized eigenvalue problem. *Japan Journal of Industrial and Applied Mathematics, 28*(1), 153-170.

Diago, L., Yang, Y., Abe, H. & Hagiwara, I.. (2018). NeuroFaceLab : A new framework for passengers analysis in autonomous driving. En *Proceedings of the 31st International Computational Mechanics Symposium - CMD2018(286).* Japan: JSME.

Fukushima, H., Kamada, Y., & Hagiwara, I. (2004). Optimum engine mounting layout using MPOD. Nippon Kikai Gakkai Ronbunshu, C Hen/Transactions of the *Japan Society of Mechanical Engineers, Part C, 70*(1), 54-61.

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM computing surveys (CSUR), 46*(4), 1-37.

He, J., Liu, D., Wan, Z., & Hu, C. (2014). A noninvasive real-time driving fatigue detection technology based on left prefrontal Attention and Meditation EEG. In *2014 International Conference on Multisensor Fusion and Information Integration for Intelligent Systems* (MFI) (pp. 1-6). IEEE.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735-1780.

Khan, M. Q. y Lee, S. (2019). A Comprehensive Survey of Driving Monitoring and Assistance Systems. Sensors (Basel, Switzerland), 19(11), 2574. https://doi.org/10.3390/s19112574

Kudo, M., Toyama, J., & Shimbo, M. (1999). Multidimensional curve classification using passing-through regions. *Pattern Recognition Letters, 20*(11-13), 1103-1111.

Roche, F., Somieski, A., & Brandenburg, S. (2019). Behavioral changes to repeated takeovers in highly automated driving: effects of the takeover-request design and the nondriving-related task modality. *Human factors, 61*(5), 839-849.

SAE On-Road Automated Vehicle Standards Committee. (2018). *Taxonomy and definitions*

for terms related to driving automation systems for on-road motor vehicles. SAE International: Warrendale, PA, USA.

Sutherland, J. G. (1990). A holographic model of memory, learning and expression. International *Journal of Neural Systems, 1*(03), 259-267.

Wu, Y., Kihara, K., Hasegawa, K., Takeda, Y., Sato, T., Akamatsu, M., & Kitazaki, S. (2020). Age-related differences in effects of non-driving related tasks on takeover performance in automated driving. Journal of safety research, 72, 231-238.