

Sistema de monitoreo de conductores de vehículos a partir de análisis de expresiones faciales

Vehicle driver monitoring system based on facial expression analysis

DOI: <http://doi.org/10.17981/ingecuc.16.2.2020.14>

Artículo de Investigación Científica. Fecha de Recepción: 26/07/2020. Fecha de Aceptación: 16/09/2020.

Juan Felipe Cordoba Fuzga

Universidad, Politécnico Colombiano Jaime Isaza Cadavid. Medellín (Colombia)
juan_cordoba91111@elpoli.edu.co

Ruben Dario Vasquez Salazar 

Universidad, Politécnico Colombiano Jaime Isaza Cadavid. Medellín (Colombia)
rdvasquez@elpoli.edu.co

Henry Omar Sarmiento Maldonado 

Universidad, Politécnico Colombiano Jaime Isaza Cadavid. Medellín (Colombia)
hosarmiento@elpoli.edu.co

Para citar este artículo:

J. Cordoba Fuzga, R. Vasquez Salazar & H. Sarmiento Maldonado, “Sistema de monitoreo de conductores de vehículos a partir de análisis de expresiones faciales”, *INGE CUC*, vol. 16, no. 2, pp. 192–201, 2020. DOI: <http://doi.org/10.17981/ingecuc.16.2.2020.14>

Resumen

Introducción— Al conducir, la persona se encuentra expuesta a diferentes estímulos que pueden llevar a que se ocasionen accidentes. Aunque numerosas propuestas tecnológicas se han presentado para mantener monitoreado al conductor, estas han pasado por alto el estado anímico en el que este se encuentra, el cual podría generar efectos negativos en la capacidad de reacción al conducir.

Objetivo— Buscar diferentes alternativas de inteligencia artificial para el análisis permanente de rostros de conductores, con el fin de encontrar un buen modelo de clasificación de expresión facial (feliz, enojo, sorpresa, neutral).

Metodología— La metodología utilizada consiste en la selección de una base de datos que es preprocesada, para posteriormente entrenar diferentes modelos y realizar comparaciones de precisión entre ellos.

Resultados— Se logra encontrar una precisión mayor al 80% en la detección del estado anímico del usuario. Y se logra migrar el modelo a un sistema de monitoreo portátil.

Conclusiones— En este caso particular los métodos de aprendizaje de máquina tradicionales (machine learning) consumen menos tiempo a la hora de clasificar, sin embargo, estos son superados en precisión por un aprendizaje profundo.

Palabras clave— Expresiones faciales; inteligencia artificial; aprendizaje de máquina; aprendizaje profundo; visión artificial

Abstract

Introduction— When driving, any person is exposed to different stimuli that can lead to accidents. Although numerous technological proposals have been presented to keep the driver monitored, these have overlooked the state of mind in which they driver is, which could have negative effects on the ability to react when driving.

Objective— Find different artificial intelligence alternatives for the permanent analysis of drivers' faces, in order to find a good model for classifying facial expression (happy, angry, surprise, neutral).

Methodology— The methodology proposed consists in the selection of a database that is pre-processed, in order to later train different models and make precision comparisons between them.

Results— It is possible to find a precision greater than 80% in the detection of the user's mood and then the model is migrated to a portable monitoring system.

Conclusions— In this particular case, traditional machine learning methods consume less processing time when classifying, however, they are exceeded in precision by deep learning.

Keywords— Facial expressions; artificial intelligence; machine learning; deep learning; computer vision

I. INTRODUCCIÓN

Al momento de conducir, el ser humano está expuesto a numerosos riesgos que lo pueden llevar a perder el control del vehículo, por ello es necesario que el conductor se permita atender a más de un estímulo al mismo tiempo para posteriormente seleccionar la información más relevante y continuar con su labor. Todo ello debe de realizarse en el menor tiempo posible para evitar problemas en el desempeño de la tarea. Sin embargo, existen estados anímicos o mentales como la desconcentración que perjudican esta capacidad de reacción, llevando a desempeñar las tareas de forma menos efectivas, generando así riesgos en las vías [1]. Tanto el enojo como la risa pueden ser distractores a la hora de realizar una tarea que necesita gran concentración.

Existen diferentes formas como las autoridades encargadas han buscado reducir los riesgos de accidentes en las vías. Entre ella se encuentra el uso de retenes y cámaras de foto multas, sin embargo, la reducción de riesgos no ha presentado resultados satisfactorios [2]. Por otro lado, los recientes avances tecnológicos han permitido a los fabricantes de automóviles incorporar en sus vehículos sensores y mecanismos de protección, en donde evitan que conductores en mal estado estén tras el volante, lo que ha permitido reducir la accidentabilidad en vehículos de última generación.

La clasificación de los niveles de alerta de un conductor es una necesidad que puede ser atendida, ya que ésta determina el tiempo de una persona procesa y reacciona de acuerdo a la situación. Aunque parezca que dicha reacción normalmente es inmediata, en realidad el tiempo de esta oscila entre 0.5 s y 1 s [3], aunque este intervalo puede variar dependiente del estado actual de la persona.

A pesar que para el ser humano resulte una tarea fácil y cotidiana la posibilidad de reconocer una de las cuatro emociones básicas en una expresión, los avances tecnológicos han permitido desarrollar diferentes métodos de reconocimiento de expresión por medio de una computadora, ya que esto permitiría desarrollar estudios específicos en temas como la publicidad, o la psicología [4]. Uno de estos métodos consiste en la implementación de algoritmos inteligentes que aprenden el comportamiento de las expresiones fáciles, para posteriormente clasificar la emoción según el posicionamiento de puntos clave en el rostro.

II. METODOLOGÍA

Para el correcto desarrollo de este artículo, se aplicó una metodología dividida en diferentes etapas, entre las cuales se consultan y exploran diferentes herramientas. En la primera de las etapas se definen las expresiones faciales más relevantes que expresa un usuario al momento de conducir, especialmente aquellas que generan algún tipo de peligro.

La segunda parte de este artículo, busca una base de datos adecuada que dé solución al problema, esta debe de contener diferentes fotografías de rostros o simplemente de personas en diversos estados anímicos, para que de esta manera sea posible realizar una clasificación. En caso que ninguna base de datos cumpla con la cantidad o características esperadas, se optará por realizar una base de datos propia.

Al seleccionar una base de datos que cumpla con los requerimientos, se realizará la próxima etapa, la cual consiste en la elaboración de un algoritmo que detecte rostros, con el propósito de seleccionar y segregar los puntos de interés facial de cada una de las imágenes. De esta manera se logra construir una nueva base de datos modificada de la original, en donde se eliminan elementos que no son requeridos para el entrenamiento, por último, se etiquetan de acuerdo a la expresión facial y/o estado anímico.

Con una serie de datos relevantes para un entrenamiento, se procede con el entrenamiento de diferentes algoritmos de aprendizaje de máquina, los cuales se le suministrarán diversos datos de entrenamiento para poder evaluar su desempeño y resultados, entre estos datos se encuentran los cálculos de los *momentos de Hu* [9] y los *landmarks* de cada una de las imágenes previamente etiquetadas. De esta manera se busca encontrar una precisión lo suficientemente alta que logre acercarse a los mejores algoritmos de identificación de estado anímico.

A. *Expresiones faciales destacadas*

Las emociones y/o estados en el que se encuentra un ser humano ha dejado de tener un punto de vista negativo, de ser un aspecto a reprimir o evitar ser detectado, a ser consideradas la base central de la persona, permitiendo identificar sus deseos, pensamientos, gustos y actitudes [5]. Estas emociones pueden ser detectadas de diversas maneras, desde el comportamiento hasta cambios físicos de la persona.

En busca de detectar las emociones de una persona, se han desarrollados diferentes estudios que permiten globalizarlas, entre ellos se encuentra el sistema FACS (Facial Action Coding System), el cual permite evaluar la expresión a partir de la acción combinada de una serie de unidades de acción específicas; entendiendo por unidad de acción la actividad de uno o varios músculos faciales que otorguen un cambio en la apariencia facial [6]. Puesto que el rostro posee más de 40 músculos, es factible observar los cambios físicos de este para determinar la emoción de una persona, realizando énfasis en puntos específicos que menciona el FACS, como lo es el movimiento de las cejas, la posición de los labios y hasta la apertura de párpados.

El Sistema de Codificación Facial (FACS), describe los diferentes cambios físicos en el rostro, categorizados en ocho grupos de emociones (feliz, asco, miedo, orgullo, triste, avergonzado, sorprendido, enojado), sin embargo, muchas de estas emociones pueden ser confundidas entre ellas debido a que los puntos de interés no poseen una variación significativa [7]. Por esta razón se limitan a cuatro emociones básicas (feliz, enojo, sorpresa, neutral), que pueden ser interpretadas sin la necesidad de profundizar en aspectos situacionales o factores fuera del rango físico, adicionalmente, son emociones que se maneja en cualquier cultura, puede ser identificada con facilidad y son expresadas por todo tipo de persona, sin importar su edad o estilo de vida. Estas cuatro expresiones, de persistir en el tiempo y presentarse de forma intensa durante la conducción, resultan ser un factor de riesgo durante la ejecución de esta actividad [8]. Por ello se basará en estas expresiones para entrenar un modelo con el propósito del proyecto.

B. *Selección de base de datos*

Una base de datos apropiada para el entrenamiento del algoritmo que se planea realizar, debe contener una cantidad basta de fotografías de diferentes personas en diversos estados anímicos, sin embargo, esto no basta, es requerido que las personas en esta base de datos pertenezcan a diferentes etnias, edades y género, esto debido a que una alta diversidad le permitirá al algoritmo ser más preciso, y detectar puntos faciales de diversos individuos.

Actualmente existen diversas bases de datos libres en internet, entre ellas se encuentra “JAFFE” [10], la cual es un conjunto de 217 imágenes pertenecientes a 10 modelos femeninas japonesas, las cuales presentan las expresiones fáciles basadas en las principales emociones. El hecho que represente las emociones faciales más relevantes es de gran ayuda para el entrenamiento, sin embargo, solo posee fotografías de 10 individuos, lo que limita mucho la diversidad que se busca tener, teniendo en cuenta que todas las imágenes son de cultura asiática.

Otro de las bases de datos encontradas es “FEI Face Database” [11], la cual contiene un conjunto de fotografías tomadas a 200 individuos en Brasil, en donde se posee la misma cantidad de imágenes de hombres y de mujeres, su color de piel, ojos y cabello presenta gran diversidad; Por otro lado, las expresiones faciales se limitan a felicidad y neutralidad (seriedad), por lo que limitaría el algoritmo de entrenamiento a dos emociones.

Ejemplo de esto es la base de datos “Real-world Affective Faces (RAF)” [12], la cual es una base de datos de expresiones faciales a gran escala, con alrededor de 10 mil imágenes de rostros extraídos de internet, donde poseen gran variabilidad de edad, genero, etnia, posturas de cabeza, iluminación y escenarios que permiten suministrar un alto rango de características al algoritmo computacional.

Seleccionando la base de datos “RAF”, debido a su variedad de individuos en las imágenes, se inicia con la elaboración del algoritmo de procesamiento de imágenes, el cual consiste en extraer las características más relevantes que disminuyan la información suministrada para el entrenamiento.

C. Preprocesamiento de imágenes

Las imágenes que presenta la base de datos “RAF” no necesariamente se encuentran centradas en el rostro del individuo (Fig. 1), por lo que, al suministrar este tipo de imágenes a un algoritmo de inteligencia artificial, este trataría de encontrar patrones en el fondo, ropa o elementos que no son del interés en esta ocasión. Para poder evitar este tipo de problemas, es necesario recortar la imagen para extraer únicamente el rostro, sin embargo, visto que el tamaño de estas es variable, se requiere editarlo y mantener un estándar en tamaño y tonalidad, a este proceso se le conoce como preprocesamiento de imágenes [13].



Fig. 1. Ejemplos de imágenes correspondientes a la base de datos “RAF”.
Fuente: Autores.

En la Fig. 2 se logra visualizar el cambio al que fueron expuestas las imágenes de la Fig. 1. En ella se logran apreciar el cambio de tamaño y el enfoque realizado en el rostro, sin embargo, debido a la cantidad de colores que poseen las diferentes fotografías, basta con realizar una transformada a escala de grises para unificar la tonalidad, y tener completamente normalizada la imagen (en color y tamaño). Este proceso se realiza con software libre, en lenguaje Python y la librería OpenCV.



Fig. 2. Ejemplo de uso del Algoritmo para selección y normalización de los rostros de cada fotografía de la base de datos “RAF”.
Fuente: Autores.

Una vez se encuentre la base de datos normalizada y ordenada, basta con etiquetar cada una de las imágenes, en donde se realizan grupo dependiendo de la emoción a la que pertenezca, permitiendo entrenar un algoritmo de forma supervisada.

Puesto que el modelo que se busca entrenar, requiere de un alto costo computacional, solo se usaran cuatro de las emociones etiquetadas en la base de datos seleccionada (feliz, enojo, sorpresa, neutral).

Es posible que la información aportada por cada imagen aun posea datos irrelevantes, por ello se reduce dicha información a parámetros mas precisos o simplificados, de esta forma se facilita el costo del entrenamiento.

Para realizar diferentes pruebas, se realiza una nueva base de datos en donde únicamente se encuentran los puntos de referencia facial, proporcionando información de apertura de boca, inclinación de cejas y amplitud de parpados, para ser dibujados en un lienzo negro [14].

De esta forma, se reemplaza cada una de las imágenes originales por una correspondiente a sus puntos de referencia facial (Fig. 3), creando una nueva base de datos con la que se logra realizar las pruebas correspondientes. En donde el algoritmo de aprendizaje, solo tendrá como entrada una imagen binaria (blanco y negro), reduciendo significativamente la información, resaltando solo las posiciones de las características que según los criterios FACS son fundamentales para identificar la expresión del rostro.

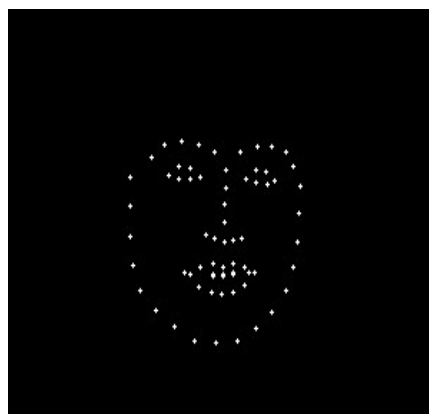


Fig. 3. Ejemplo de Landmarks en rostro (Imagen Binaria).
Fuente: Autores.

Aunque una base de datos de imágenes binarias puede representar un grupo que pueda ser aprendido fácilmente, es posible simplificar aún más la base de datos para reducir el número de entradas en el algoritmo de aprendizaje, por ello se calculan los momentos de Hu usando los puntos Landmarks (calidades como la posición de las cejas, la boca y la nariz), buscando la magnitud que representa dichas características con el propósito de reconocer un estado anímico sin importar su inclinación, rotación o tamaño de rostro, por ello se usan las imágenes binarias, las cuales tienen resaltadas las posiciones fundamentales del rostro.

D. Algoritmo clasificador

Con un total de tres bases de datos derivadas de una misma, se procede a realizar un algoritmo de inteligencia artificial que este en la capacidad de clasificar la imagen según su estado anímico. Para realizar esto, se pretende hacer diferentes pruebas con las diversas bases de datos, y determinar que procesamiento de imagen y que algoritmo clasificador es más preciso, teniendo en cuenta que cada base de datos posee las mismas etiquetas, y un aproximado de cinco mil muestras por cada uno (este número varía según el estado anímico, ya que la base “RAF” no posee la misma cantidad de muestra en cada etiqueta).

El primer paso para poder realizar el algoritmo clasificador es dividir la base de datos para entrenamiento y prueba. Esto se realiza barajando las diferentes muestras, generando un orden aleatorio, de esta manera de extraer el 25% de datos para uso de pruebas y poder obtener una medición de la precisión del algoritmo [15].

En el presente trabajo se plantean algoritmos de entrenamiento para los siguientes clasificadores: vecinos cercanos (Knn), clasificador Bayesiano (naive_bayes), máquinas de soporte vectorial (svm), árbol de decisiones (decision_tree), bosques aleatorios (random_forest), perceptrón multicapa (mlp). Cada uno de ellos fue manipulado en diferentes pruebas para generar un resultado preciso, tanteando sus características de forma adecuada.

Iniciando con la elaboración de Knn, en donde se realizaron pruebas con diferentes funciones de peso utilizadas en la predicción, adicionalmente se presentaron variaciones en el número de vecinos permitidos para su entrenamiento. Las diferentes combinaciones, permitieron una precisión mínima de 50%, sin embargo, se obtiene una mejora a 57%, aumentando su número de vecinos a 10.

Por otro parte, el clasificador Bayesiano, se diseña sin generar prioridad a ninguna de las clases, sin embargo, se ajusta la variación más grande de todas las características que se agrega para la estabilidad del cálculo en $1e-09$, permitiendo que su precisión aumente a 53%.

El algoritmo para el entrenamiento de máquinas vectoriales, se diseña probando diferentes funciones de kernel, en donde los mejores resultados se reflejan con una función Polinomial de grado 2, permitiendo elevar la precisión del algoritmo a un 72%.

En otro orden de ideas, el diseño del árbol de decisiones, se implementa con la función “gini”, la cual permite medir la cantidad de divisiones y criterios admitidos en las ramificaciones, limitando en dos, el número mínimo de muestras necesarias para su división. Sin embargo, la mejor precisión no supero el 57%.

Al igual que el algoritmo anterior, la elaboración de bosques aleatorios presentó mejores resultados con una función “gini”, sin embargo, la precisión llego a un 70% implementando un número de árboles de hasta 200 ramificaciones.

Luego de un entrenamiento, se procede a conocer su precisión y tiempo de clasificación, guardando registro de cada uno de los resultados.

TABLA 1. MACHINE LEARNING CON IMÁGENES ROSTRO A ESCALA DE GRISES.

Clasificador	Precisión	Tiempo Test (segundos)
Knn	57%	0.023
Naïve_bayes	53%	0.001
Svm	72%	0.009
Decision_tree	57%	0.001
Randomforest	69%	0.027
mlp	78%	0.002

Fuente: Autores.

TABLA 2. MACHINE LEARNING CON PUNTOS LANDMARKS.

Clasificador	Precisión	Tiempo Test (segundos)
Knn	60%	0.025
Naïve_bayes	60%	0.006
Svm	64%	0.020
Decision_tree	51%	0.001
Randomforest	63%	0.021
mlp	59%	0.048

Fuente: Autores.

TABLA 3. MACHINE LEARNING CON MOMENTOS DE HU.

Clasificador	Precisión	Tiempo Test (segundos)
Knn	43%	0.003
Naïve_bayes	50%	0.001
Svm	48%	0.001
Decision_tree	45%	0.001
Randomforest	50%	0.019
mlp	42%	0.001

Fuente: Autores.

Con resultados de aprendizaje de máquina registrados, se procede a usar un algoritmo más robusto, buscando aumentar la precisión y evaluando el tiempo de ejecución al clasificar una imagen. Se utilizará el Aprendizaje Profundo (Deep Learning) debido a que ha probado ser muy eficiente para el aprendizaje de las características y así poder clasificar imágenes, evitando pasos como la extracción de momentos de Hu u otros parámetros sobre forma, textura, color, etc., pasos que son aprendidos automáticamente en las capas convolucionales de las redes neuronales.

Para su construcción se utiliza una estructura que se ha convertido en una de las más utilizadas en aprendizaje profundo, Xception, el cual fue propuesto por el propio François Chollet, el creador y jefe de mantenimiento de la biblioteca Keras [16], la cual es una extensión de la arquitectura “Inception” que reemplaza los módulos Inception estándar con convoluciones separables en profundidad.

La arquitectura Xception tiene 36 capas convolucionales que forman la base de extracción de características donde es posible insertar capas completamente conectadas antes de la capa de regresión logística, que se explora en sección de evaluación experimental; Las 36 capas convolucionales están estructuradas en 14 módulos, todos los cuales tienen conexiones residuales lineales a su alrededor, excepto el primer y el último módulo.

La construcción de la estructura Xception en Python se realiza mediante la librería Keras, añadiendo capa tras capa utilizando la función relu como función de activación no lineal en cada capa, adicionalmente, es aquí donde se decide el número de datos de entrada y la cantidad de etiquetas de clasificación.

Finalmente se realiza una compilación o retroalimentación del sistema en busca de disminuir su error al momento de entrenar el sistema, en donde se suministra los datos de entrenamiento (base de datos y etiquetas). En este caso particular se realizan 500 iteraciones para generar el mejor resultado posible y poder evaluar la precisión por medio de los datos de prueba, sin embargo, solo se utilizará la base de datos de imágenes con rostro a escala de grises, ya que, como se mencionó anteriormente, la estructura de aprendizaje profundo aprende automáticamente los patrones o descriptores que se requieren para su clasificación.

TABLA 4. DEEPLARNING CON IMÁGENES DE ROSTRO A ESCALA DE GRISES.

Precisión	Tiempo Test (segundos)
84%	0.080

Fuente: Autores.

E. Análisis en procesamiento portátil

Se desarrollaron pruebas en dispositivo portátil para poder medir la velocidad de reacción de este con los algoritmos previamente entrenados, de esa manera sería posible determinar si los tiempos de procesamiento llevan a realizar pruebas dentro de un automóvil.

Estas pruebas se realizan en una Raspberry Pi, la cual es una pequeña computadora basada en una placa de circuito impreso y equipado con un procesador ARM, sistema de video integrado HDMI, conexión Ethernet y entradas USB para conectar periféricos y pantallas. También cuenta con un conector para tarjeta SD mediante el cual se puede dotar al Raspberry Pi de la capacidad para leer y ejecutar un sistema operativo del tipo Linux [18].

Puesto que la precisión obtenida en los algoritmos de entrenamiento usando imágenes de rostro a escala de grises, supera los resultados de los momentos de Hu e imágenes Landmarks, se procede a realizar la prueba en la Raspberry, únicamente con las imágenes originales en escala de grises, lo que permitirá descartar algoritmos de entrenamiento que requieran mayor tiempo para clasificar.

TABLA 5. PRUEBAS DE DESEMPEÑO EN RASPBERRY PI CON DEEP LEARNING.

Clasificador	Tiempo Test (segundos)
Knn	0.035
Naïve_bayes	0.086
Svm	0.060
Decision_tree	0.023
Randomforest	0.041
mlp	0.016
DeepLearning	0.499

III. ANÁLISIS DE RESULTADOS

Con los diferentes sistemas de clasificación debidamente entrenados, se procede a determinar cuál de los algoritmos presenta resultados más óptimos. Según los resultados de las Tabla 1, Tabla 2 y Tabla 3, se encuentra que los algoritmos con mayor precisión se encuentran con un entrenamiento de imágenes de rostro a escala de grises, en donde solo se extraen los rostros de cada una, dando una precisión de hasta 78% con las máquinas de soporte vectorial, la cual es un resultado aceptable, si se toma en cuenta los resultados de los concursos elaborados por “Kaggle”, recientemente, que llegan hasta un 85% [17], en donde se usa una base de datos más completa.

Aunque la precisión del algoritmo de SVM es la más alta entre las pruebas con aprendizaje de máquina, también es necesario destacar que el tiempo de ejecución al usar la base de datos de imágenes de rostro a escala de grises es superior en comparación a los demás algoritmos de clasificación con las diferentes bases de datos, lo que se genera debido a que la cantidad de información procesada es mucho mayor que si se usaran los momentos de Hu o los puntos Landmarks.

Los modelos entrenados en Machine Learning presentaron resultados favorables, sin embargo, el modelo de Deep Learning presenta más beneficios en la precisión (84%) sacrificando algo de tiempo de ejecución, aunque, este aumento en tiempo no genera ningún tipo de dificultad a la hora de ejecutar el sistema en el procesador portátil (Raspberry), como se observa en la [Tabla 5](#).

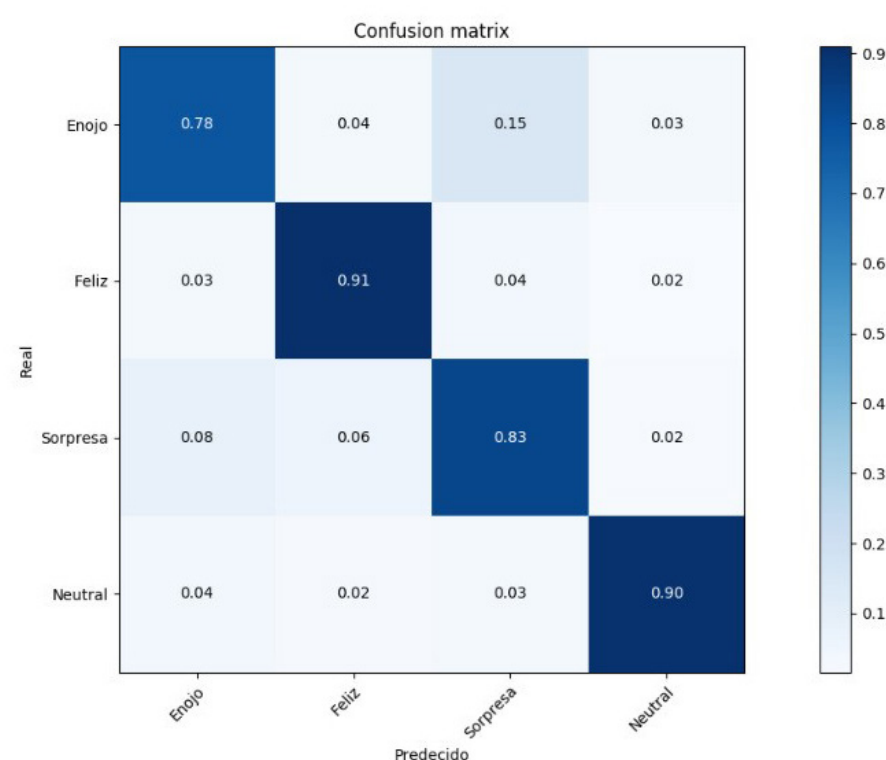


Fig. 4. Matrix de confusion Deep Learning.
Fuente: Autores.

IV. CONCLUSIONES

La base de datos más apropiada para esta aplicación fue la RAF, debido a la diversidad de individuos y condiciones en que fueron tomadas las fotografías. Sin embargo, se requirió aplicar visión artificial para preprocesar las imágenes, extrayendo sólo el rostro y convirtiéndolas a escala de grises para facilitar su análisis.

Utilizando código libre, principalmente como base el lenguaje Python y las librerías OpenCV, DLIB, Scikit-Learn y Keras, fue posible implementar una solución para detectar las cuatro expresiones faciales más importantes durante la actividad de conducción. Esto permitirá implementar este desarrollo en un equipo portable y compacto, sin requerir software licenciado, bajando los costos del producto final sin poner en riesgo su precisión y eficiencia.

De los diferentes algoritmos de aprendizaje de máquina tradicionales implementados (machine learning), el que presentó mayor precisión fue el perceptrón multicapa, alcanzando un 78% y permitiendo que, en un computador personal promedio, se puedan procesar imágenes a velocidad superior a los 30 cuadros por segundo.

En la búsqueda de una mayor precisión, el aprendizaje profundo logra llegar un 84% ([Tabla 4](#)), y aunque su costo computacional aumenta en comparación a los modelos de máquina tradicional, éste no presenta cambios significativos a la hora de indicar la predicción en un procesador portable.

Es necesario ajustar la base de datos para poseer la misma cantidad de muestras por cada etiqueta, de esta manera es posible que la precisión mejore en cada una ([Fig. 4](#)). Esto puede ser posible con la librería Keras, la cual proporciona funciones para estos casos particulares.

FINANCIAMIENTO

Este trabajo fue realizado en la fase de entrenamiento de algoritmos de aprendizaje profundo (Deep learning), utilizando una Titan XP GPU financiada en calidad de donación por NVIDIA Corporation, y recibiendo el apoyo de la universidad Politécnico Colombiano Jaime Isaza Cadavid. Año de inicio: 2019, año de finalización: 2020.

REFERENCIAS

- [1] A. González, F. J. Martínez, A. V. Pernía, F. A. Elías, M. Castejón, J. Ordieres & E. Vergara, *Técnicas y Algoritmos Básicos de Visión Artificial*. RI, Es.: U de la Rioja, 2006.
- [2] F. Chollet, “Deep Learning With Python” *Vol. 1*, GMT, USA: Manning Publications Co, 2017.
- [3] V. Torres & J. Alejandra, “Las Compañías más Grandes en Seguros Generales en Colombia: Un Análisis Comparativo en el Período 2015-2017,” *Trabajo de grado*, UC, Bog., Co., 2017.
- [4] W. Arias & S. B. Chaves, “Analysis Of Fatalities For Crash Accidents in Colombia Between 2011-2015,” *Redes de Ingenieria-Rompiendo las Barreras del Conocimiento*, vol. Especial, pp. 226–239, 2017. Available: <http://revistas.udistrital.edu.co/ojs/index.php/REDES/index>
- [5] E. G. Fernández-Abascal & M. Chóliz, *Expresión Facial de la Emoción*, Mad., Esp.: UNED, 2001.
- [6] A. Freitas-Magalhães, *Facial Action Coding System-Manual of Scientific Codification of the Human Face*. RJ. Br.: Escrytos, 2018.
- [7] R. Amini, C. Lisetti & G. Ruiz, “Hapfacs 3.0: Facs-Based Facial Expression Generator for 3d Speaking Virtual Characters,” *IEEE Trans Affect Comput*, vol. 6, no 4, pp. 348–360, 2015. <https://doi.org/10.1109/TAFFC.2015.2432794>
- [8] E. A. Morris & J. A. Hirsch, “Does Rush Hour See A Rush Of Emotions? Driver Mood in Conditions Likely To Exhibit Congestion,” *Travel Behav Soc*, vol. 5, pp. 5–13, Sep. 2016. <https://doi.org/10.1016/j.tbs.2015.07.002>
- [9] A. Carneiro, P. Cortez & R. Costa, Reconhecimento de Gestos da Libras com Classificadores Neurais a Partir dos Momentos Invariantes de Hu, in *Interaction Sudamerica/09SP*. Br.: IXTA-SP, pp. 193–198, 2009.
- [10] M. G. Kamachi, J. Gyoba & M. J. Lyons, “The Japanese Female Facial Expression (Jaffe) Database, in “*Proceedings of Third International Conference on Automatic Face and Gesture Recognition*,” Data Set, Zenodo, pp. 14–16, 14 Abr. 1998. <https://doi.org/10.5281/zenodo.3451524>
- [11] C. E. Thomaz, “Fei Face Database,” *Fei Face Databaseavailable*, 2012. Available: <https://fei.edu.br/~cet/facedatabase.html>
- [12] S. Li, W. Deng & J. P. Du, “Reliable Crowdsourcing And Deep Locality-Preserving Learning For Expression Recognition In The Wild,” presented at *Proceedings of the Ieee Conference on Computer Vision and Pattern Recognition*, CVPR, Hnl., Hi., USA, pp. 2852–2861, 21-26 Jul. 2017. <https://doi.org/10.1109/CVPR.2017.277>
- [13] N. Pellejero, G. L. Grinblat & U. Lucas, “Análisis Semántico en Rostros Utilizando Redes Neuronales Profundas,” presentado al *XVIII Simposio Argentino de Inteligencia Artificial*, ASAI-JAIIO 46, Cord., Arg., pp. 62–71, Sept. 2017. Disponible en <http://sedici.unlp.edu.ar/handle/10915/65945>
- [14] M. Köstinger, P. Wohlhart, P. M. Roth & H. Bischof, “Annotated Facial Landmarks In The Wild: A Large-Scale, Real-World Database for Facial Landmark Localization,” en *2011 IEEE International Conference on Computer Vision Workshops, ICCV Workshops*, Bcn., Es., pp. 2144–2151, 6-13 Nov, 2011. <https://doi.org/10.1109/ICCVW.2011.6130513>
- [15] S. Marsland, *Machine Learning: An Algorithmic Perspective, 2. Ed.*, CRC Press, Br. Fl. USA, 2015. <https://doi.org/10.1201/b17476>
- [16] K. Simonyan & A. Zisserman, “Very Deep Convolutional Networks For Large-Scale Image Recognition,” *Arxiv Preprint*, ICLR, Baff., Ca., pp. 1409–1556, 2014. Available: <https://arxiv.org/pdf/1409.1556.pdf>
- [17] R. S. Jurecki, T. L. Stańczyk & M. J. Jaśkiewicz, “Driver’s Reaction Time In A Simulated, Complex Road Incident,” *Transport*, vol. 32, no. 1, pp. 44–54, 2017. <https://doi.org/10.3846/16484142.2014.913535>
- [18] B. C. C. Cerezo & E. I. Ortega, “Diseño de un prototipo de casilleros inteligentes, utilizando el carnet universitario y clave de seguridad para el control de acceso a través de una red inalámbrica, mediante la administración de página web y base de datos aplicando tecnología de bajo costo (raspberry y arduino), para beneficio de los estudiantes de la carrera de ingeniería en networking y telecomunicaciones de la Universidad de Guayaquil,” *tesis grado*, dpto. FCFM, UG, Gquil., Ec., 2018. Disponible en <http://repositorio.ug.edu.ec/handle/redug/32565>

Juan Felipe Cordoba Fuzga es Ingeniero de Instrumentación y Control de la Universidad Politécnico Colombiano Jaime Isaza Cadavid. Actualmente se encuentra estudiando la Maestría en Ingeniería con énfasis en Automatización en la misma institución. En este tiempo se dedicó a la Automatización de procesos industriales, y actualmente se encuentra trabajando en Sistemas de inteligencia artificial y manejo de base de datos en el área en la que se ha enfocado.

Rubén Vásquez Salazar es Ingeniero de Control de la Universidad Nacional de Colombia (Medellín), y posteriormente el título de Magíster en Ingeniería – Automatización Industrial en el año 2009 de la misma universidad sede Manizales. Su investigación actual se centra principalmente en los sistemas de visión artificial, especialmente investigación aplicada para la solución de problemas de la sociedad y el país. <https://orcid.org/0000-0002-1690-8393>

Henry Sarmiento Maldonado es Ingeniero Electricista, Especialista en Ciencias Electrónicas e Informática (énfasis en Automatización Industrial), Magíster en Ingeniería (énfasis en Energética) y Doctor en Ingeniería Electrónica (línea de énfasis en Automatización y Control) de la Universidad de Antioquia (Colombia). Docente ocasional de la Universidad de Antioquia desde el año 1996 hasta el año 2007, y actualmente docente asociado del Politécnico Colombiano Jaime Isaza Cadavid donde actualmente es Director del Grupo de Investigación en Instrumentación, Control Automático y Robótica ICARO. Áreas de interés: técnicas de inteligencia artificial, y el control y automatización de procesos. <https://orcid.org/0000-0001-8011-1293>