

MÉTODO PARA IDENTIFICAR ESTADOS COGNITIVOS A PARTIR DE DATOS MULTIMODALES DE COMPORTAMIENTO APLICANDO TÉCNICAS DE VISIÓN ARTIFICIAL Y ALGORITMOS DE APRENDIZAJE AUTOMÁTICO SUPERVISADO

METHOD FOR IDENTIFYING COGNITIVE STATES FROM MULTIMODAL BEHAVIORAL DATA USING COMPUTER VISION TECHNIQUES AND SUPERVISED MACHINE LEARNING ALGORITHMS

Maricarmen Toribio-Candela ^a, Gabriel González-Serna ^b, Andrea Magadan-Salazar ^c, Nimrod González-Franco ^d, Máximo López-Sánchez ^e

Abstract:

Human intelligence is a psychological quality that encompasses learning from experience, adapting to new circumstances, processing complex abstract concepts, and using knowledge to interact with and modify the environment. These cognitive states are reflected in responses that influence human behavior, manifesting through facial expressions, body movement, and emotional reactions to situations that impact cognitive stability. Including cognitive state detection during user experience (UX) assessment represents a valuable opportunity to improve the efficiency and quality of products or services. The multimodal extraction strategy includes detecting 46 points related to head movement, hand position, and facial expressions, three supervised machine learning algorithms Random Forest, KNN, and SVM were analyzed. Two image datasets were used for training, Cam3D and Pandora, obtaining an accuracy of 98% with Random Forest, 97% with KNN, and 95% with SVM, for the detection of three cognitive states, attention, concentration, and distraction.

Keywords:

UX, Machine Learning, Facial recognition, Multimodal data, cognitive states

Resumen:

La inteligencia humana es una cualidad psicológica que abarca la capacidad de aprender de la experiencia, adaptarse a nuevas circunstancias, procesar conceptos abstractos complejos y emplear el conocimiento para interactuar con el entorno y modificarlo. Los estados cognitivos se reflejan en respuestas que influyen en el comportamiento humano, manifestándose a través de expresiones faciales, movimiento corporal y reacciones emocionales ante situaciones que impactan la estabilidad cognitiva. La inclusión de la detección de estados cognitivos durante la evaluación de la experiencia del usuario (UX) representa una oportunidad valiosa para mejorar la eficiencia y calidad de productos o servicios. La estrategia de extracción multimodal incluye la detección de 46 puntos relacionados con movimiento de cabeza, posición de manos y expresiones faciales, se analizaron tres algoritmos de aprendizaje automático supervisado, Random Forest, KNN y SVM. Para el entrenamiento se utilizaron dos data-set de imágenes Cam3D y Pandora, obteniendo una precisión del 98% con Random Forest, 97% con KNN y 95% con SVM, para la detección de tres estados cognitivos atención, concentración y distracción.

Palabras Clave:

UX, Aprendizaje Máquina, Reconocimiento facial, Datos Multimodales, estados cognitivos

^a Autor de Correspondencia, TecNM-CENIDET, <https://orcid.org/0009-0004-9089-0058>, Email: m22ce054@cenidet.tecnm.mx

^b TecNM-CENIDET, <https://orcid.org/0000-0002-1874-9402>, Email: gabriel.gs@cenidet.tecnm.mx

^c TecNM-CENIDET, <https://orcid.org/0000-0001-5474-4150>, Email: andrea.ms@cenidet.tecnm.mx

^d TecNM-CENIDET, <https://orcid.org/0000-0002-1051-1379>, Email: nimrod.gf@cenidet.tecnm.mx

^e TecNM-CENIDET, <https://orcid.org/0000-0002-4130-3564>, Email: maximo.ls@cenidet.tecnm.mx

Introducción

De manera natural, el ser humano expresa emociones y estados cognitivos mediante expresiones corporales, sin embargo, el reconocimiento de las expresiones basadas en su comportamiento relacionadas con estados cognitivos, como la concentración, la atención y la distracción, el reconocimiento automático mediante técnicas de visión artificial es un problema desafiante, de manera particular, porque requiere métodos para procesar y analizar datos biométricos multimodales (expresiones faciales, postura de la cabeza y de las manos).

El estado cognitivo de un usuario se refiere a los procesos mentales relacionados con el aprendizaje, tales como, la memoria, la atención, el lenguaje, la resolución de problemas y la planificación, a partir de estos procesos es posible identificar varios estados cognitivos tales como: la carga cognitiva, el estrés mental, la fatiga mental, la concentración, la atención, la distracción, entre otros, los cuales incluye variables multimodales que describen procesos cognitivos, como el movimiento ocular, dilatación de la pupila, expresión facial, postura de la cabeza y/o movimiento de manos.

En el contexto del reconocimiento y clasificación de estados cognitivos utilizando algoritmos de aprendizaje automático (ML) y aprendizaje profundo (DL), se realizó una revisión sistemática de la literatura (RSL), en la que se analizaron los trabajos [1] - [20]. En (1) el análisis del movimiento de las manos, los movimientos de la cabeza y los ojos son fundamentales para el reconocimiento de estados cognitivos en conjunto con las expresiones faciales. El estado del arte se divide en dos secciones: 1) Detección de datos multimodales y 2) Reconocimiento de estados cognitivos.

En la primera sección, se analizaron los artículos (2), (3) y (4) que proporcionaron ideas relevantes para la localización y el preprocesamiento de análisis multimodal en función de la calidad de las bases de datos, extractores de características y clasificadores de patrones. En el segundo apartado del análisis del estado del arte, se estudiaron los trabajos [1] - [20], su objetivo principal fue realizar el reconocimiento de estados cognitivos. Los trabajos que destacaron utilizaron técnicas de preprocesamiento y algoritmos de ML y DL. Estos trabajos ayudaron a definir la arquitectura del sistema de clasificación de estados cognitivos. Trabajos como (5) y (6) reportan haber implementado arquitecturas de redes neuronales demostrando que esta técnica no mejora el resultado, ya que tienden a igualarse en cuanto a resultados mostrados en trabajos que utilizaron ML. Analizando los resultados de las técnicas implementadas, se concluyó que utilizar redes neuronales implica un mayor costo computacional, por lo que para esta

investigación se optó por utilizar algoritmos clásicos de ML; adicionalmente se consideraron los trabajos (1), (7) y (8) que utilizan técnicas geométricas para el reconocimiento de estados cognitivos.

En los trabajos [1] - [20], existe una tendencia a explorar las diferencias de expresión principalmente en el espacio de píxeles de la imagen facial completa. En este enfoque, la expresión facial se considera como una característica expresiva global. Sin embargo, algunos investigadores han buscado mejorar la precisión de la identificación de expresiones faciales mediante técnicas complementarias, en las que seleccionan áreas faciales específicas y aplican métodos de extracción de características locales para calcular descriptores específicos para estas áreas, tales métodos incluyen las cascadas de Haar (9) y el uso del histograma de gradientes orientados (HOG) (8). Estos enfoques se centran en características locales de la cara, lo que puede ayudar a capturar detalles específicos de las expresiones faciales y mejorar la precisión del reconocimiento.

Es importante destacar que el método basado en la apariencia, si bien puede capturar información detallada y sutil de la expresión facial, la postura de la cabeza y la posición de las manos presenta ciertas limitaciones. Estas características son altamente sensibles a factores como la luminancia y la oclusión, lo que puede dificultar su fiabilidad en algunas situaciones.

Por otro lado, la segunda categoría de métodos se puede considerar como un enfoque basado en la geometría. Este enfoque a menudo implica la marcación de puntos de rasgo facial, cuyos movimientos correspondientes pueden ser útiles para capturar rasgos expresivos faciales y corporales. Esto es similar a lo que se ha utilizado en referencias anteriores (10) y (11) para el análisis de expresiones faciales y posturas corporales. Este enfoque basado en la geometría puede ofrecer una mayor robustez en términos de la influencia de factores externos como la luminancia y la oclusión.

El identificar automáticamente estados cognitivos expresados por el usuario de manera inconsciente, proporciona información valiosa para mejorar los niveles de usabilidad, en comparación con los métodos convencionales basados en cuestionarios auto informados, en donde los resultados dependen de la capacidad del usuario para evaluar pos-interacción y de la interpretación del evaluador de UX, como se muestra en (figura 1).

Por lo anterior, el objetivo de esta investigación es desarrollar un método para identificar automáticamente estados cognitivos durante el proceso de evaluación UX, la cual, se define en el estándar ISO 9241-210, como "Las percepciones y respuestas de un usuario que resultan del uso o uso anticipado de un sistema, producto o servicio; involucra varias facetas (instrumental,

emocional-afectiva, cognitiva y experiencia) y es consecuencia del estado interno del usuario, las características del sistema o aplicación y el contexto donde la interacción ocurre".

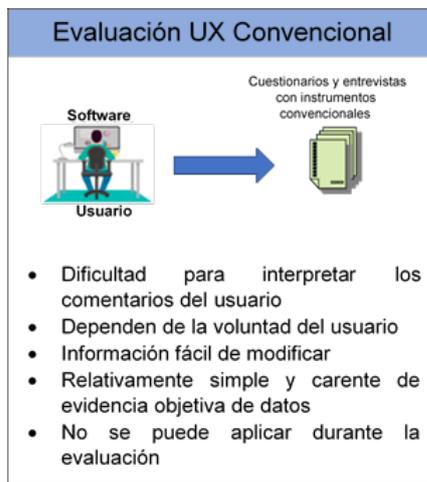


Figura 1. Herramientas utilizadas en una evaluación UX convencional

Identificar estados cognitivos en el proceso UX permite identificar qué tan fácil es utilizar un producto digital, lo cual está directamente relacionado con la carga mental, lo cual se puede traducir en niveles de satisfactoria y en última instancia, a productos digitales que se alinean de manera óptima con las necesidades y preferencias del usuario.

En resumen, en este trabajo se evalúa e implementa una solución innovadora en el área de la UX, mediante el análisis de la expresión facial, la posición de las manos y la postura de la cabeza, ya que se complementa con las unidades de acción (AU) de Ekman y se aplican en la identificación de tres estados cognitivos específicos. El propósito de este enfoque es evaluar la experiencia del usuario (UX) en relación con la interacción con un producto digital para obtener información sobre su impacto cognitivo. Esta información es de gran utilidad para los evaluadores de UX, ya que les permite identificar aspectos positivos o negativos relacionados con la interfaz de usuario (UI). Para validar los resultados de esta investigación, se realiza una evaluación del rendimiento de tres algoritmos de ML utilizando métricas de reconocimiento tradicionales.

Método

Las expresiones corporales se consideran como la principal fuente visual para mostrar estados cognitivos, además, es el componente más importante y complejo de la comunicación no verbal. Los estados cognitivos son un conjunto de destrezas y competencias que permiten al ser humano esbozar y establecer relaciones lógicas (12). Jaume y Munar (2014) Plantea que la atención es

fundamental para la interacción con el mundo exterior y para comprender cómo se organizan los pensamientos y acciones en consecuencia. (Cañizares, 2009) describe el estado de concentración como la intensidad con la que una persona concentra su atención en una actividad, objeto o tarea determinada, haciendo abstracción de todo lo demás. (DeGroot, 1990) describe la distracción cuando gastamos recursos y esfuerzo para enfocar nuestra atención en estímulos relevantes y luego tenemos dificultades para volver a nuestra tarea original, lo que reduce nuestro rendimiento.

Unidades de acción

Los componentes del reconocimiento facial observable se conocen como Unidades de Acción (UA). De esta manera, la expresión facial de un estado cognitivo se puede descomponer en una o varias UA que representan diferentes aspectos de uno o varios estados cognitivos.

La forma más efectiva para identificar un estado cognitivo en el rostro humano es a través de un sistema de codificación de acciones faciales (AF) que relaciona el movimiento de regiones específicas de del rostro con identificadores (UA), permitiendo que los algoritmos de clasificación (ML) aprendan de esta codificación.

El sistema de codificación de AF se ha establecido como un estándar para medir las expresiones faciales. El psicólogo Paul Ekman y Wallace V. Friesen adoptaron una taxonomía de movimientos faciales humanos conocida como el Sistema de Codificación de Acciones Faciales (FACS) (13).

Este sistema se basa en un método originalmente desarrollado por el anatomista sueco Carl-Herman Hjortsjö. FACS se ha convertido en una herramienta comúnmente utilizada para categorizar de manera sistemática las expresiones faciales de estados mentales (emociones y cognitivas) y ha demostrado ser valioso tanto para psicólogos como para animadores.

Hand-over-face

Durante el proceso de comunicación Humano-Humano, las manos se utilizan para comunicarse de manera no verbal. Esto abarca desde acciones simples, como señalar objetos, hasta gestos más complejos que expresan emociones y estados cognitivos. Además, acercar las manos al rostro, se ha empleado para inferir estados cognitivos.

(Ambady y Rosenthal 1992) determinan que en el proceso de comunicación el 35% implica interpretar expresiones faciales y movimiento corporal, postura del cuerpo y movimiento de manos. La interpretación humana de las diferentes interacciones sociales en una variedad de situaciones es más precisa cuando las personas pueden observar tanto la cara, como la postura y movimiento del cuerpo (Ekman y Friesen 1969).

Puntos de referencia facial

Un aspecto clave para el desarrollo del método desarrollado en esta investigación para identificar de manera automática estados cognitivos, se enfoca en la descripción detallada del rostro.

Esto implica la identificación precisa de puntos de referencia en la cara, como los ojos, la nariz, la boca y otras estructuras faciales. Además, se consideran tareas adicionales, como la alineación facial, la estimación de la postura de la cabeza y la identificación de expresiones faciales relacionadas con estados cognitivos (14).

Solución Propuesta

Para el desarrollo del método, se propuso el esquema de solución que se muestra en (Figura 2), el cual consta de 8 etapas, las primeras cinco definen el método para la extracción de características basado en el cálculo de distancias geométricas y las otras 3 etapas corresponden al entrenamiento, validación y prueba del algoritmo de clasificación. Cada una de las etapas propuestas se describe a continuación.

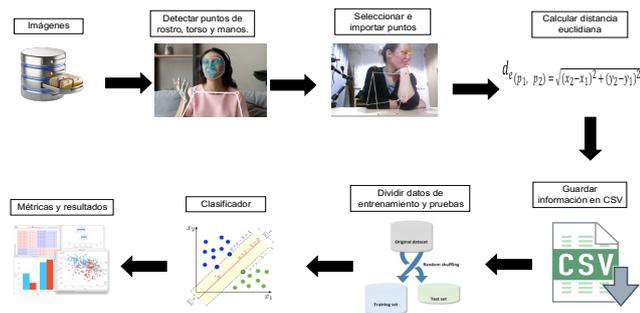


Figura 2. Esquema del sistema de clasificación de estados cognitivos.

Etapas 1. Imágenes

En esta etapa, se realizó una revisión sistemática de la literatura (RSL) para identificar y seleccionar data-sets públicos de imágenes relacionadas con estados cognitivos. Se identificaron y seleccionaron dos, Cam3D (15) y Pandora (16). La Tabla 1 muestra las características de los conjuntos de datos seleccionados respectivamente de cada uno de los data-set.

Tabla 1. Conjuntos de datos para entrenar y validar los algoritmos.

Data-set	Estados cognitivos	Imágenes por clase
Cam3D	Atención, distracción, concentración	600
Pandora	Atención y distracción	700

En el proceso de análisis de cada data-set, se identificó variación lumínica, por otro lado, se tomó en cuenta del torso a la cabeza, las imágenes se tomaron en ambientes controlados con luminosidad suficiente para poder reconocer las expresiones faciales. Los ejemplos de cada conjunto de datos se muestran en (Figuras 3 y 4).



Figura 3. Estados cognitivos atención, distracción y concentración respectivamente del conjunto de datos Cam3D (15).



Figura 4. Estados cognitivos atención, distracción y concentración respectivamente del conjunto de datos Pandora (16).

Cada data-set de imágenes se dividió en carpetas con una etiqueta numérica del 0 al 2 que corresponde a los 3 estados cognitivos: 0 = Atención, 1 = Distracción, 2 = Concentración.

El sistema inicia su proceso localizando el rostro, torso y mano en una imagen para así extraer los rasgos necesarios para el reconocimiento de los estados cognitivos. Para llevar a cabo esta detección, las imágenes son convertidas a niveles de gris, lo que implica la eliminación de la información de color. Luego, se utiliza un detector de puntos de referencia holístico previamente entrenado, el cual forma parte de la biblioteca MediPipe (17). Esta biblioteca, presentada por (18), está diseñada para estimar la ubicación de 468 puntos o UA en las estructuras faciales de la cara, 21 puntos en la estructura de cada mano y 33 puntos en la estructura de la pose de la cabeza. Estos puntos de referencia son esenciales para la posterior descripción y análisis de la cara, las manos y el hombro en la imagen, basándose en distancias geométricas entre ellos. La transformación de la imagen a niveles de gris y la detección de puntos de referencia permiten al sistema capturar información relevante para el reconocimiento de estados cognitivos.

La descripción de las expresiones faciales se calcula mediante las distancias euclidianas de ciertos movimientos musculares especificados por Ekman (13) relacionada con la representación de cada expresión para identificar los estados cognitivos. Para llevar a cabo este proceso, las coordenadas de los rasgos faciales, posición de la mano y posición de cabeza, se definen de la manera: [punto de partida, punto final]. Es decir, para que el detector haga la predicción de los puntos, las imágenes

en escala de grises se pasan como parámetro de entrada, se ubica rostro, mano y hombro obteniendo los puntos predictores en las partes de interés del usuario que está presentando un estado cognitivo en la (Figura 5).

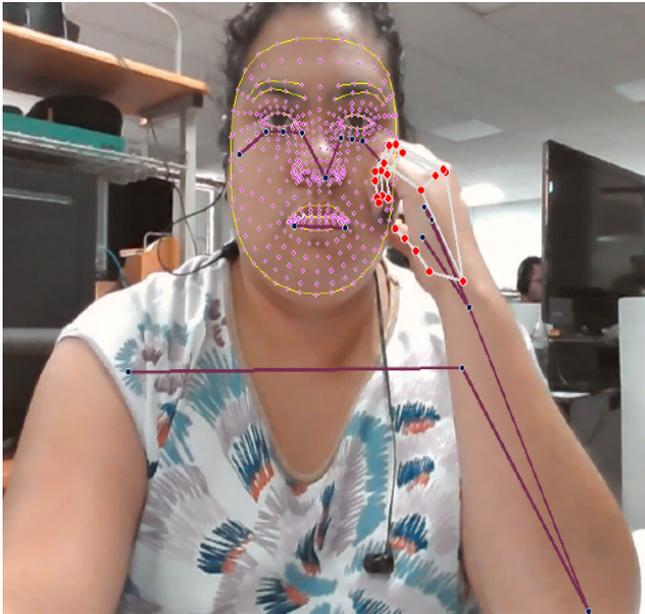


Figura 5. Los 543 puntos colocados en la cara, mano y torso que presentan estado cognitivo atención.

Etapas 2 y 3. Selección de puntos característicos importantes.

En esta etapa se seleccionaron 47 puntos clave de un conjunto de 543 para calcular distancias esenciales. Estas distancias se almacenan en un archivo para identificar estados cognitivos. Además, calculamos el ángulo de rotación de la cabeza usando puntos faciales y puntos corporales, al igual que evaluamos la posición de las manos. Estos puntos desempeñan un papel central en nuestro análisis, los puntos se muestran en la (Tabla 2).

Tabla 2. Conjuntos de datos para entrenar y probar los algoritmos

No.	Regiones de la cara	Puntos faciales
1	Parpado Derecho	[258,386,259,295,283]
2	Parpado Izquierdo	[29,159,53,28,66]
3	Ceja derecha	[276,342,384,441,285,8]
4	Ceja izquierda	[46,113,157,221,55,8]
5	Boca	[11,16, 2,164]
6	Nariz	[6]
7	Mano izquierda	[12,0]
8	Mano Derecha	[12,0]
9	Pose	[12,11,13,14,0]
10	Ojos	[145,159,374,386,474,476, 362,469, 471,133]

Los puntos seleccionados pueden visualizar (Figura 6).

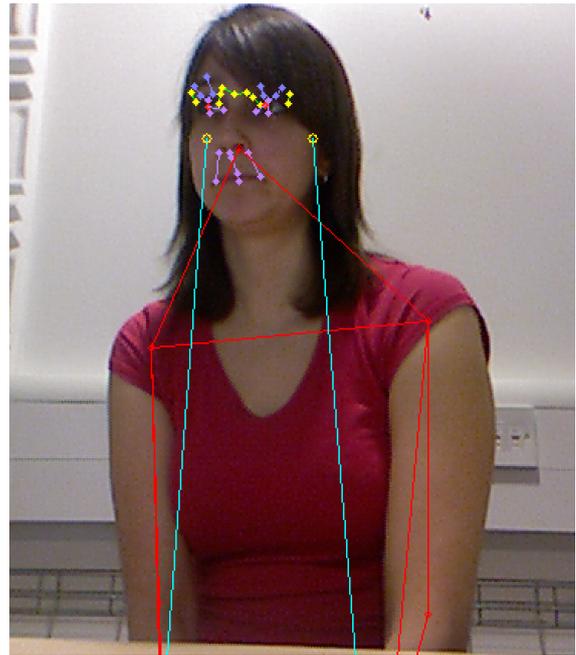


Figura 6. Los 47 puntos de referencia faciales y corporales colocados en la cara y cuerpo de una de las personas de la base de datos Cam3D (15) expresión de concentración.

La selección de estos puntos se basó en el sistema de codificación de acción facial para estados cognitivos propuesto por Ekman, que proporciona las UA que caracterizan cada estado cognitivo, las unidades de acción se refieren al hecho de que Ekman asignó un identificador con las siglas AU# a los movimientos de cada uno de los músculos de la cara que corresponden a cada emoción, por ejemplo, en el estado cognitivo atención, están involucradas las unidades de acción AU1, AU2, AU5 y AU17.

El primero corresponde al levantamiento de la ceja exterior, el segundo corresponde a levantar ceja interior, el tercero corresponde a elevar el parpado superior y el cuarto corresponde a fruncir el mentón.

Las Tablas 3, 4 y 5 muestran la relación de las unidades de acción (UA) con puntos de referencia faciales seleccionados para los estados cognitivos atención, concentración y distracción.

Tabla 3. Relación entre puntos faciales y unidades de acción del estado cognitivo atención

No.	Combinación	Puntos faciales
1	Ceja_superior_izq	[8,55,157,221]
2	Ceja_superior_der	[8,285,384,441]
3	Ceja_exterior_izq	[46,113]
4	Ceja_exterior_der	[342,276]
5	Parpado_superior_izq	[28,29,53,66,159]

6	Parpado_superior_ der	[258,259,283,295,386]
7	Mentonera	[2,11,164]
8	comisura_der	[291,327]
9	comisura_izq	[61,98]
10	Mano_izq	[8,22,12,14,16]
11	Mano_der	[21,7,11,13,15]

Tabla 4. Relación entre los puntos faciales y las unidades de acción de estado cognitivo distracción

No.	Combinación	Puntos faciales
1	Labios	[11,16]
2	Ojo_der	[386, 374]
3	Ojo_izq	[159,145]
4	cabeza	[12,0,11]
5	Mano_izq	[8,22,12,14,16]
6	Mano_der	[21,7,11,13,15]
7	Iris_izq	[469,471,133]
8	Iris_der	[474, 476, 362]

Tabla 5. Relación entre los puntos faciales y las unidades de acción de estado cognitivo concentración

No.	Combinación	Puntos faciales
1	boca	[11,16]
2	Ojo_der	[386, 374]
3	Ojo_izq	[159,145]
4	cabeza	[12,0,11]

Etapa 4. Cálculo de la distancia euclidiana.

Una vez identificadas las UA's, se requiere calcular la distancia entre estos puntos utilizando la ecuación de distancia euclidiana indicada en la fórmula 1, obteniendo 31 combinaciones de puntos, que interfieren en cada una de las Unidades de Acción de cada emoción a reconocer, determinadas por el psicólogo Paul Ekman, en su Sistema de Codificación de la Acción Facial [24].

$$d_e(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

Donde $d_e(p_1, p_2)$ es la distancia que se pretende calcular desde un punto característico del (rostro, mano o cabeza) hasta otro punto característico del (rostro, mano o cabeza); por ejemplo, para calcular la longitud de abertura de ojo izquierdo, se utilizó la distancia entre el punto 159 y 145 como se ve en (Figura 7).

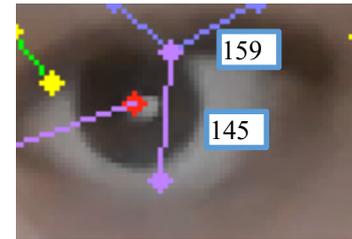


Figura 7. Puntos 159 y 145 del ojo izquierdo.

Etapa 5. Almacenar la información en un archivo CSV.

Una vez realizado el cálculo de las distancias entre los puntos, se obtiene un vector que se almacena en un archivo con el que posteriormente se entrenará el algoritmo de clasificación. El conjunto de datos contiene 31 combinaciones de puntos que serán las características y una columna llamada clase que es la etiqueta del estado cognitivo que corresponde a cada imagen.

Etapa 6. Entrenar al clasificador.

Como resultado de la revisión sistemática de la literatura, se identificaron y seleccionaron tres algoritmos para realizar la evaluación, los cuales fueron Random Forest, SVM y KNN. Random Forest es una técnica de conjunto (ensemble) que combina múltiples árboles de decisión para mejorar la precisión y la robustez del modelo. Los árboles de decisión se utilizan para dividir los datos en subconjuntos en función de las características para hacer predicciones (19). El algoritmo KNN vecinos más cercanos o K-Nearest Neighbors es un algoritmo de Machine Learning (ML) que pertenece a los algoritmos de aprendizaje supervisado simples y fáciles de aplicar que pueden ser utilizados para resolver problemas de clasificación y de regresión, busca los "k" puntos de datos más cercanos en función de una métrica de distancia, como la distancia euclidiana. Estos "k" vecinos más cercanos votan para determinar la clase del nuevo punto de datos (20). Support Vector Machine (SVM), es un algoritmo de aprendizaje automático que se utiliza tanto para clasificación como para regresión. El objetivo principal es encontrar un hiperplano que maximice la separación entre las clases de datos. Un hiperplano es una superficie en un espacio de características que divide los puntos de datos de diferentes clases (21).

Para comenzar el proceso de entrenamiento de los algoritmos de clasificación, se procedió a ingresar un conjunto de imágenes previamente etiquetadas del dataset Cam3D el cual contiene 600 imágenes por cada estado cognitivos, para cada clase se tomaron 350 imágenes, para evitar un desequilibrio de clases en el data-set, posteriormente se establecieron los hiperparámetros óptimos para cada clasificador que

permitieron obtener una precisión del 98% con Random Forest, 97 % con KNN y 95% con SVM. Los tres algoritmos de clasificación se entrenaron con el mismo conjunto de datos. La tabla 6 muestra los resultados obtenidos para cada algoritmo.

Resultados

Esta sección presenta la precisión obtenida de cada algoritmo clasificador, ver Tabla 6 y la comparación con otros trabajos similares, ver Tabla 7.

Tabla 6 Precisión de cada algoritmo.

Algoritmos	Exactitud
Soporte Vector Machine	99.7 %
Random Forest	99.6 %
K-Vecinos más cercanos	99.6 %

Tabla 7 Comparación de exactitud de clasificación de estados cognitivos con otros autores

Autor	Algoritmo	Exactitud
(1)	SVM	90.51%
(7)	SVM	95.74%
(8)	Naive Bayes	
	SVM	95%
	SVM	
(22)	Random Forest	80%
	KNN	
	SVM	99.7%
Seleccionados	Random Forest	99.6%
	KNN	99.6%

Métricas.

Las métricas que se aplicaron en este experimento fueron precisión, Recall, F1-Score y AUC. La tabla 8 muestra los resultados obtenidos de cada métrica, con cada uno de los algoritmos analizados.

Tabla 8. Resultados de métricas de cada algoritmo

Modelo	AUC	F1-Score	Precisión	Recall
SVM	99.7	99.7	99.7	99.7
Random Forest	99.8	99.6	99.6	99.6
KNN	99.8	99.6	99.6	99.6

Como se puede observar en la figura (8) área bajo la curva (AUC), se obtuvieron buenos resultados con los tres algoritmos, pero, aunque se obtuvieron buenos resultados, se puede identificar el mejor algoritmo de clasificación que obtuvo mejores resultados en todas las

métricas fue el clasificador SVM. El desarrollo se llevó a cabo utilizando la librería OpenCV, MediaPipe, y el lenguaje de programación Python para la obtención de los datos. Se utilizó Orange para obtener las métricas de clasificación de cada algoritmo seleccionado.

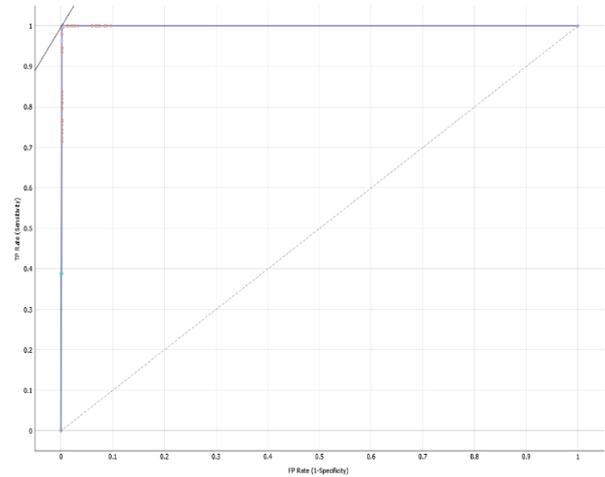


Figura 8. AUC de los algoritmos SVM, Random Forest y KNN.

Conclusión

En este trabajo, se describe el desarrollo de un método prometedor para el reconocimiento de estados cognitivos a través del análisis de datos multimodales, expresiones faciales, posición de manos y posición de cabeza, tomando como referencia algunos puntos clave en la mano, la cara y el cuerpo. El método fue evaluado utilizando el conjunto de imágenes, Cam3d, los resultados obtenidos hasta el momento son alentadores, demostrando mejorar los resultados reportados en la literatura para la detección de estos estados cognitivos. Sin embargo, es importante destacar que aún queda un importante paso por dar: la validación y análisis del sistema en un entorno diferente utilizando videos en lugar de imágenes estáticas.

Este proceso es crucial para determinar si nuestro método mantiene la misma precisión o si podríamos enfrentar desafíos adicionales en un contexto más dinámico. Este será un foco clave de las actividades futuras. Se plantea adaptar y probar el método en diversos entornos y condiciones para evaluar su robustez y versatilidad. Una vez completada esta fase de experimentación y validación, tenemos la intención de implementar el sistema en la plataforma de evaluación de la experiencia del usuario conocida como QUXLab. Esto permitirá identificar estados cognitivos en los usuarios que participan en la evaluación de la experiencia del usuario de productos digitales, lo que podría tener un impacto significativo en la comprensión y mejora de la interacción humano-computadora en diferentes contextos.

Agradecimientos

El presente trabajo fue realizado con el apoyo del proyecto modalidad investigador de la convocatoria 2023 del Tecnológico Nacional de México número 16973.23-P.

Referencias

- [1] Ardendu Behera, P. M. (2020). Associating Facial Expressions and Upper-Body Gestures with Learning Tasks for Enhancing Intelligent Tutoring Systems. *International Journal of Artificial Intelligence in Education*, 236-270 <https://doi.org/10.1007/s40593-020-00195-2>.
- [2] Ashwin T. S., R. M. (2019). Automatic detection of students' affective states in classroom environment using hybrid convolutional neural networks. *Springer Science+Business Media*, Doi : s10639-019-10004-6.
- [3] DataScientest. (enero de 2023). <https://datascientest.com/es/random-forest-bosque-aleatorio-definicion-y-funcionamiento>.
- [4] Ekman, P. &. (1978). Facial Action Codign System (FACS). <https://doi.org/10.1037/t27734-000>.
- [5] Francesca D'Erricoa, M. P. (2018). Cognitive Emotions in E-Learning Processes and Their Potential Relationship with Students. *International Journal of Emotional Education*, 89-111 ISSN 2073-7629.
- [6] Gama Velasco, A. (2018). Desarrollo de un Sistema interactivo basado en vision artificial para la rehabilitacion del tobillo. Morelos, Cuernavaca: CENIDET.
- [7] H.S Gunavathi, M. e. (2018). Towards Cognitive State Detection Using Facial Expression And Hand-Over-Face Gesture. *IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology*. Bangalore, India: IEE.
- [8] <https://datascientest.com/es/que-es-el-algoritmo-knn>. (Diciembre de 2022). *DataScientest*.
- [9] <https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/>. (junio de 2017). *monkeylearn*.
- [10] <https://psy.takelab.fer.hr/datasets/all/pandora/>. (s.f.). *PSY.TXT*.
- [11] Hu, G. L. (2018). Deep Multi-Task Learning to Recognise Subtle Facial Expressions of Mental States. *Conference on Computer Vision* (págs. 106 - 123). European: https://doi.org/10.1007/978-3-030-01258-8_7.
- [12] Indhumathi R, G. A. (2019). Emotional Interfaces for Effective E-Reading using Machine Learning Techniques. *International Journal of Recent Technology and Engineering (IJRTE)*, 4443-4449 ISSN: 2277-3878 (Online) DOI:10.35940/ijrte.D8391.118419.
- [13] King, D. (2015). *dlib C++ Library*. Obtenido de www.Dlib.Net
- [14] Mahdi Pourmirzaei, G. A. (2021). Customizing an Affective Tutoring System Based on Facial Expression and Head Pose Estimation. *Human Computer Interaction Cornell university*, 17.
- [15] MediaPipe. (2020). Obtenido de <https://developers.google.com/mediapipe>
- [16] MediaPipe. (s.f.). https://developers.google.com/mediapipe/solutions/vision/face_landmarker.
- [17] Shan Li, S. P. (2020). Automated Detection of Cognitive Engagement to Inform the Art of Staying Engaged in Problem-solving. *Journal Pre-proof*, DOI: <https://doi.org/10.1016/j.compedu.2020.104114>.
- [18] Shivalila Hangaragi, T. S. (2023). Face Detection and Recognition Using Face Mesh and Deep Neural Network. *Department of Electronics and Communication Engineering, Amrita School of Engineering, Bengaluru, Amrita Vishwa Vidyapeetham, India*, 741-723; Doi: 10.1016/j.procs.2023.01.054.
- [19] Tzvi Michelson, H. A. (2022). A Peek at Peak Emotion Recognition. <https://arxiv.org/abs/2205.09791>.
- [20] Ullari, J. (2013). Procesos Metacognitivo del Pensamiento Lógico Matemático:Razonamiento Hipotético. Ecuador: Universidad de Cuenca.
- [21] Universidad de Cambridge, R. U. (2011). *Corpus 3D de estados mentales complejos espontáneos*. (Computación Afectiva e Interacción Inteligente) Obtenido de <https://www.cl.cam.ac.uk/research/rainbow/projects/cam3d/>.
- [22] Valderrama Cárdenas, W. (2019). Reconocimiento automático del rostro para verificación de identidad para evaluación en línea. *Tesis de Maestría*. Cuernavaca, Morelos., México: CENIDET.