

Método para identificar estados mentales aplicando técnicas de visión artificial y aprendizaje supervisado

Gabriel González-Serna
Departamento de Ciencias
Computacionales
TecNM/CENIDET
Cuernavaca, México
gabriel.gs@cenidet.tecnm.mx

Luis Hernández-Pérez
Departamento de Ciencias
Computacionales
TecNM/CENIDET
Cuernavaca, México
m20ce038@cenidet.tecnm.mx

Andrea Magadan-Salazar
Departamento de Ciencias
Computacionales
TecNM/CENIDET
Cuernavaca, México
andrea.ms@cenidet.tecnm.mx

Nimrod González-Franco
Departamento de Ciencias
Computacionales
TecNM/CENIDET
Cuernavaca, México
nimrod.gf@cenidet.tecnm.mx

Olivia Frago-Díaz
Departamento de Ciencias
Computacionales
TecNM/CENIDET
Cuernavaca, México
olivia.fd@cenidet.tecnm.mx

Juan Gómez-Ramírez
Departamento de Ciencias Básicas
TecNM/ITAcapulco
Acapulco, México
juan.gr@acapulco.tecnm.mx

Resumen— Las expresiones faciales resultan de gran importancia cuando se trata de la interacción entre personas, ya que revelan estados emocionales y cognitivos complejos, que son visualmente expresados. El rostro humano es considerado el principal sistema visual para mostrar emociones, además de ser el área visual más importante y compleja de la comunicación no verbal y la que más se observa y analiza durante cualquier tipo de interacción social. Este trabajo presenta el resultado de la implementación de un método para el reconocimiento de emociones mediante el análisis de expresiones faciales y algoritmos de aprendizaje automático supervisado. Se utilizó un modelo pre-entrenado para identificar emociones que identifica 68 puntos en el rostro, de los cuales se seleccionaron 17 para calcular la distancia geométrica entre puntos específicos. El algoritmo de clasificación se entrenó con un data-set propio. Las pruebas para validar el método desarrollado en esta investigación se realizaron con el conjunto de datos CK+ que se encuentra en la literatura, así como con videos. Los resultados mostraron una precisión del 84,52% en el reconocimiento de cuatro emociones.

Keywords— classifier, Emotions, artificial vision, Facial Action Code System, facial recognition.

I. INTRODUCCIÓN

A. Antecedentes

Durante el proceso de comunicación persona-personas, las expresiones faciales son cruciales porque revelan estados emocionales y cognitivos complejos que son visualmente expresados [1]. Además, es el sistema visual más importante y complejo de la comunicación no verbal, el rostro humano es el sistema visual que más se observa y analiza durante cualquier tipo de interacción social entre personas.

En este trabajo se implementó un método para identificar automáticamente estados mentales, se utilizaron algoritmos de Machine Learning (ML), se realizó una revisión sistemática de la literatura (RSL), en la que se revisaron trabajos relacionados con este tema. En [14] se analizan las expresiones faciales y datos relacionados con gestos de las manos sobre la cara, el movimiento y postura de la cabeza y de comportamiento ocular para el reconocimiento de estados mentales.

Se revisaron los trabajos [10], [11] y [19], en los que se identificaron aspectos relevantes para localizar y pre-procesar características del rostro del usuario, la selección de

características relevantes y el uso de algoritmos de aprendizaje automático supervisado. En la segunda etapa de la RSL se revisaron y clasificaron trabajos que realizaban el reconocimiento de emociones y estados cognitivos mediante algoritmos de ML y DL. Estos trabajos se tomaron como base para la construcción del método para reconocer automáticamente estados mentales. Por otro lado, los trabajos [2] y [4] presentan arquitecturas basadas en redes neuronales, reportando buenos resultados, sin embargo, el costo computacional es alto, por lo que para esta investigación se decidió utilizar algoritmos clásicos de aprendizaje máquina supervisado, como el utilizado en [5]; además de considerar los trabajos [6] y [7] que aplican técnicas geométricas para el reconocimiento de emociones reportando buenos resultados.

B. Motivación

El proceso de evaluación de la experiencia del usuario (UX por sus siglas en inglés) permite evaluar los atributos de calidad de un producto digital. El objetivo es identificar y corregir problemas de usabilidad e interacción que afecten el nivel de aceptación y en consecuencia aumentar el nivel de insatisfacción o rechazo del producto digital. Para identificar problemas relacionados con la usabilidad, se utilizan instrumentos convencionales para recolectar y analizar información (cuestionarios o entrevistas). Sin embargo, estos instrumentos presentan limitantes al momento de precisar la retroalimentación del usuario, ya que son auto informados, es decir, dependen de la voluntad y de la habilidad del usuario para expresar cómo se sintió antes, durante y después del proceso de interacción con la interfaz digital de un producto software; entre sus limitantes encontramos que los datos registrados pueden ser manipulados e interpretados de manera diferente por los evaluadores de usabilidad (ver Fig. 1). Esta limitante se puede mejorar utilizando métodos para reconocer automáticamente estados mentales, utilizando técnicas de visión artificial para analizar las expresiones faciales del usuario durante la interacción con la interfaz digital de un software, debido a que revelan estados emocionales y cognitivos que los usuarios expresan de manera inconsciente y que proporcionan información valiosa para la toma de decisiones en el proceso de evaluación de usabilidad en la evaluación de la experiencia del usuario UX.

Con respecto a la complejidad computacional, el reconocimiento automático de emociones y de estados cognitivos, mediante el análisis de expresiones faciales, presenta varios retos descrito en [1], por ejemplo, la oclusión del rostro, el contraste de la imagen, la postura u orientación del rostro, entre otros, dificultan la extracción de las características relevantes para el proceso de reconocimiento. Otro reto está relacionado con los data-set utilizados para el entrenamiento de los algoritmos de clasificación supervisado, los cuales contienen imágenes clasificadas por expertos con emociones positivas y negativas, dichas imágenes se obtienen en ambientes controlados y las fuentes visuales para inducir estados mentales, se basan en imágenes etiquetadas como emociones positivas o negativas, que difieren de las emociones inducidas por la interfaz digital de un software.

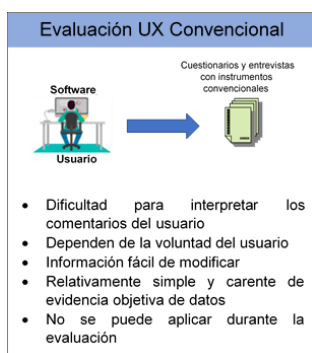


Fig. 1 Técnicas utilizadas en el proceso de evaluación UX convencional.

Los trabajos [1] al [19] se clasifican en dos categorías; la primera se relaciona con la modalidad de texturas, que se utiliza para el enfoque basado en la apariencia, para examinar las diferencias de expresión en el espacio de píxeles. Considerando la imagen facial completa como la característica expresiva, algunos trabajos mapean el vector de alta dimensión en el subespacio de baja dimensión aplicando un algoritmo de reducción de dimensionalidad, como el análisis de componentes principales (PCA) [9] y [11]. Algunos trabajos seleccionan áreas faciales específicas para el análisis y sugieren técnicas para encontrar descriptores de estas áreas. Los métodos incluyen Haar Cascades [19], filtro de extracción de bordes (EEF) [4] e Histograma de gradientes orientados (HOG) [9]. El método basado en la apariencia puede capturar la expresión facial detallada y sutil, pero las características son sensibles al nivel de iluminación, la variación de la postura de la cabeza y la oclusión. La segunda categoría se puede considerar como un método de base geométrica, este método requiere marcar un punto de característica facial cuyos movimientos pueden ayudar a capturar las características expresivas como los empleados en [5], [6] y [7].

C. Objetivos

Analizar y evaluar la eficacia de técnicas visión por computadora para identificar automáticamente estados mentales experimentados por un usuario durante el proceso de interacción con la interfaz digital de un producto software.

Como resultado se identificaron automáticamente tres emociones, aplicando técnicas de análisis de expresión facial, algoritmos de aprendizaje automático supervisado y la extracción de características con un método de base geométrica, para aplicarlo en el proceso de evaluación de usabilidad y de la UX para obtener información sobre el impacto emocional del usuario con respecto al uso de la interfaz digital de un producto software.

II. MÉTODO

Los estados mentales son los pensamientos e ideas que acompañan al estado de ánimo, se divide en dos categorías: los estados cognitivos y estados afectivos [27]. Un estado cognitivo se refiere a la condición neuropsicológica que presenta un sujeto mientras realiza una o más tareas cognitivas, que incluyen los procesos necesarios para el cálculo, la concentración y la memoria, mientras que un estado afectivo se refiere a la condición neuropsicológica que presenta un sujeto. Los estados afectivos se clasifican en emociones y sentimientos Soriano Terrazas [27], una categorización de estados mentales según la neurociencia se muestra en (Fig. 2).

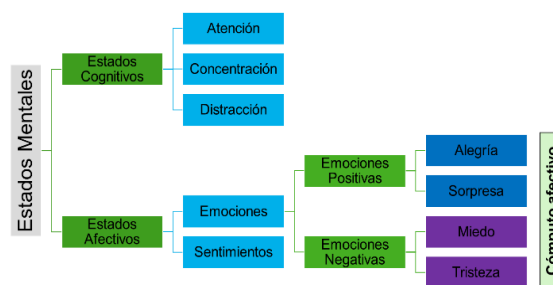


Fig. 2. Clasificación de estados mentales [27]

A. Unidades de Acción

Cada movimiento facial observable recibe el nombre de Unidad de Acción (AU) por lo que la composición de la expresión facial de una emoción se puede dividir en una o varias AU para sus respectivas variantes. De acuerdo con el EMFACS (Sistema de codificación de acciones faciales emocionales) que es un acrónimo de “Emotion FACS” este se basa en la utilización del Sistema de Codificación de la Acción Facial para detectar las emociones a través de la expresión de la cara. EMFACS (Emotion FACS) es una aplicación selectiva de la calificación FACS, en la que los codificadores solo califican el comportamiento que probablemente tenga un significado emocional. En este sistema sólo se consideran aquellas unidades de acción que se asocian a determinadas emociones, las conocidas como emociones básicas. Dichas emociones son la alegría, la sorpresa, el miedo, la ira, el disgusto y el desprecio [24].

Este sistema fue desarrollado por Paul Ekman y Wallace V. Friesen en la década de los 80 P.E [28] con la misión de simplificar la puntuación de las unidades de acción de la cara, centrándose únicamente en los signos de la emoción como se ve en la Tabla I.

TABLA I. UNIDADES DE ACCIÓN QUE CONFORMAN CADA EMOCIÓN EN EL EMFACS [28]

| Nombre | Unidad de acción |
|-----------|--|
| Sorpresa | AU 1+2+5+26 |
| Miedo | AU 1+2+4+5+20+25, 26 o 27 |
| Alegría | AU 6+12 |
| Tristeza | AU 1+4+15 con o sin 54+64 |
| Asco | AU 9+17 |
| Ira | AU 4+5+7+17+23 |
| Desprecio | AU R12/L12 o R14/L14. Las letras R o L indican el lado de la cara donde se produce la acción muscular, siendo esta Left (izquierda) o Right (derecha). |

B. Puntos de referencia facial.

Se realiza mediante la detección de puntos de referencia clave. Este proceso requiere evaluar la alineación facial, estimar la postura de la cabeza, detectar parpadeos y detectar expresiones faciales. Es fundamental identificar las estructuras faciales clave utilizando técnicas de predicción, en el contexto de los puntos de referencia faciales. Este proceso se realiza en dos etapas: 1) Identificar el rostro en la imagen y 2) analizar e identificar las estructuras faciales como la nariz, la mandíbula, los ojos, la ceja y la boca.

Para el desarrollo del algoritmo se planteó el esquema de solución que se muestra en la (Fig. 3), el cual, consta de 10 fases, las primeras cinco son el método para la extracción de características con base al cálculo de distancias geométricas (1,2,3,4 y 5) y las otras cinco fases (6, 7, 8 ,9 y 10) corresponden al entrenamiento, validación y pruebas del clasificador.

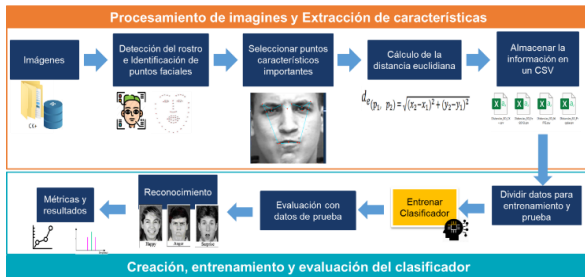


Fig. 3 Arquitectura propuesta

C. Selección de repositorio de imágenes emocionales

En la revisión sistemática de la literatura se identificó el repositorio de acceso abierto más utilizado, Cohn-Kanade extended (CK+) [20], adicionalmente se creó un data-set propio con la finalidad de robustecer el entrenamiento del algoritmo utilizado. La Tabla II muestra las características.

TABLA II. REPOSITARIOS PARA ENTRENAMIENTO Y PRUEBA DE LOS ALGORITMOS.

| Dataset | Características | Emociones | Cantidad por clase |
|----------|-----------------------------------|-------------------------------------|--------------------|
| Ck+ [20] | 432, Frontal escala Gris, 640X480 | Felicidad, Sorpresa, Enojo, Neutral | 108 |
| Propio | 432, Frontal RGB, 968X720 | Feliz, Enojo, Sorpresa, Neutral | 108 |

Este repositorio de imágenes propio se creó para entrenar y evaluar los algoritmos de clasificación de emociones, utilizando rostros de personas mexicanas, ya que los repositorios antes mencionados contienen imágenes con rostros de otras regiones geográficas. Por esta razón surgió la

necesidad de crear un repositorio relacionado con las características faciales mexicanas, el repositorio contiene 532 imágenes de rostros en formato RGB con un tamaño de 968 X 720 pixeles de los cuales se seleccionaron 432 imágenes para realizar el entrenamiento y las pruebas, los detalles del repositorio se describen en la Tabla III.

TABLA III. REPOSITARIO DE IMÁGENES PROPIO.

| Etiqueta | Emoción | Personas | Hombres | Mujeres |
|----------|----------|----------|---------|---------|
| 0 | Alegría | 149 | 59% | 49% |
| 1 | Enojo | 115 | 54% | 46% |
| 2 | Sorpresa | 119 | 47% | 53% |
| 3 | Neutral | 149 | 64% | 36% |

Para que el repositorio estuviera balanceado, las imágenes se clasificaron en cuatro clases, cada conjunto de imágenes se almacenó en cuatro repositorios que corresponden a cada emoción, como se muestran en (Fig. 4).



Fig. 4 Rostros con las emociones de Felicidad, Enojo, Sorpresa y Neutral respectivamente del conjunto de datos PROPIO

Las imágenes de repositorio CK+ muestran una variación de iluminación mínima porque se tomaron en ambientes controlados, es decir, en un área bien iluminada. Los ejemplos se muestran en (Fig. 5).



Fig. 5 Rostros con las emociones de Felicidad, Enojo, Sorpresa y Neutral respectivamente del conjunto de datos CK+ [20].

Cada conjunto de imágenes se dividió en carpetas con números que van del 0 al 3 que se corresponden con las cuatro emociones que se intenta predecir. El número 0 indica alegría, el número 1 enojo, el número 2 sorpresa y el número 3 neutralidad.

D. Detección de Rostro e identificación de puntos faciales.

El algoritmo 1 muestra el pseudocódigo para estimar los puntos de referencia en la cara desarrollado por Sullivan, 2014 e implementado en [22] en su librería Dlib.

Algoritmo 1 Entrenamiento de árboles de regresión para localización de puntos faciales

Entrada: Tener datos de entrenamiento $\{(I_{\pi_i}, \hat{s}_i^{(t)}, \Delta S_i^{(t)})\}_{i=1}^N$ y el aprendizaje tasa (factor de contracción) $0 < v < 1$

Salida: $r_i(I, \hat{S}^{(t)}) = f_k(I, \hat{S}^{(t)})$

1. Inicializar

$$f_0(I, \hat{S}^{(t)}) = \arg \min_{\gamma \in \mathbb{R}^{2p}} \sum_{i=1}^N \|\Delta S_i^{(t)} - \gamma\|^2$$

2. para $k = 1, \dots, K$: hacer

(a). asignar

para $i = 1, \dots, N$ hacer

- $r_{ik} = \Delta \hat{S}^{(t)} - f_{k-1}(I_{\pi i}, \hat{S}_i^{(t)})$
- (b). Ajuste un árbol de regresión a los objetivos r_{ik} dando una función de regresión débil $g_k(I, \hat{S}^{(t)})$.
 - (c). Actualizar $f_k(I, \hat{S}^{(t)}) = f_{k-1}(I, \hat{S}^{(t)}) + v g_k(I, \hat{S}^{(t)})$

Fin para **Fin para**

3. **Salida** $r_t(I, \hat{S}^{(t)}) = f_k(I, \hat{S}^{(t)})$

El sistema de detección de expresiones faciales comienza localizando el rostro en una imagen para identificar los rasgos o características faciales necesarios para reconocer emociones. Para la detección del rostro, las imágenes se redimensionan a 280x280 píxeles y se transforman a niveles de gris. La información de color se elimina porque la descripción del rostro se realiza mediante distancias geométricas, entre algunos puntos de los componentes del rostro. Posteriormente, se localiza el rostro en la imagen mediante el detector de puntos de referencia facial, previamente entrenado, que se incluye en la librería Dlib [21]. Esta librería se describe en [22] la cual estima la ubicación de 68 puntos que asignan a estructuras faciales en la cara.

Para que el algoritmo detector [22] realice la predicción de los puntos, toma como parámetro de entrada la imagen en escala de grises obteniendo las coordenadas x , y en píxeles de cada punto, mediante un bucle que va iterando punto por punto del predictor. En (fig 6) se puede observar cómo se posicionan los puntos del predictor.



Fig. 6 Los 68 puntos de referencia faciales del conjunto de datos iBUG

En (fig. 7) se muestra los puntos localizados en un rostro con una expresión de felicidad.

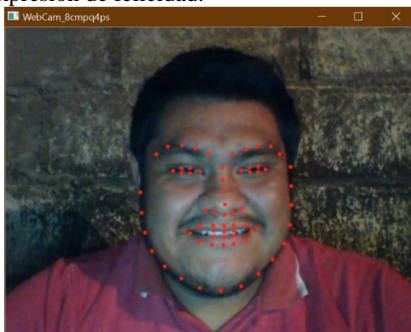


Fig. 7. Los 68 puntos de referencia faciales posicionados en el rostro que presenta la emoción de felicidad

Se realizaron pruebas con variación de iluminación y poses para determinar la limitante que la librería Dlib tiene en la detección del rostro, las fig. 8 muestran las pruebas realizadas en la detección del rostro.



Fig. 8. Localización de los puntos de referencia facial en el rostro con diferente orientación y variación de luminosidad en la imagen.

E. *Selección de puntos para identificar emociones*

De los 68 puntos propuestos en [20] solo se consideran aquellos que sirven como base para calcular las distancias necesarias que permite identificar las cuatro emociones y posteriormente guardar esas distancias en un archivo CSV, para calcular las distancias de la apertura de los parpados, se tomaron cuatro puntos de la mascarilla, quedando de la siguiente manera, los puntos [69,70] para la apertura del ojo derecho y [71,72] para la apertura del ojo izquierdo. Los 17 puntos seleccionados se describen en la tabla IV.

TABLA IV PUNTOS RELEVANTES PARA GENERAR LAS DISTANCIAS QUE NOS AYUDARAN A IDENTIFICAR LAS 4 EMOCIONES.

| No. | Región del rostro | Puntos o coordenadas |
|-----|-------------------|--------------------------|
| 1 | Ojo izquierdo | [45, 71,72] |
| 2 | Boca | [60, 64, 62, 66, 48, 54] |
| 3 | Ceja izquierda | [22, 25] |
| 4 | Nariz | [27] |
| 5 | Ceja derecha | [18,21] |
| 6 | Ojo derecho | [36, 69,70] |

La (fig. 9) muestra los 17 puntos seleccionados en el rostro, y las distancias consideradas para describir la expresión facial presente, por ejemplo, para la emoción de felicidad intervienen las unidades de acción AU6 y AU12 la primera corresponde al levantamiento de las mejillas y la segunda corresponde a la contracción de las comisuras de la boca, en esas regiones se encuentran los puntos 36, 45, 48, 60, 62, 64 y 66, de esta forma fue como se relacionó cada punto con las unidades de acción.

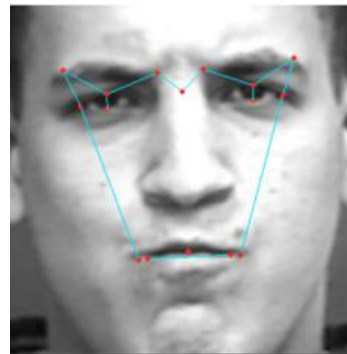


Fig. 9 Puntos faciales en un rostro con expresión de enojo del conjunto de datos CK+.

Las tablas V a la VII muestran la relación de las unidades de acción (UA) con los puntos de referencia facial seleccionados para reconocer tres emociones 1) felicidad, 2) enojo y 3) sorpresa.

TABLA V RELACIÓN ENTRE LOS PUNTOS FACIALES Y LAS UNIDADES DE ACCIÓN DE LA EMOCIÓN DE FELICIDAD

| UA | Imagen representativa | Definición | Puntos |
|------|-----------------------|---------------------------------|----------------------|
| AU6 | | Mejillas elevadas | [36, 48] [45, 54] |
| AU12 | | Comisura de los labios elevados | [60, 64] [62, 66] |

TABLA VI RELACIÓN ENTRE LOS PUNTOS FACIALES Y LAS UNIDADES DE ACCIÓN DE LA EMOCIÓN DE ENOJO

| UA | Imagen representativa | Definición | Puntos |
|------|-----------------------|--------------------------|----------------------|
| AU4 | | Cejas bajas | [22, 27] [21, 27] |
| AU5 | | Párpado superior elevado | [69, 70] [71, 72] |
| AU7 | | Párpados tensos | [60, 64] [62, 66] |
| AU23 | | Labios tensos | |

TABLA VII RELACIÓN ENTRE LOS PUNTOS FACIALES Y LAS UNIDADES DE ACCIÓN DE LA EMOCIÓN DE SORPRESA

| UA | Imagen representativa | Definición | Puntos |
|-----|-----------------------|--------------------------------|----------------------|
| AU1 | | Interior de las cejas elevadas | [18, 69] [21, 69] |
| AU2 | | Exterior de las cejas elevadas | [25, 71] [25, 45] |
| AU5 | | Párpado superior elevado | [69, 70] [71, 72] |
| AU6 | | Mandíbula caída | [60, 64] [62, 66] |

F. Cálculo de la distancia euclidiana

Una vez que se identificaron las UA, se aplica la ecuación 1 para calcular la distancia euclidiana, que se utiliza para calcular las distancias entre dos puntos. Se tienen catorce combinaciones de puntos, que intervienen en cada una de las UA relacionadas con cada emoción, de acuerdo a [23].

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

Donde:

$d_e(p_1, p_2)$ es la distancia que se pretende calcular de un punto característico del rostro a otro punto.

- x_1 es la coordenada X en pixeles del punto uno.
- x_2 es la coordenada X en pixeles del punto dos.
- y_1 es la coordenada Y en pixeles del punto uno.
- y_2 es la coordenada Y en pixeles del punto dos.

Por ejemplo, en la tabla VIII se describen los puntos y combinaciones de expresiones y sus respectivos puntos que se utilizan para calcular la distancia euclidiana de cuatro rasgos faciales, definidos por la distancia entre los puntos [36,48] mejilla elevada derecha, distancia entre los puntos [45,54] mejilla elevada izquierda, distancia entre los puntos [60,64] comisura de los labios elevados izquierdo y derecho

de la boca y la distancia entre los puntos [45,54] del labio inferior inferior y del labio superior de la boca.

TABLA VIII PUNTOS Y UA PARA CALCULAR LA DISTANCIA EUCLIDIANA DE LA EMOCIÓN DE FELICIDAD

| No | Combinación | Puntos | Imagen |
|-------------------|---|---------|--------|
| AU6 y AU12 | | | |
| 1 | Long_mejillas_elevadas_derecha | [36,48] | |
| 2 | Long_mejillas_elevadas_izquierda | [45,54] | |
| 3 | Long_comisura_interior_izquierda_y_derecha_de_la_boca | [60,64] | |
| 4 | Long_labio_interior_inferior_y_superior_de_la_boca | [62,66] | |

G. Almacenamiento de información.

Una vez calculadas las distancias los resultados se almacena en un archivo CSV, conformando el data-set para entrenar los algoritmos de aprendizaje automático supervisado. El conjunto de datos contiene catorce combinaciones de características y una columna que especifica la emoción con la que se etiqueta cada imagen (Fig. 10).

Fig. 10 Estructura del archivo CSV

H. Etapa de entrenamiento

De acuerdo con la RSL se seleccionó el algoritmo con los mejores resultados en el reconocimiento de emociones aplicando base geométrica [5], [6] y [7]. El algoritmo seleccionado fue Support Vector Machine SVM [24]. Aunque se creó originalmente como una técnica de clasificación binaria, también se puede utilizar para problemas de clasificación multiclase y de regresión.

Para iniciar con el entrenamiento del algoritmo, se dividió el data-set propio y se mezcló con el data-set FER2013[25], se tomaron aleatoriamente 108 imágenes de cada clase, para evitar un desbalanceo de clases, posteriormente se establecieron los hiperparametros óptimos con los cuales se obtuvo un 84.52% de precisión. De manera complementaría

se evaluaron otros dos algoritmos, RandomForest y KNN, se entrenaron con el mismo data-set fusionado y se seleccionaron los hiperparámetros para cada caso.

III. RESULTADOS, DISCUSIÓN Y CONCLUSIONES

Los resultados obtenidos para cada uno de los algoritmos de clasificación, se presentan en la Tabla XII y la comparación con otros trabajos similares en la Tabla XIII.

A. Métricas

La precisión, el recuerdo y el puntaje F1 fueron las métricas evaluadas en este experimento. Las Tablas IX, X y XI muestran los resultados de las métricas mencionadas y los resultados de cada emoción de cada algoritmo evaluado.

TABLA IX MÉTRICAS DEL ALGORITMO MSV.

| Emoción | Precisión | Recall | F1-Score |
|-----------|-----------|--------|-------------|
| Felicidad | 0.96 | 0.92 | 0.94 |
| Neutral | 0.76 | 0.79 | 0.78 |
| Sorpresa | 0.91 | 1.00 | 0.95 |
| Enojo | 0.71 | 0.62 | 0.67 |

TABLA X MÉTRICAS DEL ALGORITMO RANDOM FOREST.

| Emoción | Precisión | Recall | F1-Score |
|-----------|-----------|--------|-------------|
| Felicidad | 1.00 | 0.88 | 0.93 |
| Neutral | 1.00 | 0.74 | 0.72 |
| Sorpresa | 0.97 | 0.95 | 0.97 |
| Enojo | 0.57 | 0.75 | 0.65 |

TABLA XI MÉTRICAS DEL ALGORITMO KNN.

| Emoción | Precisión | Recall | F1-Score |
|-----------|-----------|--------|----------|
| Felicidad | 1.00 | 0.96 | 0.98 |
| Neutral | 0.67 | 0.92 | 0.77 |
| Sorpresa | 1.00 | 0.90 | 0.95 |
| Enojo | 0.80 | 0.50 | 0.62 |

TABLA XII EXACTITUD DE CADA ALGORITMO.

| Algoritmo | Accuracy |
|---------------|----------|
| MSV | 84.52 % |
| KNN | 84.50 % |
| Random Forest | 82.14 % |

TABLA XIII COMPARACIÓN DE CLASIFICACIÓN DE EMOCIONES CON BASE GEOMÉTRICA Y EL PROPUESTO.

| Autor | Algoritmo | Accuracy | Puntos Faciales |
|-----------|-----------|----------------|-----------------|
| [6] | FCM | 93.67 % | 44 |
| | SVM | 93.10 % | 44 |
| [5] | MLP | 88.03 % | 68 |
| | SVM | 38.46 % | 68 |
| Propuesto | SVM | 84.52 % | 17 |

Como se puede ver en las tablas, se obtuvieron buenos resultados con los tres algoritmos, sin embargo, aunque se obtuvieron buenos resultados en el Accuracy se puede identificar el mejor algoritmo en la clasificación de las cuatro emociones fue el clasificador SVM realizando un buen reconocimiento en cada emoción como se ve en (Fig 11).

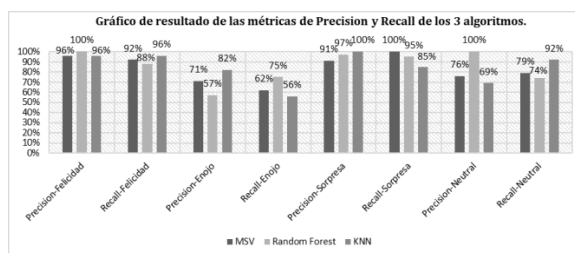


Fig. 11. Resultados generales de los tres algoritmos

En (fig. 12) se presenta el resultado general de la métrica F1-score de los tres algoritmos evaluados MVS, RF y KNN. Se puede observar que las emociones que presentan mejores resultados son felicidad y sorpresa.

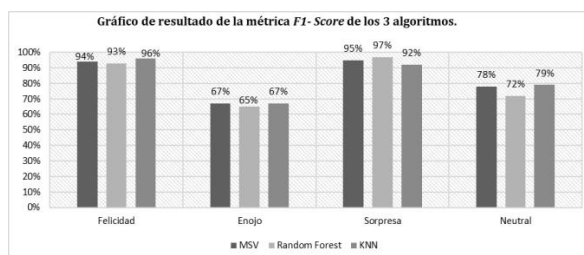


Fig. 12. Resultado de la métrica F1-score de los tres algoritmos

B. Pruebas con videos.

Se realizaron pruebas con procesamiento de video utilizando el algoritmo MVS, que dio los mejores resultados en las pruebas anteriores, obteniendo un buen porcentaje de reconocimiento para cada emoción de clasificación de un video con el porcentaje de probabilidad de reconocimiento de cada emoción.

En (Fig. 13) se presentan los resultados de las pruebas con video de la emoción de felicidad. El porcentaje de reconocimiento se muestra en cada imagen.

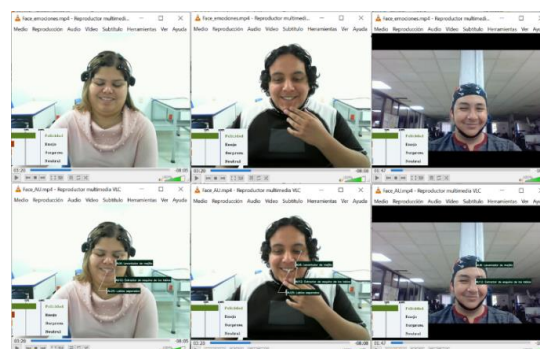


Fig. 13. Emoción de Felicidad y sus Unidades de Acción.

En (Fig. 14) se presentan los resultados del video de la emoción de enojo. El porcentaje de reconocimiento se muestra en cada imagen.

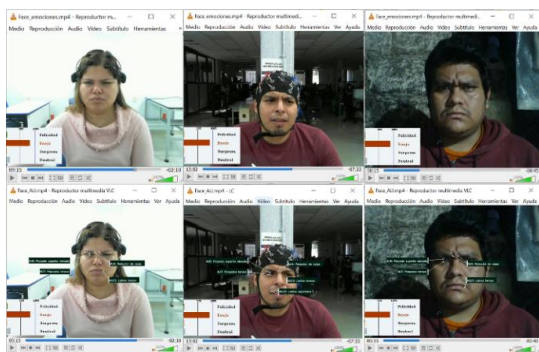


Fig. 14. Emoción de Enojo y sus Unidades de Acción.

En (Fig. 15) se presentan los resultados del video de la emoción de sorpresa. El porcentaje de reconocimiento se muestra en cada imagen.

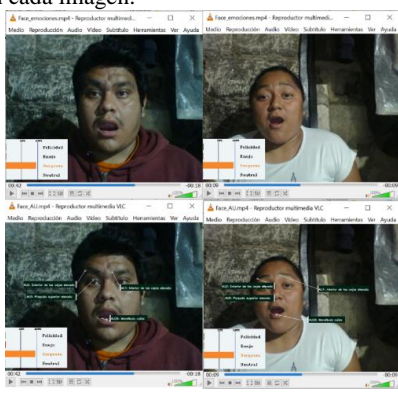


Fig. 15. Emoción de Sorpresa y sus Unidades de Acción.

En (Fig. 16) se presentan los resultados del video del estado neutral. El porcentaje de reconocimiento se muestra en cada imagen.

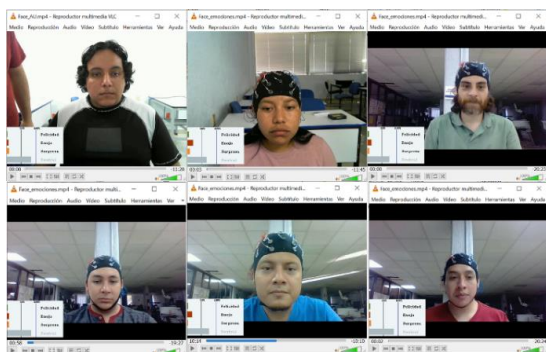


Fig. 16. Secuencia de video detectando la emoción Neutra.

Cabe mencionar que cuando la emoción se presenta de manera más pronunciada o remarcada el clasificador alcanza el porcentaje siguiente para cada emoción ver Tabla XIV.

TABLA XIV PORCENTAJE DE RECONOCIMIENTO DE EMOCIONES EN VIDEO.

| Emoción | Porcentaje de reconocimiento |
|-----------|------------------------------|
| Felicidad | 95.43 % |
| Enojo | 98.6 % |
| Sorpresa | 91.46 % |
| Neutral | 88.00 % |

C. Conclusión.

En esta investigación se presenta el resultado de la implementación de un método para el reconocimiento automático de emociones mediante el análisis de características de expresiones faciales. El reconocimiento de emociones mediante el análisis facial con base geométrica y un clasificador de aprendizaje automático supervisado SVM que presentó los mejores resultados y un costo computacional bajo. Para validar la efectividad del método desarrollado se utilizaron dos data-set de imágenes, uno de acceso público CK+ y otro propio, generado con características faciales de personas mexicanas. El resultado obtenido fue aceptable, sin embargo, todavía hay trabajo por hacer para mejorar la precisión de reconocimiento, que es del 84,52%. Estos resultados se pueden mejorar mediante el procesamiento de imágenes, aplicando filtros de iluminación y contraste, realizando una alineación de rostros, obtener imágenes con una mayor nitidez y que los rostros tengan expresiones más pronunciadas para cada emoción, también se requiere aumentar el número de imágenes para cada clase. La relación que se hizo con los puntos de referencia facial y el sistema de codificación de acción facial de emociones (EmFACS) para la extracción de características, permitió que la clasificación fuera menos costosa computacionalmente. Por otra parte, también fue posible demostrar que el análisis facial para la clasificación de emociones durante la evaluación de la experiencia del usuario (UX) proporcione información valiosa con respecto a lo que el usuario siente durante la interacción con la interfaz digital de un producto software ya que en conjunto con los datos fisiológicos (EEG, ECG, GSR y Eye-tracking) se puede realizar un mejor análisis de cada evaluación y cuando la clasificación de los datos fisiológicos y seguimiento ocular no proporcione información clara, el método basado en comportamiento podrá complementar la información, ya que por medio de las expresiones del rostro del usuario se puede notar fácilmente lo que el usuario va experimentando durante la interacción. Este método se integró en una plataforma de evaluación de la experiencia del usuario llamada QUXBox en el futuro. El objetivo es identificar los estados mentales de los usuarios que participen en el proceso de evaluación de la experiencia de compra de productos digitales.

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.

REFERENCES

- [1] Sunitha A, P. Ajay Kumar Reddy, S.Nanda Kishore, G.N Kodanda Ramaiah (2017), Recognition of Facial Emotions Based on Sparse Coding. Journal of Engineering Research and Application, ISSN: 2248-9622.
- [2] Nwosu, L, Wang, H., Lu, J, Unwala, I, Yang, Zhang, T.(2017, Septiembre). deep convolutional neural network for facial expression recognition using facial parts. DOI 10.1109/DASC-PICOM-DataCom-CyberSciTec.2017.213
- [3] Krestinskaya, et al. (2017, septiembre 13-16). Conferencia Internacional sobre Avances en Informática, Comunicaciones e Informática (ICACCI) Udupi, India, Facial Emotion Recognition using Min-Max Similarity Classifier. IEEE. 10.1109/ICACCI.2017.8125932
- [4] Liu Xiao, Lee Kiju. (2018, August 15-17). IEEE Games, Entertainment, Media Conference (GEM), Galway, Irlanda, Optimized

- Facial Emotion Recognition Technique for Assessing User Experience. *IEEE*, 10.1109 / GEM.2018.8516518
- [5] M. Alvarez V, et al. (2018, August 22-24). Conferencia Internacional sobre Investigación en Inteligencia y Computación en Ingeniería (RICE), San Salvador, El Salvador, Facial Emotion Recognition: A Comparison of Different Landmark-Based Classifiers. *IEEE*. 10.1109 / ARROZ.2018.8509048
- [6] Dewi, Y.L., Widyano, M., Basaruddin, T. (2018). Geometric facial components feature extraction for facial expression recognition.
- [7] Mangal, Divya., Prajwala.(2018, Abril 3-5). International Conference on Communication and Signal Processing, Facial expression recognition by calculating euclidian distance for Eigenfaces using pca.
- [8] Kartali, et al. (2018, November 20-21).14th Symposium on Neural Network and Aplication (NEUREL), Belgrade, serbia real-time algorithms for facial emotion recognition: a comparison of different approaches.*IEEE*.
- [9] Rabhi Yassine, et al. (2018, March 21-24). 4a Conferencia internacional sobre tecnologías avanzadas para el procesamiento de señales e imágenes (ATSIP), Susa, Túnez, A Real-time Emotion Recognition System for disabled persons.*IEEE*. 10.1109 / ATSIP.2018.8364339
- [10] Cadena Moreano, J.A, La Serna Palomino, N., Llano Casa, A. C. (2019, September). A Facial recognition technique using SVM: A comparative analysis. e-ISSN: 1390-6542 / p-ISSN: 1390-9363.
- [11] Sawhney Shreyak, et al. (2019, January 10-11). IX Conferencia Internacional sobre Cloud Computing, Data Science & Engineering (Confluence), Noida, India, India Real-Time Smart Attendance System using Face Recognition Techniques.*IEEE*. 10.1109/ CONFLUENCE.2019.8776934
- [12] Lacort, J. (21 agosto 2017). Las claves de los sistemas de reconocimiento facial: ¿cuál es su verdadero nivel de seguridad? Recuperado el 2 de junio del 2020 de <https://www.xataka.com/seguridad/las-claves-de-los-sistemas-de-reconocimiento-facial-cual-es-su-verdadero-nivel-de-seguri>
- [13] Hussain, J., Ali Khan, W., Hur, T., Bilal, H., Bang, J., Hassan, A. . . . Lee, S. (2018, Mayo 18). A Multimodal Deep Log-Based User Experience (UX) Platform for UX Evaluation. *Sensors*, 18(5). doi:10.3390/s18051622.
- [14] Behera Ardhendu, et al (2020, abril), International Journal of Artificial Intelligence in Education. Associating Facial Expressions and Upper-Body Gestures with Learning Tasks for Enhancing Intelligent Tutoring Systems, <https://doi.org/10.1007/s40593-020-00195-2>
- [15] Jain Udit, et al. (2018, mayo), 2do Congreso Internacional de Tendencias en Electrónica e Informática (ICOEI). Tirunelveli, India. Analysis of Face Detection and Recognition Algorithms using Viola Jones Algorithm with PCA and LDA. 10.1109 / ICOEI.2018.8553811
- [16] Sharma S., et al. (2016, mayo 25-27). Conferencia Internacional sobre Tecnologías Avanzadas de Computación y Control de Comunicaciones (ICACCCT). Ramanathapuram, India, FAREC - CNN Based Efficient Face Recognition Technique using Dlib. *IEEE*. 10.1109 / ICACCCT.2016.7831628
- [17] James Garrett, Jesse (2011) The Elements of User Experience: User-Centered Design for the Web and Beyond, Second Edition, ISBN 13: 978-0-321-68368-7
- [18] Michael J. Lyons, Shigeru Akamatsu, Miyuki Kamachi, Jiro Gyoba. Coding Facial Expressions with Gabor Wavelets, 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205 (1998). <http://doi.org/10.1109/AFGR.1998.670949> Open access content available at: <https://zenodo.org/record/3430156>
- [19] Yang, D, Alsadoon, A., P.W.C, Prasad, Singh, A, K, Elchouemi, A. (2017, Dic), International Conference on Smart Computing and Communications. Kurukshetra, India. An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment Networks. <https://doi.org/10.1016/j.procs.2017.12.003>
- [20] Valderrama Cárdenas, W. (2019). Reconocimiento automático del rostro para verificación de identidad para evaluación en línea (tesis de maestría). Morelos, México: CENIDET.
- [21] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 94-101.
- [22] D. King, "dlib C++ Library," 2015. [Online]. Available: www.dlib.net.
- [23] Kazemi, V, Sullivan, J. (23-28 June 2014). IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. One millisecond face alignment with an ensemble of regression trees. *IEEE*. 10.1109/CVPR.2014.241
- [24] P.E Group "FACS Archives - Paul Ekman Group, LLC.," 2016 [Online]. Available: <http://www.paulekman.com/product-category/facs/>
- [25] Joaquín A.R. (2021) Máquinas de Vector Soporte (SVM) con Python available under a Attribution 4.0 International (CC BY 4.0) at <https://www.cienciadedatos.net/documentos/py24-svm-python.html>
- [26] Sambare, M. (2020, Julio 19). FER-2013 Learn facial expressions from an image. Version 1. <https://www.kaggle.com/msambare/fer2013/metadata>
- [27] Jeovanny Soriano Terrazas (2018). Metodología para caracterizar e inducir estados mentales a través de realidad virtual inmersiva e interfaz cerebro computadora (Tesis de maestría). CENIDET, Morelos, México
- [28] P.E Group " Facial Action Coding System - Paul Ekman Group, LLC.," 2016 [Online]. Available: <https://www.paulekman.com/facial-action-coding-system/>