Universidad Politécnica de Puebla

Programa Académico de Posgrado

Sistema de apoyo para la evaluación psicológica a través del reconocimiento de expresiones y micro-expresiones faciales

TESIS QUE PARA OBTENER EL GRADO DE

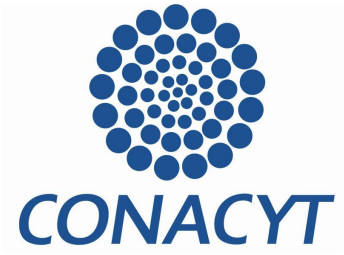
Maestría en Ingenería en Sistemas y Cómputo Inteligente

Presenta:

SONIA MARIBEL GONZÁLEZ LOZOYA

Director: Dr. Jorge de la Calleja Mora
Co-Director: Dr. Hugo Jair Escalante Balderas (INAOE)

Juan C. Bonilla, Puebla, México, Agosto 2017.



El presente trabajo fue realizado en el Laboratorio de Investigación y Posgrado del Departamento de computo de la Universidad Politécnica de Puebla, ubicada en Tercer carril del Ejido "Serrano" S/N, San Mateo Cuanalá, Municipio Juan C. Bonilla, Puebla, CP 72640. Apoyo del CONACYT, Beca No. 71150, Programa de Maestría perteneciente al Programa Nacional de Posgrados de Calidad (PNPC-CONACYT)

Universidad Politécnica de Puebla
Maestría en Ingeniería en Sistemas y Cómputo Inteligente

Sistema de apoyo para la evaluación psicológica a través del reconocimiento de expresiones y micro-expresiones faciales

Tesis realizada por:

SONIA MARIBEL GONZÁLEZ LOZOYA

Aprobada por ... Agosto, 2017.

Profesor **(Firma)**

Dr. Jorge de la Calleja Mora .....

Dr. Hugo Jair Escalante Balderas (INAOE) .....

Dr. Juan Manuel González Calleros (BUAP) .....

Dr. Antonio Benítez Ruiz .....

Juan C. Bonilla, Puebla, México, Agosto 2017.

Universidad Politécnica de Puebla
Maestría en Ingenería en Sistemas y Cómputo Inteligente

Copyright © – All rights reserved. Universidad Politécnica de Puebla, 2030.

Copyright statement

(Firma)

.....
Sonia Maribel González
Lozoya

Resumen

El objetivo de este proyecto es desarrollar un sistema computacional para el apoyo en la evaluación psicológica, a tráves del reconocimiento de expresiones y micro-expresiones faciales usando una representación basada en aprendizaje profundo con redes neuronales convolucionales.

En el desarrollo experimental, se adquirieron bases de datos con ejemplos de expresiones y micro-expresiones faciales para formar conjuntos de entrenamiento, se emplearon redes neuronales convolucionales para la extracción de características y se utilizó la máquina de soporte vectorial para el proceso de aprendizaje, obteniendo seis modelos binarios que reconocen seis expresiones faciales básicas, con un promedio de exactitud del 92 % y tres modelos binarios que reconocen tres clases de micro-expresiones faciales, con un promedio de exactitud del 77 %. Estos resultados son comparables con los obtenidos en el estado del arte. Mientras en la construcción del prototipo se empleó Matlab para el desarrollo de la interfaz gráfica, que permite la selección de un video, el cual será analizado con el propósito de encontrar las expresiones y micro-expresiones faciales.

Obteniendo finalmente, un prototipo eficaz y fácil de utilizar, para reconocer seis clases de expresiones y 3 clases de micro-expresiones faciales de un video, con la finalidad de que especialistas apoyados con este tipo de información puedan determinar un diagnóstico acertivo y proporcinar un tratamiento adecuado para el paciente.

Palabras claves

Evaluación psicológica, reconocimiento de expresiones faciales, reconocimiento de micro-expresiones faciales, aprendizaje profundo.

Agradecimientos

A mi familia:

Por todo el amor, apoyo, comprensión y confianza que han depositado en mí, que ha servido en mi desarrollo personal y profesional.

Dr. Jorge de la Calleja y Dr. Hugo J. Escalante:

Por darme la oportunidad de realizar el proyecto de tesis bajo su supervisión, por la confianza, tiempo y apoyo a través de sus consejos.

Dr. Antonio Benitez, Dr. Luis I. Olivos, Dra. Leticia Ramírez y Dra. María A. Medina:

Por compartir sus experiencias y conocimiento en clase.

Sonia Maribel González Lozoya

Contenido

Abstract 7

Agradecimientos 9

I Planteamiento del problema de investigación 17

1. Planteamiento del problema de investigación 19

1.1. Introducción . . . 19
1.2. Objetivos . . . 20
1.2.1. Objetivo general . . . 20
1.2.2. Objetivos específicos . . . 20
1.3. Justificación . . . 20
1.4. Organización de Tesis . . . 21

II Marco teórico y estado del arte 23

2. Marco Teórico y estado del arte 25

2.1. Antecedentes . . . 25
2.2. Visión por computadora . . . 25
2.2.1. Procesamiento digital de imágenes . . . 26
2.2.2. Extracción de características . . . 27
2.2.3. Reconocimiento de objetos . . . 27
2.2.3.1. Reconocimiento facial . . . 28
2.2.3.2. Reconocimiento de expresiones y micro-expresiones faciales 29
2.3. Aprendizaje automático . . . 30
2.3.1. Aprendizaje supervisado . . . 31
2.3.1.1. Clasificación . . . 31
2.3.1.2. Evaluación de modelos . . . 32
2.3.2. Aprendizaje profundo . . . 33
2.3.2.1. Redes convolucionales . . . 34
2.4. Evaluación psicológica . . . 35
2.4.1. Valoración del paciente . . . 35
2.4.2. Comunicación no verbal . . . 35
2.4.2.1. Expresiones y micro-expresiones faciales . . . 35
2.4.2.2. Antecedentes de expresiones faciales . . . 36
2.4.2.3. Antecedentes de micro-expresiones faciales . . . 36
2.5. Trabajos relacionados . . . 37
2.5.1. Trabajos de reconocimiento de expresiones faciales . . . 37
2.5.2. Trabajos de Reconocimiento de Micro-expresiones faciales . . . 39

III Metodología **43**

3. Metodología **45**

3.1. Etapa experimental 45

3.1.1. Método para el reconocimiento de expresiones faciales 48

3.1.2. Método para el reconocimiento de micro-expresiones faciales en secuencia de imágenes 50

3.2. Etapa de contrucción del prototipo 52

IV Resultados **55**

4. Resultados **57**

4.1. Resultados de la etapa experimental 57

4.1.1. Resultados del método de reconocimiento de expresiones faciales . . 57

4.1.1.1. Experimento 1: Clasificación Binaria vs Multi-clase 57

4.1.1.2. Experimento 2: Estrategias para reducir sobreajuste 61

4.1.2. Resultados del método de reconocimiento de micro-expresiones faciales 64

4.1.2.1. Experimento 3: Reconocimiento de Micro-expresiones Faciales 64

4.2. Resultados de la etapa de contrucción del prototipo 67

4.2.1. Obtención de los modelos aprendidos 67

4.2.2. Estructura del prototipo e integración de los modelos 68

4.2.3. Diseño de la interfaz gráfica 68

4.2.4. Funcionamiento del prototipo 71

V Conclusiones **75**

5. Conclusiones **77**

5.1. Conclusiones 77

5.2. Trabajo futuro 77

Bibliografía **83**

Lista de Figuras

2.1. Ejemplos de aplicaciones de aprendizaje profundo para tareas en visión por computadora . . . 25
2.2. Esquema general de visión por computadora . . . 26
2.3. Proceso de operación puntual [1] . . . 26
2.4. Proceso de detección para identificar características rectangulares discriminatorias y determinar los clasificadores débiles, tomada de [2] . . . 28
2.5. Ejemplos de imágenes de expresiones faciales etiquetadas con su respectiva emoción tomada de http://k46.kn3.net/taringa/1/1/3/E/C/2/ robernetit 123/EB0.png . . . 29
2.6. Ejemplo de una secuencia de imágenes que contiene una micro-expresión, tomada de [3] . . . 30
2.7. Adquisición del conocimiento, tomada de [4] . . . 30
2.8. Ejemplo de aprendizaje supervisado . . . 32
2.9. Ejemplo de un diagrama de una red neuronal profunda, tomada de https:// www.extremetech.com/computing/227634-qualcomm-announces-new- deep-learning-sdk-with-support-for-snapdragon-820-heterogeneous-compute . . 34
2.10. Ejemplo gráfico de una red neuronal convolucional, tomada de https://www. linkedin.com/pulse/why-convolutioanl-neural-networks-shamane-siriwar dhana . . . 34

3.1. Metodología empleada. . . . 45
3.2. Pasos de la etapa experimental . . . 45
3.3. Ejemplos de tareas en la etapa del pre-procesamiento de datos . . . 46
3.4. Arquitectura general de CNN para la extracción de características de una imagen . . . 47
3.5. Marco de trabajo para la clasificación . . . 47
3.6. Marco de trabajo para la evaluación . . . 48
3.7. Marco de trabajo usado para el pre-procesamiento de imágenes en el reconocimiento de expresiones faciales . . . 48
3.8. Arquitectura empleada para la extracción de características de una imagen 49
3.9. Marco de trabajo usado para el pre-procesamiento de imágenes en el reconocimiento de expresiones faciales . . . 50
3.10. Arquitectura empleada para la extracción de características en una secuencia de imágenes . . . 51
3.11. Datos a considerar para la estructura del prototipo . . . 52
3.12. Interfaces para el desarrollo del prototipo . . . 52

4.1. Seis expresiones básicas de la BD CK+ [5] . . . 58
4.2. Ejemplos de la clasificación de los modelos binarios obtenidos del primero y segundo experimento . . . 65
4.3. Estructura del prototipo . . . 69

4.4. Interfaces obtenidos para el prototipo 70
4.5. Imágenes de ejemplo del funcionamiento del prototipo 72
4.6. Resultados del segmento 73
4.7. Expresiones encontradas en el video con la persona 1 73
4.8. Expresiones encontradas en el video con la persona 2 74
4.9. Expresiones encontradas en el video con la persona 3 74

Lista de Tablas

2.1. Matriz de confusión para dos clases [6] 33
2.2. Medidas de funcionamiento para problemas de dos clases. 33
2.3. Comparación de modelos para el reconocimiento de expresiones faciales . 39
2.4. Comparación de modelos para el reconocimiento de micro-expresiones faciales 40
2.5. Comparación de base de datos de micro-expresiones faciales 40

4.1. Distribución de ejemplos CK+ y JAFFE 58
4.2. Resultados de exactitud de modelos binarios 59
4.3. Resultados de precisión de modelos binarios 59
4.4. Resultados de recuerdo de modelos binarios 59
4.5. Resultados de la medida F de modelos binarios 60
4.6. Resultados multi-clase 60
4.7. Matriz de confusión de la clasificación multi-clase 60
4.8. Distribución de ejemplos de cada base de datos 62
4.9. Número de ejemplos utilizados para las configuraciones propuestas 62
4.10. Resultados de exactitud en el reconocimiento de expresiones faciales . . . 63
4.11. Resultados de precisión en el reconocimiento de expresiones faciales . . . 63
4.12. Resultados de recuerdo en el reconocimiento de expresiones faciales 63
4.13. Resultados de medida F en el reconocimiento de expresiones faciales . . . 63
4.14. Distribución de ejemplos de cada base de datos 66
4.15. Resultados de exactitud en el reconocimiento de micro-expresiones faciales 66
4.16. Resultados de precisión en el reconocimiento de micro-expresiones faciales 66
4.17. Resultados de recuerdo en el reconocimiento de micro-expresiones faciales 67
4.18. Resultados de medida F en el reconocimiento de micro-expresiones faciales 67

I

Planteamiento del problema de investigación

Capítulo 1: Planteamiento del problema de investigación

1.1. Introducción

Los problemas mentales pueden afectar el equilibrio de vida de un individuo y de las personas que lo rodean. Aproximadamente el 25 % de las personas en Europa y EUA han experimentado un trastorno alguna vez en su vida [7], mientras que en México existe una estimación del 15 % de la población adulta y solo el 3 % es atendido por un médico especialista mediante sesiones psicológicas [8].

El estudio psicológico se basa en la observación y valoración del individuo. Generalmente se lleva a cabo una evaluación basada en la comunicación experto-paciente con apoyo de cuestionarios y actividades encaminadas a la comprensión y solución a los problemas demandados. En esta etapa se busca obtener los datos de la personalidad del individuo, como pueden ser sus ideales, intereses, aptitudes, afectos, impulsos, entre otros aspectos. Cerrada la etapa de evaluación se realiza un informe con los datos relevantes, donde posteriormente con base al informe, el experto pueda establecer conclusiones y tomar decisiones de manera que le convenga al paciente. Por lo tanto, es necesario recalcar que, entre mayor sea la información que se tenga del paciente, habrá mayor probabilidad de que el experto realice las decisiones correctas para solucionar los problemas mentales [9].

Por otro lado, investigadores en el área de psicología han estudiado a las expresiones y micro-expresiones faciales, definidas como medios de comunicación no verbales que se encuentran asociados con las emociones. La diferencia entre las expresiones y las micro-expresiones son: que las expresiones faciales se manifiestan voluntariamente y es fácil disfrazarlas con el fin de ocultar las emociones reales, mientras que las micro-expresiones faciales se caracterizan por ser de corta duración, baja intensidad y se presentan de forma involuntaria revelando una emoción genuina [10].

La comunicación no verbal, como expresiones y micro-expresiones, pueden ser de gran utilidad dentro de la sesión experto-paciente, ya que permiten obtener otros datos que el paciente no transmite al experto de forma verbal. Sin embargo, incluso para expertos es complicado interpretar este tipo de comunicación, pues conlleva a identificarlas en movimientos fugaces de poca intensidad donde muchas veces no suelen percibirse a simple vista.

Una manera de apoyar al experto a recolectar estos datos de comunicación no verbal transmitidos por el paciente, es con el uso de la tecnología: un sistema computacional puede llevar a cabo la detección y reconocimiento de expresiones y micro-expresiones faciales de forma automática, además de reducir la subjetividad. Realizar dicha automatización involucra a dos grandes áreas de investigación: visión por computadora y aprendizaje automático.

La aplicación de técnicas de aprendizaje automático en visión por computadora han generado grandes avances, por ejemplo el caso de reconocimiento de rostros, seguimiento y detección de objetos [11], entre otros.

La arquitectura general para resolver una tarea de visión por computadora involucra

la aplicación de diversas etapas. Por ejemplo en clasificación de imágenes, se llevan a cabo su procesamiento digital y la extracción de características, además, de la etapa de aprendizaje automático.

Por otro lado, existen diferentes tipos de enfoques para realizar aprendizaje automático. Recientemente, el aprendizaje profundo es uno de los enfoques que ha mostrado mayor efectividad para diversas tareas que involucran imágenes. El aprendizaje profundo permite modelar abstracciones de alto nivel en datos usando arquitecturas compuestas de transformaciones no-lineales múltiples [12]. Este tipo de aprendizaje puede detectar los leves movimientos en la imagen y tomarlos en cuenta para su clasificación.

En este proyecto de investigación se presenta el desarrollo de un prototipo computacional para reconocer expresiones y micro-expresiones faciales automáticamente que permitan al especialista tener indicios sobres las emociones que manifiesta un paciente en una sesión. El sistema se basa en aprendizaje profundo aplicado a imágenes y segmentos de video, y sé encuentra compuesto por una interfaz, donde como dato de entrada se solicita un ingreso de video de un paciente, este video es examinado por dos modelos aprendidos obtenidos por el algoritmo de aprendizaje máquina de soporte vectorial (SVM en inglés *Support Vector Machine*) obteniendo un promedio de 92 % de exactitud para reconocer expresiones faciales y un 77 % de exactitud para reconocer micro-expresiones, en una base de datos que se armó para este estudio. Finalmente, como dato de salida se obtiene una lista, en donde se detalla la ubicación e identificación de las expresiones encontradas.

1.2. Objetivos

1.2.1. Objetivo general

Desarrollar un sistema de reconocimiento automático de expresiones y micro-expresiones faciales para la evaluación psicológica, usando aprendizaje profundo.

1.2.2. Objetivos específicos

- Desarrollar un método para reconocer las seis expresiones faciales básicas, usando imágenes una bases de datos.
- Desarrollar un método para reconocer tres micro-expresiones faciales, usando segmentos de video de una base de datos.
- Implementar un prototipo que integre los métodos desarrollados para facilitar al especialista la búsqueda de expresiones contenidas en un video.

1.3. Justificación

Un sistema capaz de reconocer expresiones y micro-expresiones faciales, es una herramienta de gran utilidad para apoyo al experto, que permite complementar información

dentro de la etapa de evaluación psicológica del paciente. Además, la información proporcionada por el sistema antes mencionado tiene las siguientes ventajas:

- Es de utilidad para conocer aspectos psicológicos del paciente, por ejemplo: intereses, afectos, impulsos, etc.
- Sirve para detectar incongruencia entre las respuestas del paciente, por medio de la correlación de comunicación verbal y no verbal.
- Reduce la subjetividad en los datos adquiridos durante la evaluación psicológica.
- Proporciona información complementaria para la etapa de evaluación dando mayor confiabilidad al momento de que el experto tenga que llegar a una conclusión y realizar alguna decisión para la solución del problema.
- La información puede ser de utilidad para otras aplicaciones, como en interrogatorios policiales, estudios de mercadotecnia y para varios tipos de entrevistas, como la selección de personal.

1.4. Organización de Tesis

La organización del documento de tesis es la siguiente: en el capítulo 1 además del planteamiento del problema presentado, se encuentra el objetivo general, objetivos específicos y justificación; posteriormente en el capítulo 2 se describen los temas relacionados a este proyecto; el capítulo 3 contiene la estructura detallada utilizada para el desarrollo experimental y contrucción del prototipo; en el capítulo 4 se muestran los resultados de experimentos y del prototipo obtenido; y por último en el capítulo 5 las conclusiones y el trabajo futuro que se obtiene a partir de los resultados.

II

Marco teórico y estado del arte

Capítulo 2: Marco Teórico y estado del arte

2.1. Antecedentes

En los últimos años, el aprendizaje profundo se ha categorizado como uno de los mejores enfoques para la resolución de tareas dentro del área de visión por computadora. La Figura 2.1 muestra algunas de las muchas aplicaciones en las que ha conseguido sobresalientes resultados como: reconocimiento de caracteres ópticos [13], reconocimiento facial [14], análisis de videos, [15], vehículos autónomos [16], entre otras.

En este proyecto de tesis se utiliza el aprendizaje profundo para la extracción de características en imágenes faciales y algoritmos de aprendizaje automático para la clasificación de expresiones y micro-expresiones, con la finalidad de apoyar en la interpretación de la comunicación no verbal para las evaluaciones piscológicas.

Este capítulo introduce de manera general a los temas relacionados con el proyecto de tesis. A continuación, en la subsección 2.2 se revisa el área de visión por computadora enfatizando técnicas utilizadas en esta tesis. La subsección 2.3 aborda de igual forma el área de aprendizaje automático. Por último la subsección 2.4 describe el proceso de la evaluación psicológica.

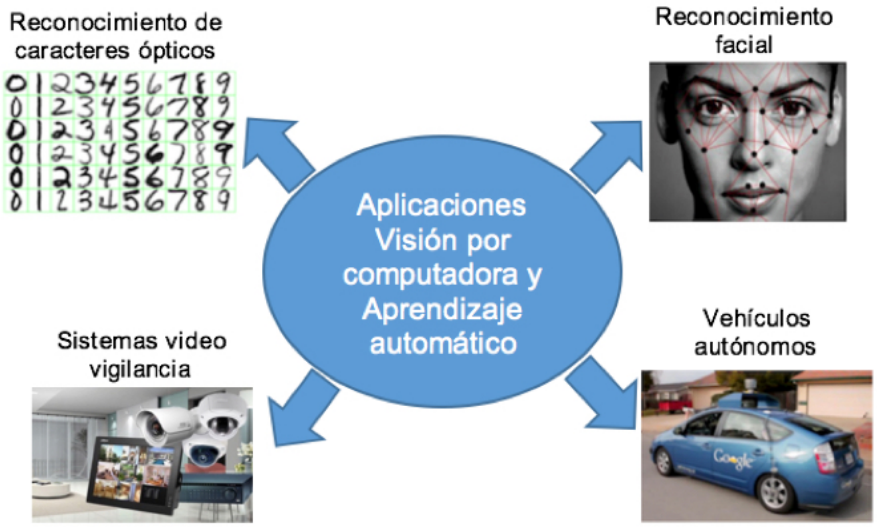


Figura 2.1: *Ejemplos de aplicaciones de aprendizaje profundo para tareas en visión por computadora*

2.2. Visión por computadora

La visión por computadora (VC) es una área de investigación que pretende imitar la capacidad de visión de un ser humano. Una forma de llevarlo a cabo es por medio del entendimiento de una imagen digital, donde se busca producir una descripción útil, la cual a su vez no contenga información irrelevante [1].

Una arquitectura general de la aplicación de visión por computadora, se aprecia en la Figura 2.2, donde se muestran tres procesos principales que son llevados a cabo:

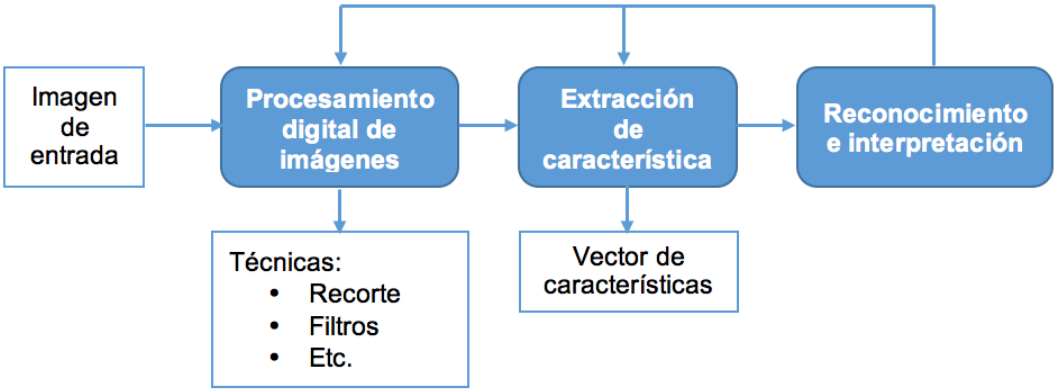


Figura 2.2: *Esquema general de visión por computadora*

A continuación, se describen técnicas principales en cada uno de los procesos antes descriptos dentro de VC.

2.2.1. Procesamiento digital de imágenes

El procesamiento digital de imágenes consiste en realizar tareas básicas donde las entradas y salidas son imágenes. Las imágenes digitales constan de un número de elementos finitos conocidos como píxeles. Cada píxel tiene un valor de intensidad o nivel de gris y una localización (x, y), donde 'x' y 'y' son coordenadas de una matriz bidimensional [17].

El objetivo general es mejorar la imagen, ya sea eliminando ruido, problemas de movimiento o desenfoque, o extraer atributos de la imagen resaltando propiedades como color, contraste, contorno, etc.; esto último dependiendo del tipo de técnica a aplicar. Las técnicas de procesamiento de imágenes más utilizadas son: operaciones puntuales, transformaciones de intensidad y filtrado [1].

- Operaciones puntuales: transforman una imagen de entrada a una imagen de salida, donde cada píxel de entrada es evaluado por una función definida para obtener el valor del pixel de salida, como se muestra en la Figura 2.3. Este tipo de técnicas solo es usada para imágenes monocromáticas.

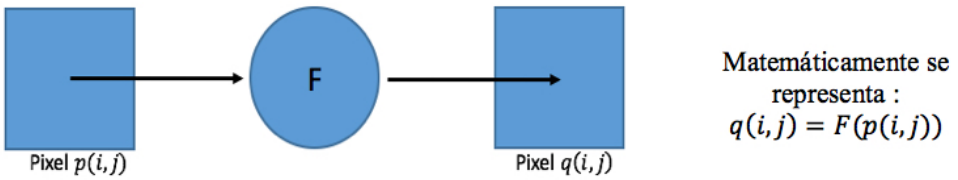


Figura 2.3: *Proceso de operación puntual [1]*

- Transformaciones de intensidad: consisten en modificar los valores de intensidad de los píxeles de acuerdo a una función de transformación, representado en la siguiente Ecuación 2.1. Estas funciones pueden ser lineales como aumento de contraste y no lineales como ecualización de histograma.

$$p(x, y) = T(p(x, y)) \tag{2.1}$$

- Filtrado: consiste en aplicar una transformación a una imagen de entrada, para obtener una imagen de salida con el fin de resaltar o acentuar ciertas características de la imagen. El filtrado se define mediante la siguiente Ecuación 2.2:

$$g(x, y) = T[f(x, y)] \tag{2.2}$$

2.2.2. Extracción de características

Mientras que la etapa de procesamiento de imágenes digitales trabaja directamente con los pixeles, donde la entrada y salida son imágenes. En la etapa de extracción de características es necesario transformar la imagen de manera conveniente para que la computadora lleve a cabo procesos que permitan reconocer objetos. La función de la extracción de características es representar la imagen de forma numérica o simbólica, generalmente mediante un vector que representa a todas aquellas características relevantes de la imagen [2].

Los autores Grauman y Leibe [2] concuerdan que el objetivo es proporcionar una representación que contenga la esencia de un objeto, con el fin de reconocer objetos entre varias imágenes. Los extractores de características deben cumplir con los siguientes criterios:

- La representación de características debe ser repetible y precisa, para reconocer algún objeto (por ejemplo, objeto igual a perro).
- A la vez, distintivo, de tal manera que la imagen reconozca el objeto pero distinga diferentes tipos de objetos (por ejemplo, objeto igual a perro y distinguir diferentes tipos de razas de perros).

Una forma de llevar a cabo la extracción de características es con el uso de descriptores visuales diseñados manualmente, como SIFT y SURF, que realizan representaciones compactas que resumen el contenido de una imagen digital [11]. Mientras que por otro lado, existen técnicas de aprendizaje profundo como la Red Neuronal Convolucional (CNN siglas en inglés *Convolucional Neuronal Network*) que por medio de varias capas de convolución permiten obtener una representación de imagen más detallada y sin la necesidad de recurrir a descriptores diseñados manualmente 2.3.2.1.

2.2.3. Reconocimiento de objetos

En esta etapa de reconocimiento de objetos se considera el nivel más alto de la VC. En esta etapa se utilizan los datos extraídos por las etapas anteriores: procesamiento de imágenes digitales y extracción de características. En esta subsección se dará un énfasis a las tareas de reconocimiento facial y reconocimiento de expresiones y micro-expresiones, así como métodos utilizados en el estado del arte.

2.2.3.1. Reconocimiento facial

El detector Viola-Jones es uno de los algoritmos más populares en el estado del arte para el reconocimiento facial [2]. Este algoritmo utiliza una ventana a manera de filtro que permite reconocer patrones de textura 2D de varias instancias de rostro. Se lleva a cabo un aprendizaje que discrimina y detecta regularidades, mediante un conjunto de clasificadores débiles. Estos clasificadores obtienen características sencillas basados en contraste. A pesar que el algoritmo de entrenamiento es costoso y consume mucha memoria, tiene la ventaja de ser rápido una vez que el modelo es aprendido, por ello es que el algoritmo es utilizado para aplicaciones de búsqueda de personas en imágenes y videos. A continuación, se describen las etapas del proceso de este detector, con el propósito de ejemplificar cómo funcionan los detectores de objetos, [2].

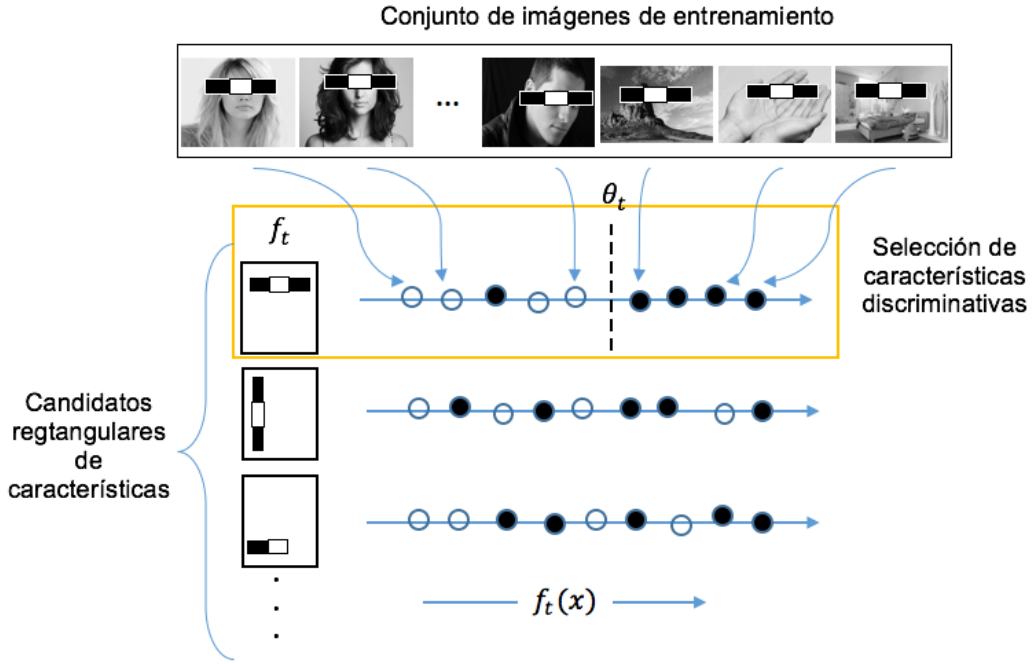


Figura 2.4: *Proceso de detección para identificar características rectangulares discriminatorias y determinar los clasificadores débiles, tomada de [2]*

1. Proceso de entrenamiento

   - Recolectar un conjunto de imágenes con ejemplos de rostros y otro conjunto de imágenes con ejemplos que no contiene rostros, a éstos se le conoce como ejemplos positivos y negativos, respectivamente.
   - Seleccionar varios candidatos rectangulares de características, donde cada candidato aplica el filtro Haar en la posición del rectángulo en la imagen que se desliza por la imagen. En la Figura 2.4, se muestra en columnas tres tipos de candidatos rectangulares y como salida puntos de intensidades blancos y grises, sumando al final los puntos de intensidades grises y restando los blancos, esto tiene el fin de obtener un banco de características para el conocimiento del modelo aprendido.

- Calcular una imagen integral para cada entrada.
- Calcular las respuestas de funciones rectangulares en todas las imágenes de entrenamiento.
- Utilizar el algoritmo AdaBoost [18] y las respuestas de características calculadas anteriormente para seleccionar secuencialmente un conjunto de características rectangulares discriminativos.

2. Proceso de reconocimiento

- Al entrar una imagen de prueba, el detector utiliza la ventana deslizante para la búsqueda de rostro.
- Al localizar un rostro en la imagen, calcula los resultados rectangulares de las características, utilizando una imagen integral.
- Clasifica los resultados de las ventanas si es cara o no cara de acuerdo a la salida binaria.

2.2.3.2. Reconocimiento de expresiones y micro-expresiones faciales

El reconocimiento de expresiones faciales de forma automática ha sido tema de investigación ya por varios años dentro del área de visión por computadora, mientras que recientemente en los últimos años se han enfocado en las micro-expresiones faciales.

El reconocimiento de expresiones y micro-expresiones faciales se basa generalmente en el aprendizaje supervisado, por medio de un conjunto de ejemplos etiquetados que sirven para entrenar un algoritmo de aprendizaje, el cual proporciona un modelo aprendido que determina el tipo de clase de expresión.

Figura 2.5: *Ejemplos de imágenes de expresiones faciales etiquetadas con su respectiva emoción tomada de http://k46.kn3.net/taringa/1/1/3/E/C/2/ robernetit 123/EB0.png*

En la Figura 2.5 se muestran las seis expresiones básicas en imagenes etiquetadas con su respectiva emoción: enojo, disgusto, miedo, alegría, tristeza y sorpresa. Mientras

tanto, en la Figura 2.6 se muestra un segmento de video o una secuencia de imágenes que contiene un leve movimiento aproximadamente a los 110ms del inicio del segmento del video que corresponde a una emoción de disgusto. Por lo tanto, en este trabajo se utiliza clasificación de imágenes para reconocer expresiones y clasificación de secuencias de imágenes para reconocer micro-expresiones.

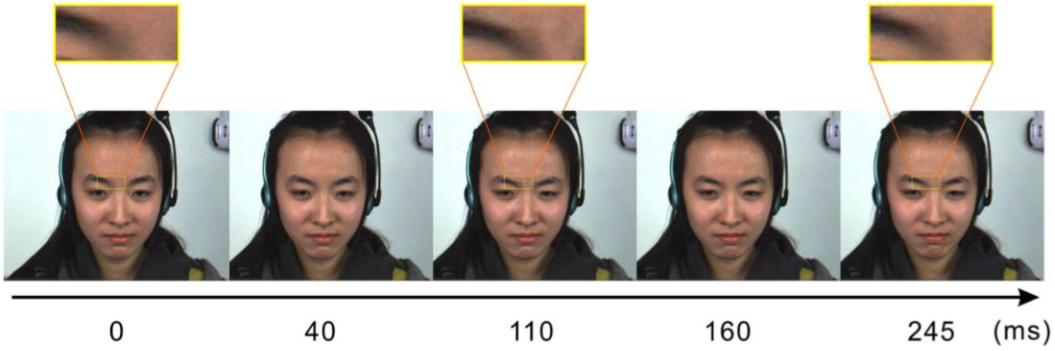


Figura 2.6: *Ejemplo de una secuencia de imágenes que contiene una micro-expresión, tomada de [3]*

2.3. Aprendizaje automático

La función principal del aprendizaje automático (AA) es desarrollar programas capaces de adaptar su comportamiento por medio de experiencia o ejemplos. Generalmente se usan conjuntos de datos de entrenamiento, a partir de los cuales se genera un modelo. Este modelo tiene el objetivo de predecir datos futuros, realizar alguna toma de decisión o ambos [4].

En la Figura 2.7 se muestra el proceso de adquisición de conocimiento de los modelos de aprendizaje.

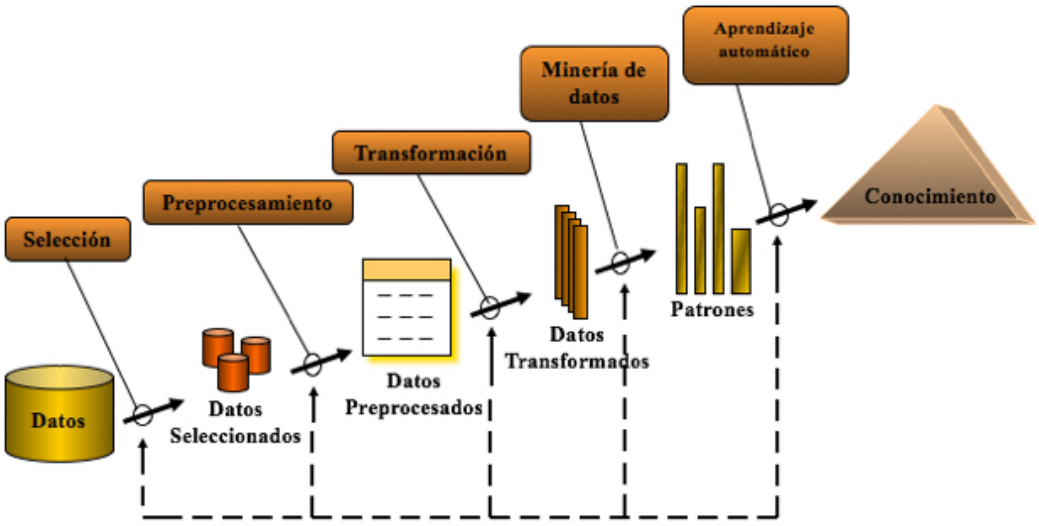


Figura 2.7: *Adquisición del conocimiento, tomada de [4]*

En la actualidad, se vive en la era de *big data* ya que existe una gran cantidad de datos.

Estos datos pueden ser de gran utilidad si son trabajados para una tarea específica y aquí es donde se basa el conocimiento de aprendizaje automático:

- Selección de datos: se crea un conjunto de ejemplos representativos.
- Pre-procesamiento: se preparan los datos para ser utilizados en pasos posteriores.
- Transformación: se representa a los datos de manera conveniente para ser analizados por procesos de minería de datos.
- Minería de datos: se analiza a los datos con el fin de encontrar patrones, que sirvan para discriminar la información.
- Aprendizaje automático: se lleva a cabo la aplicación de algoritmos que extraen modelos de conocimiento que permiten clasificar nuevos ejemplos.

Existen diferentes enfoques para realizar AA, entre ellos están el supervisado y no supervisado. A continuación se revisa el primer enfoque, siendo esté el utilizado por el presente proyecto.

2.3.1. Aprendizaje supervisado

El aprendizaje supervisado (AS) usa un conjunto de ejemplos de entrenamiento, donde cada ejemplo tiene su etiqueta que corresponde su dato de salida [4]. En este tipo de aprendizaje se emplean algoritmos cuyo objetivo es adquirir el conocimiento de la relación entre los ejemplos y su respectiva etiqueta, obteniendo un modelo aprendido definido por una función, como se muestra en la Ecuación 2.3, donde g es la función discriminante que separa los datos de ejemplo en su respectiva clase, x son las características extraídas del objeto y ∂ la etiqueta que le corresponde al objeto.

$$Y = g(x|\partial) \tag{2.3}$$

La Figura 2.8 muestra un ejemplo para aprender a reconocer un triángulo, mediante un conjunto de ejemplos de figuras geométricas, etiquetados correspondientemente con un *si*, si la figura es un triángulo y un *no*, si la figura no lo es. Cada ejemplo es analizado por un algoritmo de aprendizaje en donde obtiene características discriminativas con el fin de distingir los triángulos con las demás figuras geométricas de nuevos ejemplos.

2.3.1.1. Clasificación

Existen varios tipos de tareas en AA, uno de ellos es la clasificación. La función de un clasificador consiste en asignar un nuevo caso a una clase, asignándolas mediante reglas o por modelos aprendidos. Las reglas o características aprendidas por los modelos se les conoce como discriminantes, que son aquellas características distintivas de cada clase y ayudan a identificar a qué clase pertenece un nuevo caso.

En la actualidad existen muchos algoritmos que sirven como clasificadores, entre los más conocidos se encuentran: árboles de decisión, máquina de soporte vectorial (SVM),

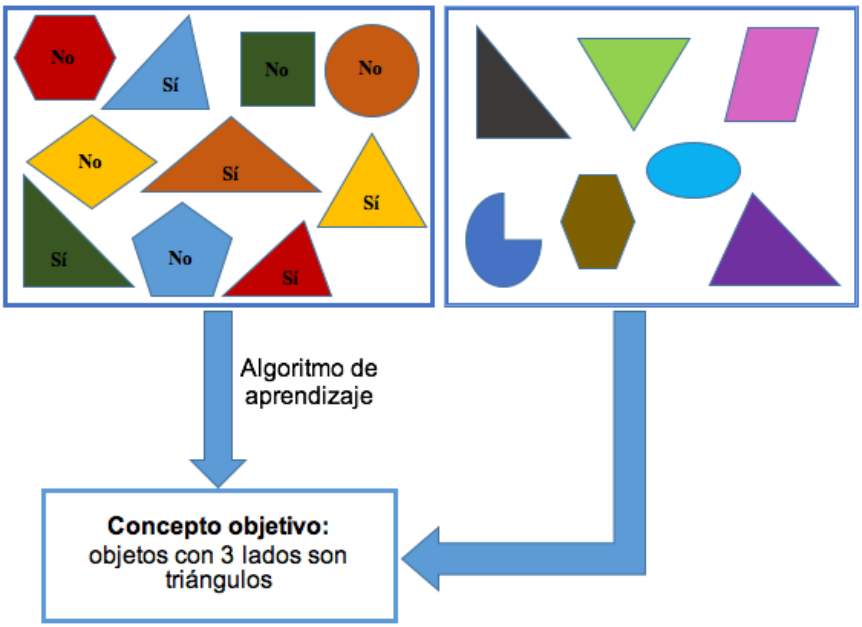


Figura 2.8: *Ejemplo de aprendizaje supervisado*

redes neuronales, redes bayesianas, entre otros. Debido a su gran éxito para resolver tareas relacionadas con VC, en la subsección 2.3.2 se aborda el aprendizaje profundo que puede ser definido como un tipo de enfoque de las redes neuronales.

2.3.1.2. Evaluación de modelos

Para evaluar el rendimiento del clasificador, es decir, si esté trabaja correctamente, se utilizan una serie de medidas [6], entre ella se tienen las siguientes:

- **Precisión** (Precision): es el número obtenido de positivo verdadero dividido por el total de resultados positivos. Si la precisión es 1, es porque no se obtuvo resultados falsos positivos.
- **Recuerdo** (Recall): es el número obtenido de positivo verdadero divido por el total resultados correctos, Si recuerdo es 1, es porque no se obtuvo resultados de falso negativo.
- **Exactitud** (Accuracy): es obtenida sumando los positivos verdaderos y los negativos verdaderos, es decir las respuestas correctas y dividir el número de ejemplo de entrenamiento.
- **Matriz de confusión** (Confusion matrix): en esta matriz es posible analizar por clase la cantidad de datos que el modelo aprendido confunde. En la Tabla 2.1, se presenta el caso para dos clases:
  - Si el nuevo caso es positivo y el resultado es positivo, se le conoce como verdadero positivo.
  - Si el nuevo caso es positivo y el resultado es negativo, se le conoce como falso negativo.

Tabla 2.1: *Matriz de confusión para dos clases [6]*

Clases Verdaderas	**Resultados Positivos**	**Resultados Negativos**	**Total**
Positivo	tp: verdadero positivo	fn: falso negativo	p
Negativo	fp: falso positivo	tn: verdadero negativo	n
Total	p'	n'	N

Tabla 2.2: *Medidas de funcionamiento para problemas de dos clases.*

Medición	**Fórmula**
error	$(fp + fn)/N$
exactitud	$\frac{(tp+tn)}{N} = 1 - error$
rango tp	tp/p
rango fp	fp/n
precisión	tp/p'
recuerdo	$\frac{tp}{p} = rangotp$

  - Si el nuevo caso es negativo y el resultado es negativo, se le conoce como verdadero negativo.
  - Si el nuevo caso es negativo y el resultado es positivo, se le conoce como falso positivo.

- **Medida F** (F-score o F-measure): es una medida que combina precisión y recuerdo llamada medida armónica, donde su fórmula es la siguiente (2.4).

$$F = 2 * \frac{precision * recuerdo}{precision + recuerdo} \tag{2.4}$$

- **Validación cruzada en k pliegues** (k- fold cross validation): una forma de validar los resultados, donde el conjunto de entrenamiento es dividido en k partes, donde su proceso es tomar $k - 1$ parte como conjunto de entrenamiento y una parte como conjunto de pruebas para obtener un resultado. Este proceso se realiza k veces tomando cada parte como el conjunto de pruebas y obtener k resultados. Estos resultados se promedian para obtener el resultado final.

2.3.2. Aprendizaje profundo

El aprendizaje profundo está basado en el concepto del algoritmo de redes neuronales artificiales (RNA) conformados por neuronas que permiten la entrada y salida de datos hacia otras neuronas. El principal próposito del aprendizaje profundo es el modelar abstracciones de alto nivel en datos usando arquitecturas compuestas de transformaciones no-lineales múltiples [12](Bengio et al., 2013). A manera de ejemplo gráfico en la Figura 2.9 se muestra una arquitectura de aprendizaje profundo.

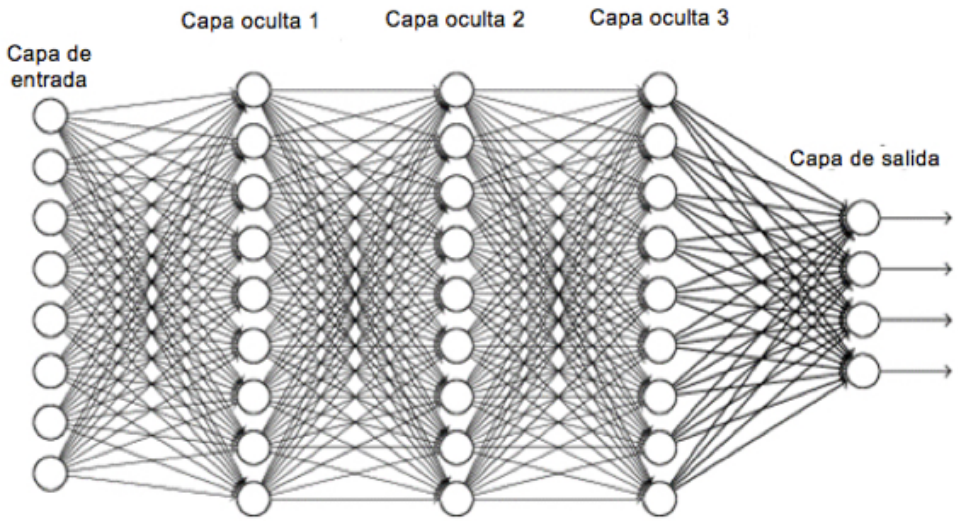


Figura 2.9: *Ejemplo de un diagrama de una red neuronal profunda, tomada de https:// www.extremetech.com/computing/227634-qualcomm-announces-new- deep-learning-sdk-with-support-for-snapdragon-820-heterogeneous-compute*

2.3.2.1. Redes convolucionales

Las redes convolucionales son usadas para procesar imágenes usando la arquitectura de aprendizaje profundo, estás aprenden relaciones de entrada-salida, donde la entrada es una imagen [19]. En la Figura 2.10, se tiene que a la imagen de entrada se le aplica una serie de filtros que son operaciones de convolución. Una convolución consiste en realizar un filtrado a una imagen usando una máscara, en donde la mácara contiene diferentes valores definidos para obtener resultados diferentes.

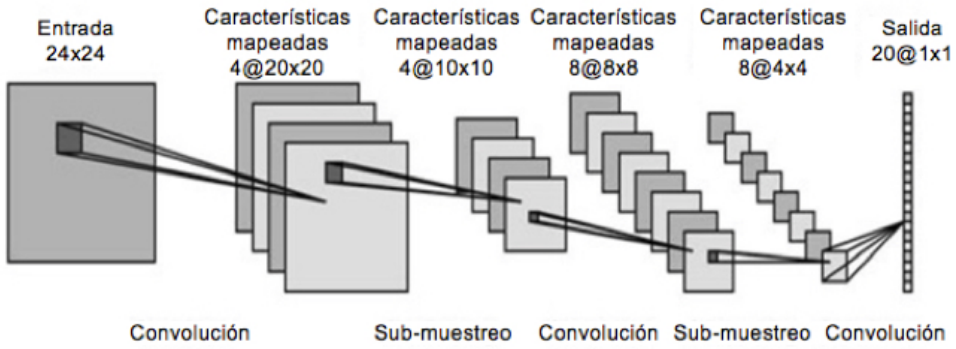


Figura 2.10: *Ejemplo gráfico de una red neuronal convolucional, tomada de https://www.linkedin.com/pulse/why-convolutioanl-neural-networks-shamane-siriwar dhana*

Las redes convolucionales son utilizadas para extraer características de un conjunto de imágenes de entrenamiento. Entre las tareas a las que se les aplican redes convolucionales están la detección o caracterización de objetos, clasificación de escenas y clasificación de imágenes en general.

2.4. Evaluación psicológica

La evaluación psicológica es la primera etapa que los psicólogos utilizan para identificar y solucionar un problema mental. En la presente subsección se describe el proceso actual que es realizado por los psicólogos para resolver problemas mentales. En las siguientes subsecciones se define en qué consiste la evaluación psicológica, la importancia de esta etapa y la confiabilidad de los datos adquiridos en la etapa de evaluación. Así mismo, se aborda la utilidad de la comunicación no verbal dentro de la evaluación psicológica.

2.4.1. Valoración del paciente

Los problemas mentales dañan el equilibrio de los individuos y de las personas que les rodean. Existen varios tipos de problemas mentales, por ejemplo: estrés, baja autoestima, fobia, personalidad múltiple, depresión, esquizofrenia, bipolaridad, entre otras. [20]

El proceso actual que los psicólogos utilizan para identificar y resolver estos problemas mentales, están basados en la observación y valoración del individuo. Este proceso inicia por la etapa de evaluación psicológica del individuo, basada en la comunicación experto-paciente, donde el experto se apoya por cuestionarios y actividades, con el objetivo de adquirir toda la información posible del paciente, para después elaborar un informe claro y preciso de los datos obtenidos en la evaluación. Con base en este informe, el especialista puede obtener conclusiones y tomar decisiones, con el objetivo de darle solución al problema o problemas mentales del paciente [9].

2.4.2. Comunicación no verbal

La expresión no verbal es una manera de comunicación considerada dentro de la psicoterapia, que nos proporciona información útil para conocer aspectos psicológicos de una persona, como son actitudes, afectos, interés, entre otros [21].

Este tipo de comunicación es transmitido por:

- Kinésica: estudia el significado de movimientos corporales, por ejemplo: la mirada, gesto, postura y contacto corporal, y transmiten aspectos afectivos.
- Paralingüística: estudia la forma o el modo del habla, por ejemplo: los tonos empleados al hablar, el ritmo, el volumen, silencio y el timbre de voz (agudo, grave, etc.)
- Proxémica: estudia el espacio personal.

Las expresiones y micro-expresiones faciales se encuentran dentro del tipo comunicación Kinésica, las cuales se asocian en transmitir emociones.

2.4.2.1. Expresiones y micro-expresiones faciales

Las expresiones y micro-expresiones son medios de comunicación de forma no verbal. Es la forma más natural e inmediata en que los individuos comunican sus emociones.

La expresión facial es una manifestación visible del estado afectivo, actividad cognitiva, intención, personalidad y psicopatológica de una persona [22]. La diferencia entre las expresiones con micro-expresiones son: que las expresiones faciales se manifiestan voluntariamente y es fácil disfrazarlas con el fin de ocultar las emociones reales, mientras que las micro-expresiones faciales se caracterizan por ser de corta duración, baja intensidad y se presentan de forma involuntaria revelando una emoción genuina [10].

2.4.2.2. Antecedentes de expresiones faciales

"*El rostro es reflejo de las emociones y éstas nos permiten inferir las causas que las originaron con una clara utilidad en diferentes ámbitos* "[23]. Por lo tanto, las expresiones faciales ejercen un papel muy importante en la interacción social, que ha atraído el interés de varios investigadores desde el año 1873 por Darwin [24], hasta en la actualidad, convirtiendosé en un campo de estudio.

Aproximadamente en los años 70's, el estudio de las expresiones faciales toma un nuevo enfoque sistemático. Este enfoque permitió que Ekman y compañeros desarrollaran un método con el objetivo de codificar ciertos músculos faciales para identificar diferentes emociones, llamado sistema de codificación de acción facial (FACS en inglés *Facial Action Coding System*) [25].Las FACS es uno de los métodos más usados actualmente para el etiquetado de las expresiones faciales y es utilizado para discriminar y clasificar expresiones, dando entrada al reconocimiento de expresiones faciales (FER en inglés *Facial Expressions Recognition*).

En los 90's el auge de los algoritmos de aprendizaje para tareas de visión por computadora, comenzaron a surgir como detección de patrones faciales, identificación facial y analisis de expresiones faciales, con el objetivo de utilizarlos en aplicaciones para identificación criminal, autenticación en sistemas seguros, etc. [26].

Actualmente, FER de manera automática es una tarea que ya ha alcanzado niveles de exactitud altos, sin embargo aún existen investigaciones que intentan mejorar su rendimiento. Las técnicas más usadas para desarrollar FER son HOG, LBP y Gabor como extractores de características, SVM y RNA como clasificadores y por último CNN utilizada para ambas. Se apoyan de bases de datos con imágenes de expresiones faciales para el conjunto de entrenamiento y su aprendizaje de manera supervisada, para reconocer emociones como enojo, sorpresa, disgusto, miedo, alegría, tristeza, depresión, vergüenza, etc., obteniendo un rendimiento superior del 90 %. Trabajos como [27], [28] y [29].

2.4.2.3. Antecedentes de micro-expresiones faciales

La idea de micro-expresión apareció en el año 1872, en una hipótesis de Darwin, en donde se sugiere que las expresiones faciales no pueden ser controladas, y pueden mostrarse expresiones involuntarias a pesar que la persona desee controlarlas [30].

Las primeras micro-expresiones fueron detectadas por los psicólogos Haggards e Isaacs en el año 1966 [31]. Durante el análisis de videos de pacientes con problemas psicológicos, se descubrieron expresiones faciales involuntarias, a estas se les llamaron expresiones micromomentarias (*Expression Micromomentary*).

Tres años después, Ekman y Frisen [21] públican el hallazgo de micro-expresiones acuñando el término. En dicho trabajo también observaron en un video a cámara lenta a un paciente con tendencias suicidas. Durante el transcurso del video el paciente se mostraba feliz, sin embargo, se detectaron leves movimientos faciales hechas en fracciones de segundo, que parecían expresar angustia. Con los primeros hallazgos de micro-expresiones, el interés de investigación fue en aumento, destacando los siguiente trabajos [32] y [33].

Posteriormente Porter y Brinke en el año 2008 [34], corroboraron la existencia de leves expresiones faciales involuntarias; ellos encontraron rasgos que no correspondían a expresiones genuinas en personas que intentaron fingir sus emociones. Reafirmado el trabajo de Ekman, que después de haber llevado a cabo varios trabajos sobre ME faciales, pública un artículo que recalca que las micro-expresiones es un gran método para la detección de mentiras [10], provocando que sea tomada en cuenta para diversas aplicaciones.

Por otro lado, investigaciones en visión por computadora ya habían realizado algoritmos eficientes con resultado mayores del 90 % sobre la detección y reconocimiento de expresiones faciales, en donde se cuenta con el apoyo de una gran variedad de bases de datos. Como es el caso de [35] y [27].

Con base al conocimiento adquirido en la detección y reconocimiento de expresiones faciales, [36] implementó un método para su detección y realizaron una modificación para incluir la detección de micro-expresiones. Apoyándose con dos bases de datos de expresiones faciales, este trabajo obtuvo resultados favorables para la detección de las expresiones, pero no en la detección micro-expresiones.

Actualmente, diversos grupos de investigación trabajan en la detección de micro-expresiones, donde aún no se ha desarrollado un algoritmo lo suficientemente efectivo para resolver la tarea. A la par, el desarrollo de BD para micro-expresiones ha ido en aumento, por ejemplo la BD de Polikovsky [37], SMIC [38], CASME [39], tres BD de SMIC (HS,VIS y NIR) [40] y CASME II [3]. Sin embargo, la complejidad de estas BD va de clasificar entre tres a cinco emociones ligadas a diversas micro-expresiones.

2.5. Trabajos relacionados

A continuación se describen trabajos relacionados con modelos de reconocimiento de expresiones faciales y micro-expresiones faciales.

2.5.1. Trabajos de reconocimiento de expresiones faciales

En el año 2015, Kumari, J., et al [41] presentaron una revisión de reconocimiento de expresiones faciales y un estudio comparativo de 4 técnicas de extracción de características: LBP, LGC, HOG y LDP. Los autores usaron un conjunto de 196 imágenes de la base de datos JAFFE. Además, implementaron los algoritmos de detección facial de Viola-Jones, el método de validación de uno afuera (leave-one-out) y KNN (K-Nearest Neighbor) como clasificador, con valor de parámetro K igual a 2. Se llevaron a cabo dos pruebas usando

imágenes 128x128 con bloques de 16x16 y bloques de 8x8. Los resultados mostraron que al LGC con una exactitud de 87.25 % y 88.77 %.

En en año 2015, De la Torre, F., et al [42] presentaron un paquete de software llamada Intraface (IF) que puede estimar posición de la cabeza, reconocer atributos faciales y analizar expresiones faciales de un video. En el reconocimiento de expresiones faciales, IF es capaz de reconocer ocho tipos de expresiones, que son: alegría, enojo, tristeza, miedo, sorpresa, disgusto, desprecio y verguenza, usando HOG para la extracción las características y máquina de tranferencia selectiva (STM en inglés *Selective Transfer Machine*) como clasificador. Utilizando dos bases de datos CK+ y FERA, obteniendo resultados de exactitud para CK+ del 96.4 % y FERA el 86.5 %.

En el año 2016, Peng, X., et al [43] propusieron una nueva base de datos construida de expresión facial natural con imágenes etiquetadas por usuarios de la web y un modelo basado en entrenamiento profundo sobre el conjunto de imágenes obtenido de la Web. Utilizaron un modelo con varios filtros para pulir el conjunto de imágenes salidas de buscadores web, conservando solo 350 imágenes para cada expresión universal (de 6 expresiones). En el reconocimiento de expresión facial, utilizaron un modelo pre-entrenado llamado VGG-face, en donde se redefinió la capa de salida y se entrenó las últimas dos capas ocultas. Los resultados fueron favorables en comparación con otras técnicas de extracción de características y clasificadores, donde la emoción de alegría tuvo el mayor exactitud del 98.3 %.

En el año 2016, Mollahosseini, A., et al [29] propusieron utilizar una arquitectura de red neuronal profunda para abordar el reconocimiento de expresión facial a través de múltiples conjuntos de datos supervisados. Para ello, usaron una CNN llamada AlexNet y una adaptación de su parte modificando su entrenamiento mediante más número de iteraciones y épocas. Utilizando varias bases de datos como MultiPIE, SFEW, MMI, CK+ y DISFA, para los seis tipos de expresiones básicas que son: disgusto, miedo, sorpresa, alegría, tristeza y enojo, obteniendo como mejor resultado su CNN propuesta en las base de datos MMI con 77.9 % contra un 56 % y con FER2013 con 66.4 % contra in 61.1 % de exactitud.

En el año 2016, Mollahosseini, A., et al [44] dieron respuesta a dos preguntas: 1. ¿cuánta consistencia existe en usar etiquetas de expresiones faciales tomadas de la web? y ¿cuánta exactitud tiene un algoritmo de clasificación de imágenes representándolas con características del estado del arte?. En la primera pregunta se recolectaron imágenes de la web a través de consultas usando seis expresiones universales. Los resultados obtenidos fueron que la mayoría de la emociones recolectadas solo 50 % hacen correspondencia con la anotación de dos expertos, con excepción de la emoción de alegría que obtuvo un 68.18 % de consistencia. En la segunda pregunta se entrenaron dos arquitecturas de redes de aprendizaje profundo llamadas AlexNet y WACV-Net. Los resultados obtenidos con el entrenamiento de AlexNet son de 82.1 % y WACV-Net de 76.52 % de exactitud, respectivamente.

Tabla 2.3: *Comparación de modelos para el reconocimiento de expresiones faciales*

Autor	Extracción de características	Clasificador	Base de Datos	Resultados
[43]	DCNN	DCNN	Imágenes Web	98.3 %
[42]	HOG	STM	CK+	96.4 %
[29]	adaptada de AlexNet	adaptada de AlexNet	CK+	93.2 %
[29]	AlexNet	AlexNet	CK+	92.2 %
[41]	LGC	KNN	JAFFE	88.77 %
[42]	HOG	STM	FERA	86.5 %
[44]	AlexNet	AlexNet	Imágenes Web	82.1 %
[29]	adaptada de AlexNet	adaptada de AlexNet	MMI	77.9 %
[29]	AlexNet	AlexNet	MMI	56.0 %

2.5.2. Trabajos de Reconocimiento de Micro-expresiones faciales

En el año 2011, Pfister, et al [38] construyeron una nueva base con micro-expresiones espontáneas (SMIC) y en su evaluación propusieron un modelo de interpolación temporal (TIM) para mejorar la detección y reconocimiento de micro-expresiones. Como resultado se obtuvo una base de datos con 77 micro-expresiones que permite reconocer emociones positivo y negativo y el primer sistema en reconocer micro-expresiones espontáneas faciales, logrando como resultados 74.3 % en detección usando RF+TIM10 y el 71.4 % en reconocimiento utilizando MKL+TIM10.

En el año 2013, Li, et al [40] propusieron una base de datos de micro-expresiones espontáneas que incluyen 164 video clips de micro-expresiones de 16 participantes. Para obtener el conjunto de micro-expresiones espontáneas se consideraron dos aspectos relevantes: primero utilizaron estímulos para provocar emociones a los participantes lo suficientemente fuertes, para que lo demostraran en el rostro y el segundo los participantes deberían sentir suficiente presión para motivarlos a ocultar sus verdaderos sentimientos. Como resultado se obtuvieron tres clases de emociones (positivo, negativo y sorpresa) y tres tipos de bases de datos SMIC: i) se utilizó una cámara de alta velocidad (HS) (PixeLINK PL-B774U, 640x480) de 100 fps; ii) utilizaron una cámara visual normal (VIS); iii) una cámara de infrarrojo cercano (NIR), ambas con 25 fps. Se obtenieron como mejores resultados de las evaluaciones en baseline en detección y reconocimiento de micro-expresiones utilizando LBP-TOP+TIM10 para extracción de características y SVM como clasificador, obteniendo lo siguiente: HS detección 65.55 % y reconocimiento 48.78 %, VIS detección 62.8 % y reconocimiento 52.11 % y por ultimo NIR en detección 59.15 % y reconocimiento 38.03 %.

En el año 2014, Yan W-J, et al [45] propusieron el desarrollo de una base de datos de micro-expresiones espontáneas realizando una rigurosa detección de micro-expresiones, codificación de la unidad activa (AU) y el etiquetado de micro-expresiones. Se seleccionaron 195 micro-expresiones para la base de datos con AUs y emociones etiquetadas. Para la evaluación, se empleo LBP-TOP para la extracción de características y SVM para el clasificador con el método de validación cruzada. El mejor resultado obtenido es 61.32 % para una clasificación de 5 clases.

Tabla 2.4: *Comparación de modelos para el reconocimiento de micro-expresiones faciales*

Autor	Extracción de características	Clasificador	Base de Datos	Resultados
[47]	HIGO+Mag*	SVM linear	SMIC-VIS	83.10 %
[38]	TIM10	MKL	SMIC	71.4 %
[3]	LBP-TOP	SVM	CASME II	63.41 %
[45]	LBP-TOP	SVM	CASME	61.32 %
[40]	LBP-TOP + TIM10	SMV	SMIC-HS	52.11 %
[40]	LBP-TOP + TIM10	SMV	SMIC-HS	48.78 %
[46]	DTSA	ELM	CASME	46.90 %
[40]	LBP-TOP + TIM10	SMV	SMIC-HS	38.03 %

Tabla 2.5: *Comparación de base de datos de micro-expresiones faciales*

Autor	Base de Datos	Número de Micro-expresiones	Número de ejemplos	Calidad de Reproducción
[45]	CASME	8	195	60 fps
[3], [47]	CASME II	5	247	200 fps
[38]	SMIC	2	77	100 fps
[40], [47]	SMIC-HS	3	164	100 fps
[40], [47]	SMIC-VIS	3	71	25 fps
[40], [47]	SMIC-NIR	3	71	25 fps

En el año 2014, Wang, et al [46] propusieron un nuevo algoritmo basado en el reconocimiento de tensor discriminante del espacio sub-análisis (DTSA) y la máquina de aprendizaje extrema (ELM). Las características discriminantes se generaron utilizando DTSA para mejorar aún más el rendimiento de la clasificación de ELM. Los resultados experimentales sobre las bases de datos faciales de ORL, Yale, YaleB y la base de datos de micro-expresión CASME muestran la eficacia del método propuesto, dando como resultado el 46.90 %.

En el año 2014, Yan W-J, et al [3] desarrollaron una base de datos de micro-expresiones espontáneas llamada CASME II, con una alta resolución temporal (200 fps) y resolución espacial (280 x 340 pixeles en el área facial). En donde incitaron expresiones faciales a los participantes en un laboratorio controlado y con una iluminación adecuada. Entre casi 3000 movimientos faciales, se seleccionaron 247 micro-expresiones para la base de datos con unidades de acción (AUs) y emociones etiquetadas. Para la evaluación se implementó LBP-TOP para la extracción de características y SVM para el clasificador con el método de validación cruzada. El mejor resultado obtenido fue 63.41 % para una clasificación de 5 clases.

En el año 2015, Li, X., et al [47] propusieron un método para la detección de las ME espontáneas en vídeos de larga duración, un marco de trabajo en reconocimiento de ME utilizando las bases de datos SMIC (NIR, VIS y HS) y CASME II. Además propusieron un sistema de análisis de ME automático (MESR) que puede detectar y reconocer las ME de datos de videos espontáneos. El mejor resultado obtenido por el método de detección de ME fue usando LBP en CASME II del 92.98 % de exactitud. Para el método de reconocimiento de ME utilizando HIGO+Mag en SMIC-VIS se obtuvo 81.69 % de exactitud y para

el sistema MESR utilizando SMIC-VIS un 42.42 % de exactitud.

A pesar de que la mayoría de los trabajos relacionados resaltan la variedad de aplicaciones de uso de las expresiones y micro-expresiones faciales, no se conoce una aplicación como herramienta de apoyo para sesiones psicológicas.

III

Metodología

Capítulo 3: Metodología

En este capítulo se describe la metodología que está dividida en dos etapas, comó se muestra en la Figura 3.1. La primera es la etapa experimental para conseguir los métodos para reconocer expresiones y micro-expresiones faciales y la segunda etapa comprende la contrucción e implementación del prototipo.

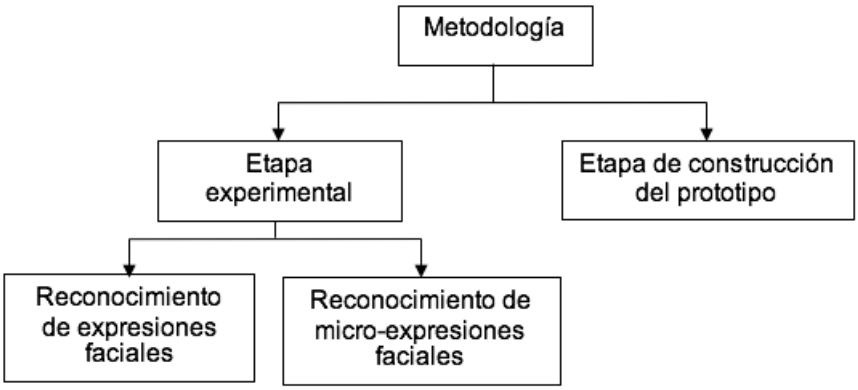


Figura 3.1: *Metodología empleada.*

3.1. Etapa experimental

En esta sección se describen los experimentos a llevar a cabo para el reconocimiento de expresiones y micro-expresiones faciales. La etapa experimental que se muestra en la Figura 3.2, consiste de cinco pasos que son: adquisición de bases de datos, pre-procesamiento de datos, extracción de características, clasificación y evaluación, y a continuación se describen cada uno de ellos:

Figura 3.2: *Pasos de la etapa experimental*

1. **Adquisición de bases de datos:** se seleccionan conjuntos de ejemplos etiquetados de expresiones y micro-expresiones faciales que servirán como entrenamiento para el aprendizaje de modelos. Los siguientes consideraciones son tomadas en cuenta para elegir las bases de datos:

   - Tipo de imagen: las imágenes deben contener el rostro frontal de una persona, con una iluminación que permita distinguir todo el área principal del rostro.
   - Diversidad: es recomendable obtener mayor variedad de imágenes, de preferencia con personas de distintas razas, edad y sexo.
   - Etiquetado: debido a que el tipo de aprendizaje tiene un enfoque supervisado, es necesario que las imágenes estén etiquetadas.
   - Cantidad de expresiones: se requiere que seis expresiones básicas estén presentes en las bases de datos: enojo, disgusto, alegría, miedo, tristeza y sorpresa.

2. **Pre-procesamiento de datos:** es necesario preparar todas las imágenes de tal manera que se facilite su posterior procesamiento. Por ejemplo: ubicar y recortar solo el área de interés de las imágenes, es decir, el rostro. En específico, en la Figura 3.3 se muestra los procesamientos realizados en la imagen:

   - Escala de color: no es necesario conocer el color de piel, ojos, cabello, etc., por lo tanto, si la imagen es de color pasa a escala de grises, para reducirla a un solo plano y minimizar recursos.
   - Detección facial: si se ubica el área del rostro, es posible minimizar los datos a analizar y descartar datos irrelevantes.
   - Recorte: con la intención de eliminar ruido y minimizar recursos, la imagen se recortará de tal manera que solo utilizaremos el rostro.
   - Estandarizar imágen: se normalizan a un mismo tamaño todas las imágenes.

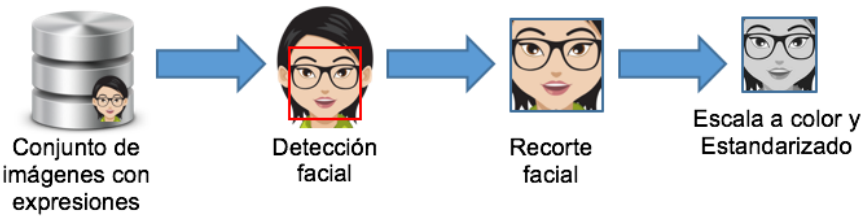


Figura 3.3: *Ejemplos de tareas en la etapa del pre-procesamiento de datos*

3. **Extracción de características:** se obtiene una selección de atributos que sirven como representación de las imágenes de los ejemplos obtenidos de las bases de datos. Se empleará una red neuronal convolucional (CNN siglas de *Convolutional Neural Network*) para llevar a cabo dicha extracción de características. El marco de trabajo para llevar a cabo este paso se muestra en la Figura 3.4 y se describe de manera general en los puntos a continuación:

- Entrada: imagen que solo contiene el rostro a analizar con un tamaño cuadrado de *nxn*.
- Procesamiento: la imagen pasa por una serie de transformaciones conformado por convoluciones y sub-muestreos.
- Salida: se obtiene un vector (1 x *m*) que resume características discriminativas de la imagen.

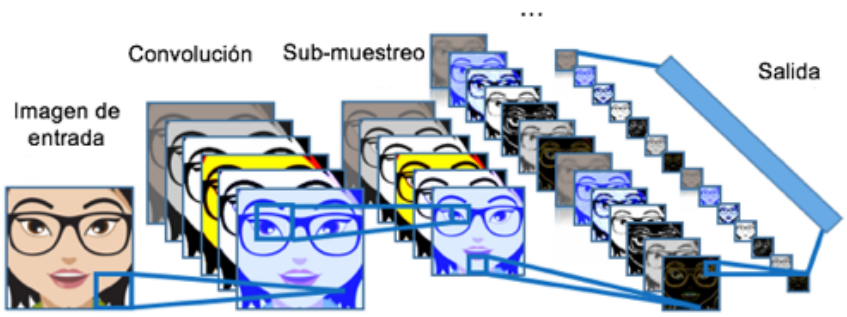


Figura 3.4: *Arquitectura general de CNN para la extracción de características de una imagen*

4. **Clasificación:** los vectores de características de cada imagen obtenidos en el paso anterior, sirven como datos de entrenamiento para los algoritmos de aprendizaje, teniendo como objetivo el de aprender modelos capaces de clasificar correctamente nuevas imágenes de expresiones y micro-expresiones faciales. El marco de trabajo para llevar a cabo este paso se muestra en la Figura 3.5 y se describe en los puntos a continuación:

   - Entrada: vectores de características de cada imagen del conjunto de ejemplos de entrenamiento.
   - Procesamiento: vectores de características pasa por un algoritmo de aprendizaje que sirve como entrenamiento para conocer y aprender el problema.
   - Salida: crea un modelo capaz de clasificar correctamente nuevos ejemplos, sin haberlos usado en el entrenamiento.

Figura 3.5: *Marco de trabajo para la clasificación*

5. **Evaluación:** por último, se evalúa el modelo aprendido usando conjuntos de imágenes de pruebas, además de aplicar medidas de evaluación. El marco de trabajo para llevar a cabo este paso se muestra en la Figura 3.6 y se describe en los puntos a continuación:

   - Entrada: conjunto de imágenes de prueba etiquetadas.
   - Proceso: las imágenes pasan por el modelo aprendido y son clasificadas.
   - Evaluación: se verifica a las clases proporcionadas por el modelo aprendido para cada imagen con su correspondencia a las etiquetas.

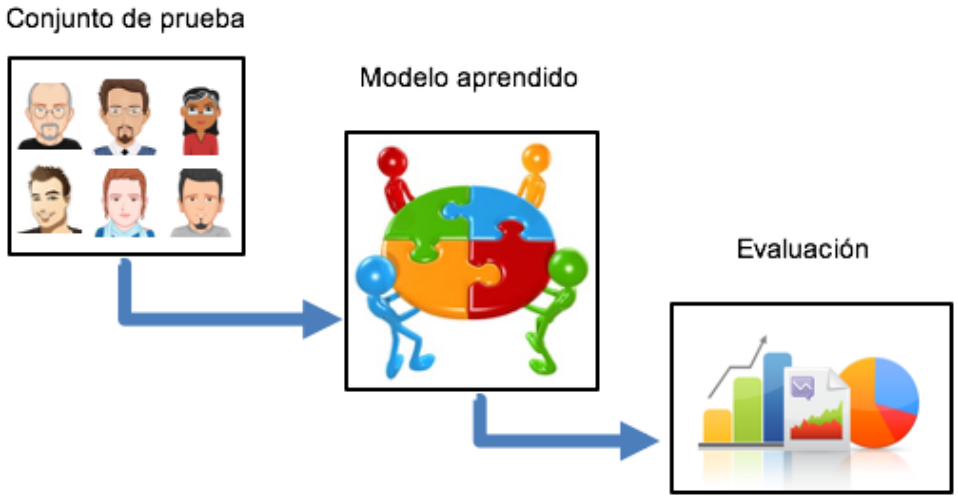


Figura 3.6: *Marco de trabajo para la evaluación*

3.1.1. Método para el reconocimiento de expresiones faciales

Empleando la metodología de la etapa experimental (Figura 3.2), a continuación se describen los pasos escenciales para llevar a cabo esta etapa:

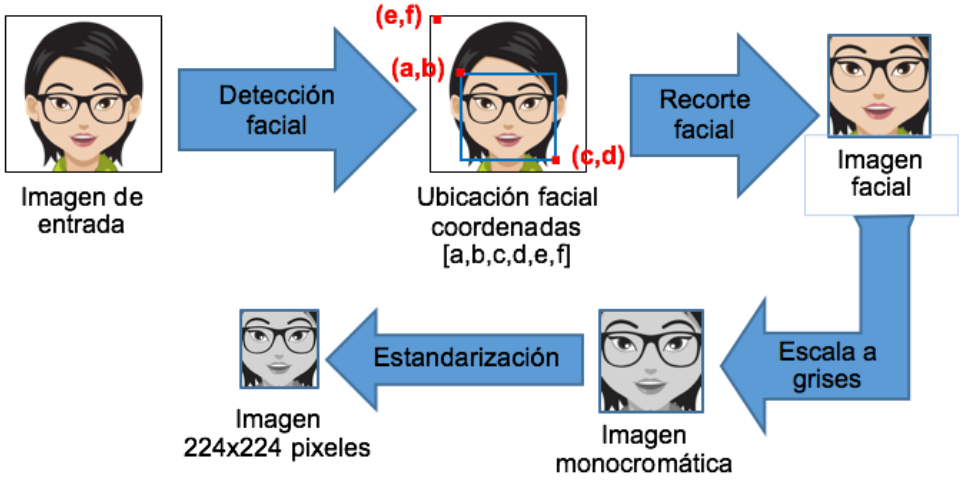


Figura 3.7: *Marco de trabajo usado para el pre-procesamiento de imágenes en el reconocimiento de expresiones faciales*

- **Pre-procesamiento de imágenes:** con base a la Figura 3.7, se muestra un marco de trabajo donde una imagen digital es procesada por un algoritmo de detección facial "*face_model*"[48], dando como dato de salida un vector de seis elementos que corresponden a tres pares de coordenadas. Estos tres pares de coordenadas indican la parte superior izquierda, la parte inferior derecha del área de rostro y el último par de coordenadas el inicio de la imagen, respectivamente. Posteriormente, con los tres pares de coordenadas se realiza el recorte para obtener una nueva imagen que contenga solo el área facial. Sobre la nueva imagen se aplica una conversión a escala de grises y se realiza una estandarización que modifica la imagen a tres planos en escala a grises a un tamaño de 224 píxeles de alto y 224 píxeles de ancho (3@224x224).

- **Extracción de características:** se utiliza la CNN "*VGG-face*"[14] entrenada con la finalidad de identificar rostros. A este proceso de reutilización de modelos entrenados se le conoce como transferencia de conocimiento (en inglés *transfer learning*). En la Figura 3.8 se muestra gráficamente su proceso, donde el dato de entrada es la imagen digital de 3@224x224 píxeles obtenida en el paso anterior, la imagen pasa por las primeras 38 capas de la CNN formada por convoluciones, sub-muestreos y capas de unidad de rectificación (ReLU). Como dato de salida se obtiene un vector de características con un tamaño de 4096 elementos que representa a la imagen.

Figura 3.8: *Arquitectura empleada para la extracción de características de una imagen*

Por último, para el vector de características obtenido de la imagen es necesario añadir su etiqueta, definido como se muestra en la Ecuación 3.1:

$$img = \begin{bmatrix} a_1 & a_2 & a_3 & \dots & a_{4096} \end{bmatrix} \cup [etiqueta] \tag{3.1}$$

donde la imagen *img* esta representada por $a_1, a_2, \dots, a_{4096}$ que corresponden a las caracteríticas extraídas, concatenando su etiqueta correspondiente.

- **Clasificación:** se emplean algoritmos de aprendizaje clásicos, como Naïve Bayes (NB), SVM, RandomForest (RF), y uno basado en *K* Instancias (IBK, por sus siglas en inglés).

- **Evaluación:** en esta etapa se utiliza validación cruzada a 10 pliegues, repitiendo

cinco veces la validación modificando la semilla aleatoria. Se usan las medidas de exactitud, precisión, recuerdo y medida F, para evaluar el desempeño.

3.1.2. Método para el reconocimiento de micro-expresiones faciales en secuencia de imágenes

Empleando la metodología de la etapa experimental (Figura 3.2), a continuación se describe a detalle los pasos principales:

- **Pre-procesamiento de imágenes:** en este paso se realizan las mismas tareas del método de reconocimiento de expresiones faciales, sin embargo la diferencia en pre-procesamiento se produce en aplicar este paso a *n* imágenes, de acuerdo al número de imágenes que contiene una secuencia. La Figura 3.9 muestra el marco de trabajo de este pre-procesamiento.

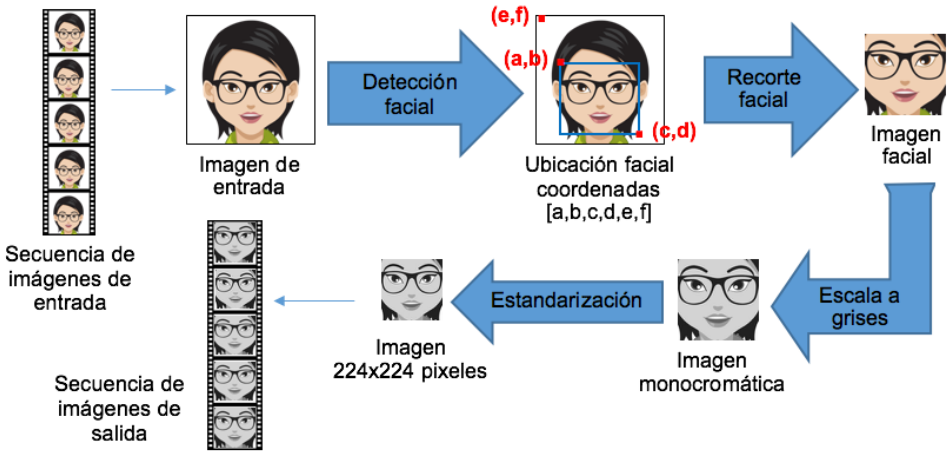


Figura 3.9: *Marco de trabajo usado para el pre-procesamiento de imágenes en el reconocimiento de expresiones faciales*

- **Extracción de características:** para esta paso se usa la CNN "*VGG-face*"[14], como se describe en la Figura 3.10, de manera general:

  1. Datos de entrada: se recibe una secuencia de *n* imágenes, donde cada imagen con un formato de 3@224x224 será procesada por la CNN.
  2. Proceso: se utilizan las primeras 38 capas de la CNN conformado por: convoluciones, sub-muestreos y capas de unidad de rectificación (ReLU).
  3. Dato de salida: para cada imagen se obtiene un vector de características de un tamaño de 4096 elementos, por lo tanto, para cada secuencia de *n* imágenes se obtiene una matriz de *n* x 4096 elementos.

  Sin embargo, para la clasificación es necesario que la representación de cada ejemplo sea un vector representativo con su etiqueta correspondiente para un aprendizaje supervisado. Por lo que se proponen tres estrategias diferentes para conocer cual representa mejor el movimiento en la secuencia de imágenes, estas tres estrategias se describen a continuación:

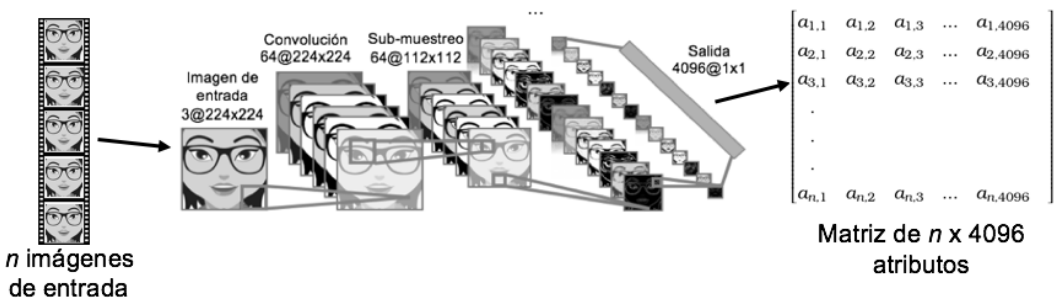


Figura 3.10: *Arquitectura empleada para la extracción de características en una secuencia de imágenes*

- **Promedio.** La primera estrategia consiste en aplicar promedio o la media a la matriz como se define en la Ecuación 3.2, es decir, se obtiene la media (m) para cada columna obteniendo un vector de 4096 elementos y se añade su etiqueta correspondiente.

$$m\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \dots & a_{1,4096} \\ a_{2,1} & a_{2,2} & a_{2,3} & \dots & a_{2,4096} \\ . & & & & \\ . & & & & \\ . & & & & \\ a_{n,1} & a_{n,2} & a_{n,3} & \dots & a_{n,4096} \end{bmatrix} = \begin{bmatrix} m_1 & m_2 & m_3 & \dots & m_{4096} \end{bmatrix} + [etiqueta] \tag{3.2}$$

- **Desviación estándar.** La segunda estrategia usa la desviación estándar de la matriz como se define en la Ecuación 3.3, es decir, se obtiene la desviación estándar (std) para cada columna obteniendo un vector de 4096 elementos y se añade su etiqueta correspondiente.

$$std\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \dots & a_{1,4096} \\ a_{2,1} & a_{2,2} & a_{2,3} & \dots & a_{2,4096} \\ . & & & & \\ . & & & & \\ . & & & & \\ a_{n,1} & a_{n,2} & a_{n,3} & \dots & a_{n,4096} \end{bmatrix} = \begin{bmatrix} std_1 & std_2 & std_3 & \dots & std_{4096} \end{bmatrix} + [etiqueta] \tag{3.3}$$

- **Fusión temprana.** La tercera estrategia es basada en fusión temprana, donde el resultado de la ecuación media (Ecuación 3.2) y la ecuación de desviación estándar (Ecuación 3.3) son concatenadas, obteniendo así un vector de 8192 elementos y se añade su etiqueta correspondiente, como se define en la Ecuación 3.4.

$$\begin{bmatrix} m_1 & m_2 & m_3 & \dots & m_{4096} \end{bmatrix} \cup \begin{bmatrix} std_1 & std_2 & std_3 & \dots & std_{4096} \end{bmatrix} = \begin{bmatrix} m_1 & \dots & m_{4096} & std_{4097} & \dots & std_{8192} \end{bmatrix} + [etiqueta] \tag{3.4}$$

Finalmente se realiza una evaluación de clasificación similar a la definida en el

método de reconocimiento de expresiones faciales empleando algoritmos clásicos.

3.2. Etapa de contrucción del prototipo

En esta sección se describen los pasos empleados para llevar a cabo la construcción del prototipo, que facilita al especialista la búsqueda de expresiones y micro-expresiones faciales en un video. La etapa consiste en cuatro pasos que se describen a continuación:

1. **Obtención de modelos aprendidos:** para implementar e integrar los modelos aprendidos obtenidos en la etapa de experimentación, se define la plataforma y el lenguaje de programación que se utilizará para la construcción del prototipo. Los modelos aprendidos en la etapa experimental son empleados en el prototipo.

2. **Estructura del prototipo e integración de los modelos:** descripción del funcionamiento del prototipo propuesto para la obtención de resultados de expresiones y micro-expresiones faciales. La Figura 3.11 muestra los datos que se debe considerar para su desarrollo, donde es necesario definir datos de entrada, proceso, integración de los modelos de expresiones y micro-expresiones y datos de salida.

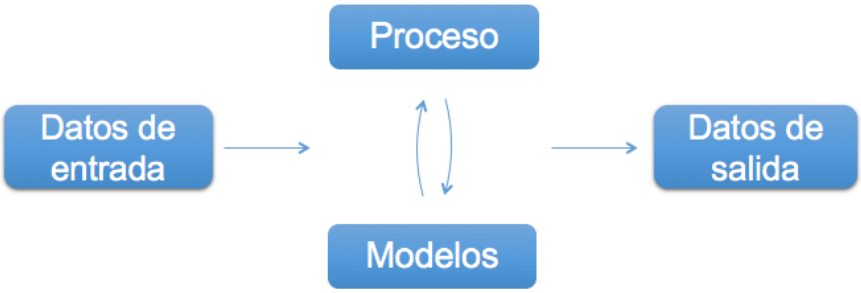
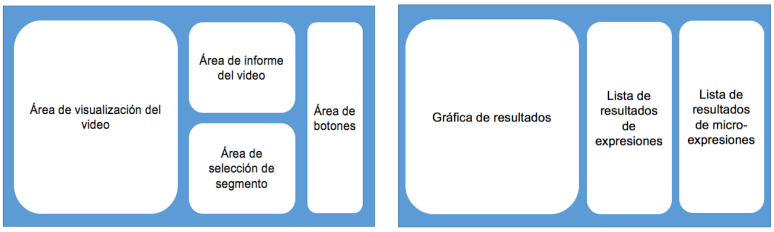


Figura 3.11: *Datos a considerar para la estructura del prototipo*

3. **Diseño de interfaz gráfica:** representación gráfica del prototipo con los elementos necesarios que se implementarán. El prototipo se concentra en dos interfaces: la primera interfaz que se muestra en la la Figura 3.12(a) contendrá los elementos de selección y visualización del video y la segunda interfaz que se muestra en la Figura 3.12(b) contendrá todos los resultados obtenidos del analisis del video.



(a) Interfaz principal (b) Interfaz de resultados

Figura 3.12: *Interfaces para el desarrollo del prototipo*

4. **Funcionamiento del prototipo:** elaboración de nuevos videos para la búsqueda de expresiones y micro-expresiones faciales por medio del protitipo.

IV

Resultados

4.1. Resultados de la etapa experimental

Esta sección muestra los resultados obtenidos de los experimentos realizados para la obtención de los modelos aprendidos. En la subsección 4.1.1 se encuentran los resultados de los experimentos basándose con el método propuesto para el reconocimiemto de expresiones, visto en la subseccion 3.1.1, y la subsección 4.1.2 muestra los resultados del experimento del método propuesto para el reconocimiento de micro-expresiones faciales, visto en la subsección 3.1.2.

4.1.1. Resultados del método de reconocimiento de expresiones faciales

Para realizar los experimentos del presente método se empleó la metodología propuesta en la Sección 3.1, donde se usó MatLab para el pre-procesamiento de las imágenes, la librería MatConvNet [49] para la extracción de características y WEKA (*Waikato Environment for Knowledge Analysis*) [50] para la clasificación.

4.1.1.1. Experimento 1: Clasificación Binaria vs Multi-clase

Este experimento tiene el próposito de comparar el desempeño de la clasificación binaria, contra la multi-clase que tiene como objetivo reconocer más de dos clases. Así como, conocer el rendimiento para resolver dichas tareas de los algoritmos de aprendizaje tradicionales.

Para el experimento se utilizaron dos bases de datos con ejemplos de imágenes que contienen diferentes expresiones faciales: enojo, disgusto, miedo, alegría, tristeza, sorpresa y neutral.

- **Cohn Kanade Extendida (CK+) [5]:** cuenta con 593 secuencias de imágenes de 123 sujetos, donde solo 309 secuencias se encuentran etiquetadas. Las secuencias de imágenes empiezan con el rostro neutral y terminan con un tipo de expresión facial. El tamaño de las imágenes son de 640x490 píxeles a escala de grises con un formato de .png. En la Figura 4.1 se muestran algunos ejemplos de las imágenes que contiene esta base de datos.

- **The Japanese Female Facial Expression (JAFFE) [51]:** cuenta con 213 imágenes de 60 mujeres japonesas. Contiene siete expresiones faciales que son enojo, disgusto, miedo, alegría, tristeza, sorpresa y neutral. El tamaño de las imágenes es de 256x256 pixeles a escala de grises y un formato de .tiff.

on la finalidad de utlizar la mayor cantidad de ejemplos para cada clase, se decidió combinar los ejemplos de ambas bases de datos para formar los conjuntos de ejemplos de entrenamiento. En la Tabla 4.1 se muestra la distribución de número de ejemplos de cada base de datos.

Figura 4.1: *Seis expresiones básicas de la BD CK+ [5]*

Los conjuntos de ejemplos para el entrenamiento se generaron de la siguiente manera: en la clasificación binaria se formaron seis conjuntos balanceados para cada expresión, tomando el total de ejemplos de cada expresión que se muestra en la Tabla 4.1 como el conjunto de ejemplos positivos y tomando de forma aleatoria el mismo número de ejemplos de otras expresiones faciales del que se va clasificar como conjunto de ejemplos negativos, y en la clasificación multi-clase se formo un conjunto desbalanceado con el total de ejemplos de expresiones de la Tabla 4.1.

Tabla 4.1: *Distribución de ejemplos CK+ y JAFFE*

Expresión	**CK+**	**JAFFE**	**TOTAL**
Enojo	45	30	75
Disgusto	59	29	88
Miedo	25	32	57
Alegría	69	31	100
Tristeza	28	30	58
Sorpresa	83	30	113
Neutral	123	30	153
Total	432	213	645

En el proceso de clasificación se utilizaron varios algoritmos de aprendizaje con valores por default como: Naïve Bayes (NB), máquinas de soporte vectorial (SVM) con LibLinear (LL) y Optimización de Secuencia Mínima (SMO) utilizando ambas funciones lineales, k-vecinos más cercanos (KNN) configurado a $k = 5$ vecinos, el algoritmo AdaBoost (AB) y árboles de decisión con RandomForest (RF).

Usando la herramienta WEKA, se ejecutó cinco veces cada algoritmo de aprendizaje, donde se modificó la semilla aleatoria en cada ejecución con el fin de poder cambiar la secuencia en la muestra a evaluar por la validación cruzada con 10 pliegues, obteniendo así los promedios de cinco resultados para cada algoritmo de aprendizaje. Las medidas de evaluación que se reportan son exactitud, precisión, recuerdo y medida F.

Por un lado, los resultados para la clasificación binaria son mostrados en las Tablas

4.2, 4.3, 4.4 y 4.5. Los mejores resultados en su mayoría son reportados por el algoritmo LL obteniendo un rendimiento del 97.70 % de exactitud y 0.98 en la medida F, siguiendo SMO con el 96.95 % de exactitud y 0.97 en la medida F, y los resultados más bajos son reportados por el algoritmo KNN con el 73.47 % de exactitud. Con base en los resultados, podemos decir, que al utilizar un vector numérico de 4096 elementos para representar un ejemplo, el mejor tipo de aprendizaje para manejar los datos extraídos es empleando una búsqueda de una función para encontrar vectores lineales y vemos que los datos extraídos no favorecen el emplear la similitud con la distancia Euclidiana para la comparación de los ejemplos.

Tabla 4.2: *Resultados de exactitud de modelos binarios*

Expresiones	**NB**	**LL**	**SMO**	**KNN**	**AB**	**RF**
Enojo	91.73	98.40	98.40	78.67	92.00	94.27
Disgusto	94.66	96.70	96.02	74.89	90.68	92.61
Miedo	88.42	95.44	95.09	70.18	89.82	91.23
Alegría	91.00	97.80	97.30	88.00	93.10	95.10
Tristeza	88.81	98.14	95.76	68.31	89.49	92.25
Sorpresa	92.92	99.73	99.12	60.80	92.92	94.25
Promedio	91.26	**97.70**	96.95	73.47	91.34	93.28

Tabla 4.3: *Resultados de precisión de modelos binarios*

Expresiones	**NB**	**LL**	**SMO**	**KNN**	**AB**	**RF**
Enojo	0.92	0.98	0.98	0.84	0.92	0.95
Disgusto	0.95	0.97	0.96	0.80	0.91	0.94
Miedo	0.89	0.96	0.95	0.74	0.90	0.92
Alegría	0.92	0.98	0.97	0.90	0.88	0.95
Tristeza	0.89	0.98	0.96	0.73	0.92	0.92
Sorpresa	0.93	0.98	0.99	0.71	0.93	0.95
Promedio	0.92	**0.98**	0.97	0.79	0.91	0.94

Tabla 4.4: *Resultados de recuerdo de modelos binarios*

Expresiones	**NB**	**LL**	**SMO**	**KNN**	**AB**	**RF**
Enojo	0.92	0.98	0.98	0.79	0.92	0.94
Disgusto	0.95	0.97	0.96	0.75	0.91	0.93
Miedo	0.88	0.96	0.95	0.70	0.90	0.91
Alegría	0.91	0.98	0.97	0.88	0.93	0.95
Tristeza	0.89	0.98	0.96	0.68	0.89	0.92
Sorpresa	0.93	1.00	0.99	0.65	0.93	0.94
Promedio	0.91	**0.98**	0.97	0.74	0.91	0.93

Por el otro lado, los resultados de la clasificación multi-clase se pueden observar en la Tabla 4.6. El algoritmo SMO obtiene los mejores resultados en exactitud con un promedio del 77.27 % y en precisión con 0.79, siguiendo el algoritmo LL con poca diferencia e iguala sus resultados con 0.77 en recuerdo y medida F. Mientras tanto, el algoritmo AB es el

Tabla 4.5: *Resultados de la medida F de modelos binarios*

Expresiones	**NB**	**LL**	**SMO**	**KNN**	**AB**	**RF**
Enojo	0.92	0.98	0.98	0.78	0.92	0.94
Disgusto	0.95	0.97	0.96	0.74	0.91	0.93
Miedo	0.88	0.96	0.95	0.69	0.90	0.91
Alegría	0.91	0.98	0.97	0.88	0.93	0.95
Tristeza	0.89	0.98	0.96	0.67	0.89	0.92
Sorpresa	0.93	1.00	0.99	0.56	0.93	0.94
Promedio	0.91	**0.98**	0.97	0.72	0.91	0.93

que obtiene el rendimiento más bajo con un promedio del 29.64 %, esto es, por que el algoritmo AB se enfoca en la clasificación binaria, ya que emplea una combinación de clasificadores lineales, fuertes para los ejemplos positivos y débiles para los ejemplos negativos, por lo tanto, no es un método favorable para la clasificación multi-clase.

Tabla 4.6: *Resultados multi-clase*

Medidas	**NB**	**LL**	**SMO**	**KNN**	**AB**	**RF**
Exactitud	51.04	77.05	**77.27**	24.90	29.64	61.21
Precisión	0.31	0.78	**0.79**	0.31	0.30	0.70
Recuerdo	0.51	**0.77**	**0.77**	0.25	0.28	0.61
Medida F	0.52	**0.77**	**0.77**	0.23	0.18	0.62

la Tabla 4.7 se muestra la matriz de confusión de mayor exactitud obtenida de las cinco ejecuciones realizadas, donde podemos apreciar que la mejor expresión clasificada fue sorpresa con un 90 % y las peores fueron enojo con un 64 % y tristeza con el 62 %, esto a causa, de que 19 ejemplos de enojo y 17 ejemplos de tristeza fueron confundidas por una expresión neutral, donde posiblemente algunos ejemplos de tristeza y enojo no muestran una alta intensidad en su expresión facial.

Tabla 4.7: *Matriz de confusión de la clasificación multi-clase*

Neu	**Eno**	**Dis**	**Mie**	**Ale**	**Tri**	**Sor**
122	12	1	2	3	12	1
19	48	3	0	0	4	1
10	1	74	1	0	1	1
4	1	0	41	6	2	3
9	1	1	2	83	3	1
17	3	0	1	1	36	1
7	0	1	1	0	2	102

Recapitulando el experimento, reconocer expresiones faciales es una tarea complicada y con la finalidad de utilizar la mayor cantidad de ejemplos se combinaron las bases de datos CK+ y JAFFE, obteniendo los siguientes resultados: el mayor rendimiento de exactitud se presentan en la clasificación binaria alcanzando un 97.70 %, mientras que la clasificación multi-clase obtiene 77.27 %. Con base a los resultados, podemos decir, que se aprende mejor en reconocer a una sola expresión a la vez, en lugar de aprender

a reconocer a varias expresiones. Mientras tanto, el algoritmo SVM LibLinear muestra ser la mejor opción de los algoritmos evaluados, teniendo la mejor exactitud para todas las expresiones faciales. Por lo tanto, en los siguientes experimentos se empleará este algoritmo para la obtención de modelos binarios.

Posteriormente, en la implementación de los modelos binarios obtenidos en esté experimento, se percató que existía un sobreajuste en ellos, es decir que solo aquellos nuevos ejemplos que tuvieran exactamente el mismo tipo de imagen con el que se entrenaron, era posible su correcta clasificación. Con la finalidad de reducir el sobreajuste de modelos binarios, se decidió realizar un segundo experimento en donde se exploran diferentes estrategias.

4.1.1.2. Experimento 2: Estrategias para reducir sobreajuste

Este experimento tiene el propósito de obtener modelos binarios con la capacidad de generalizar su discriminación disminuyendo su sobreajuste, además de buscar mantener un desempeño cercano a lo obtenido en el experimento 1. Para esto, se propone utilizar un mayor número de ejemplos de entrenamiento para cada clase, y así aumentar la diversidad de imágenes. Además, de explorar la formación de varios conjuntos de entrenamiento, considerando tres diferentes configuraciones:

- **Configuración 1:** conjunto de ejemplos de entrenamiento formado por el 25 % de ejemplos positivos y el 75 % de ejemplos negativos.
- **Configuración 2:** conjunto de ejemplos de entrenamiento formado por el 50 % de ejemplos positivos y el 50 % de ejemplos negativos.
- **Configuración 3:** conjunto de ejemplos de entrenamiento formado por el 75 % de ejemplos positivos y el 25 % de ejemplos negativos.

Para aumentar la diversidad en los ejemplos, se adquirieron tres bases de datos más aparte de las dos que se utilizaron en el experimento 1, contando con cinco bases de datos con ejemplos de imágenes con expresiones faciales y un banco de imágenes de ejemplos extraídas de la Web. A continuación se describen las bases de datos agregadas:

- **MMI Facial Expression Databases (MMI) [52]:** contiene 205 secuencias de imágenes de más de 20 sujetos. Cada secuencia de imágenes inicia y termina con una expresión neutral, donde en el intermedio capta una de las seis expresiones básicas: alegría, enojo, disgusto, tristeza, miedo y sorpresa. Se extrajeron imágenes de cada secuencia, donde se obtuvieron 11,500 imágenes con un formato de .jpg y una resolución de píxeles de 890x550.
- **Multimedia Understanding Group (MUG): [53]** consta de un aproximado de 658 secuencias de imágenes de 86 sujetos que realizan expresiones faciales. Cada secuencia de imágenes inicia y termina con una expresión neutral, donde en el intermendio capta una de las seis expresiones básicas: alegría, enojo, disgusto, tristeza, miedo, sorpresa, neutral y mixto. Se extrajeron imágenes de cada secuencia, donde se consiguió 1432 con un formato .jpg a una resolución de 896x896 píxeles.

- **Indian Spontaneous Expression Database (ISED) [54]:** contiene 428 video clips con expresiones espontáneas obtenidas por 50 sujetos mientras observaban videos emocionales. Extrajeron imágenes obteniendo un total de 425 imágenes de emociones de felicidad, disgusto, tristeza y sorpresa, con un formato .jpg y una resolución de 1920x1028 píxeles.
- **Imágenes Web (WEB):** por medio de consultas en la página de navegación de Google, se obtuvo siete conjuntos de imágenes todas las expresiones faciales básicas. Obteniendo un total de 716 imágenes con una variedad de rostros de personas de diferentes edades, sexo, raza y varios posiciones faciales.

La Tabla 4.8, contiene el número de ejemplos obtenidos de cada base de datos, así también la distribución del número de ejemplos para cada expresión facial. En total se obtuvieron 4173 ejemplos de imágenes para utilizar en la formación de los conjuntos de ejemplos de entrenamiento.

Tabla 4.8: *Distribución de ejemplos de cada base de datos*

Expresión	**CK+**	**JAFFE**	**MUG**	**ISED**	**MMI**	**WEB**	**TOTAL**
Enojo	45	30	242	0	151	104	572
Disgusto	59	29	242	78	164	101	673
Miedo	25	32	183	0	142	99	481
Alegría	69	31	259	226	164	103	852
Tristeza	28	30	202	48	154	103	565
Sorpresa	83	30	244	73	181	106	717
Neutral	123	30	60	0	0	100	313
Totales	432	212	1432	425	956	716	4173

Para formar los nuevos conjuntos de entrenamiento se toma como base el número total de ejemplos de cada expresión, definiendo los tres conjuntos con las configuraciones antes mencionadas. La Tabla 4.9 muestra la formación del número de ejemplos usados para las configuraciones propuestas para cada clase de expresión facial a clasificar, obteniendo 18 diferentes conjuntos de entrenamiento.

Tabla 4.9: *Número de ejemplos utilizados para las configuraciones propuestas*

Expresión	**Configuración 1**		**Configuración 2**		**Configuración 3**	
	Pos. 25 %	Neg. 75 %	Pos. 50 %	Neg. 50 %	Pos. 75 %	Neg. 25 %
Enojo	572	1716	572	572	572	191
Disgusto	673	2019	673	673	673	224
Miedo	481	1443	481	481	481	160
Alegría	852	2556	852	852	852	284
Tristeza	565	1695	565	565	565	188
Sorpresa	717	2151	717	717	717	239

En la clasificación se utilizó el algoritmo de aprendizaje SVM con LibLinear usando valores por default. Para cada conjunto de entrenamiento se ejecutó cinco veces modificando la semilla aleatoria, evaluando con la validación cruzada a 10 pliegues. Se obtuvó cinco diferentes resultados de exactitud, precisión, recuerdo y medida F.

Tabla 4.10: *Resultados de exactitud en el reconocimiento de expresiones faciales*

Expresión	**Configuración 1**		**Configuración 2**		**Configuración 3**	
	Exac.	Error	Exac.	Error	Exac.	Error
Enojo	92.11 %	$\pm 2{,}45^{-3}$	86.42 %	$\pm 3{,}00^{-3}$	85.71 %	$\pm 2{,}34^{-3}$
Disgusto	93.79 %	$\pm 3{,}23^{-3}$	89.97 %	$\pm 4{,}87^{-3}$	91.35 %	$\pm 5{,}38^{-3}$
Miedo	90.74 %	$\pm 2{,}19^{-3}$	84.20 %	$\pm 3{,}45^{-3}$	84.99 %	$\pm 4{,}04^{-3}$
Alegría	96.53 %	$\pm 1{,}37^{-3}$	94.05 %	$\pm 3{,}81^{-3}$	92.74 %	$\pm 4{,}01^{-3}$
Tristeza	91.71 %	$\pm 2{,}91^{-3}$	86.79 %	$\pm 5{,}51^{-3}$	86.21 %	$\pm 6{,}15^{-3}$
Sorpresa	92.22 %	$\pm 2{,}64^{-3}$	89.21 %	$\pm 4{,}79^{-3}$	89.64 %	$\pm 1{,}28^{-3}$
Promedio	**92.85** %	$\pm 2{,}46^{-3}$	88.44 %	$\pm 4{,}24^{-3}$	88.64 %	$\pm 3{,}87^{-3}$

Tabla 4.11: *Resultados de precisión en el reconocimiento de expresiones faciales*

Expresión	**Configuración 1**		**Configuración 2**		**Configuración 3**	
	Exac.	Error	Exac.	Error	Exac.	Error
Enojo	0.922	$\pm 2{,}39^{-3}$	0.857	$\pm 1{,}93^{-3}$	0.852	$\pm 2{,}39^{-3}$
Disgusto	0.938	$\pm 3{,}11^{-3}$	0.900	$\pm 4{,}82^{-3}$	0.912	$\pm 5{,}63^{-3}$
Miedo	0.909	$\pm 1{,}67^{-3}$	0.844	$\pm 3{,}36^{-3}$	0.844	$\pm 4{,}22^{-3}$
Alegría	0.965	$\pm 1{,}14^{-3}$	0.941	$\pm 3{,}56^{-3}$	0.926	$\pm 4{,}15^{-3}$
Tristeza	0.917	$\pm 3{,}05^{-3}$	0.869	$\pm 5{,}45^{-3}$	0.857	$\pm 6{,}91^{-3}$
Sorpresa	0.922	$\pm 1{,}64^{-3}$	0.893	$\pm 4{,}95^{-3}$	0.894	$\pm 1{,}22^{-3}$
Promedio	**0.929**	$\pm 2{,}17^{-3}$	0.884	$\pm 6{,}91^{-3}$	0.881	$\pm 4{,}09^{-3}$

Tabla 4.12: *Resultados de recuerdo en el reconocimiento de expresiones faciales*

Expresión	**Configuración 1**		**Configuración 2**		**Configuración 3**	
	Exac.	Error	Exac.	Error	Exac.	Error
Enojo	0.921	$\pm 2{,}65^{-3}$	0.864	$\pm 2{,}95^{-3}$	0.857	$\pm 2{,}68^{-3}$
Disgusto	0.938	$\pm 3{,}11^{-3}$	0.900	$\pm 4{,}66^{-3}$	0.913	$\pm 5{,}50^{-3}$
Miedo	0.907	$\pm 2{,}41^{-3}$	0.842	$\pm 3{,}32^{-3}$	0.850	$\pm 3{,}67^{-3}$
Alegría	0.966	$\pm 1{,}34^{-3}$	0.941	$\pm 3{,}91^{-3}$	0.927	$\pm 4{,}15^{-3}$
Tristeza	0.917	$\pm 2{,}77^{-3}$	0.868	$\pm 5{,}45^{-3}$	0.862	$\pm 6{,}26^{-3}$
Sorpresa	0.922	$\pm 1{,}64^{-3}$	0.894	$\pm 5{,}43^{-3}$	0.896	$\pm 1{,}22^{-3}$
Promedio	**0.928**	$\pm 2{,}32^{-3}$	0.885	$\pm 4{,}29^{-3}$	0.884	$\pm 3{,}92^{-3}$

Tabla 4.13: *Resultados de medida F en el reconocimiento de expresiones faciales*

Expresión	**Configuración 1**		**Configuración 2**		**Configuración 3**	
	Exac.	Error	Exac.	Error	Exac.	Error
Enojo	0.922	$\pm 2{,}30^{-3}$	0.864	$\pm 2{,}95^{-3}$	0.852	$\pm 2{,}00^{-3}$
Disgusto	0.938	$\pm 3{,}11^{-3}$	0.900	$\pm 4{,}97^{-3}$	0.912	$\pm 5{,}63^{-3}$
Miedo	0.908	$\pm 1{,}92^{-3}$	0.842	$\pm 3{,}32^{-3}$	0.845	$\pm 3{,}63^{-3}$
Alegría	0.966	$\pm 1{,}34^{-3}$	0.941	$\pm 3{,}91^{-3}$	0.926	$\pm 4{,}18^{-3}$
Tristeza	0.917	$\pm 3{,}05^{-3}$	0.868	$\pm 5{,}45^{-3}$	0.857	$\pm 6{,}69^{-3}$
Sorpresa	0.922	$\pm 1{,}87^{-3}$	0.864	$\pm 5{,}43^{-3}$	0.894	$\pm 1{,}52^{-3}$
Promedio	**0.929**	$\pm 2{,}27^{-3}$	0.885	$\pm 4{,}34^{-3}$	0.881	$\pm 3{,}90^{-3}$

La Tabla 4.10 muestra los promedios de exactitud de los cinco resultados obtenidos de cada conjunto de entrenamiento, donde el mejor resultado lo obtiene la configuración 1 con el 92.85 % y el peor resultado la configuración 3 con el 88.44 % y esto se mantiene consistente con los resultados de Tablas 4.11, 4.12 y 4.13 que contiene los promedios de precisión, recuerdo y medida F respectivamente.

Con base a los resultados mostrados, se aprecian resultados consistentes entre las tablas, y se obtuvieron modelos binarios con un desempeño competitivo, donde la mejor configuración está formada del 25 % de ejemplos positivos con el 75 % de ejemplos negativos, el cual también es el conjunto que se forma con el mayor número de ejemplos para su entrenamiento.

Comparando los modelos binarios que se obtuvieron de la configuración 1, se observa que el modelo de *Alegría* es el que mejor se desempeña con un 96.53 % de exactitud, siendo este el conjunto que contiene el mayor número de ejemplos, mientras que el desempeño del modelo de *Miedo* es el peor con un 90.74 % de exactitud, además de que es el conjunto que contiene el menor número de ejemplos para su entrenamiento. Mostrando así, que entre más número de ejemplos de entrenamiento mejora el desempeño de los modelos obtenidos.

A pesar que el rendimiento de los modelos binarios obtenidos en este experimento es un poco más bajo que el del experimento 1, se consiguió eliminar el sobreajuste y generalizar los modelos binarios. La Figura 4.2 muestra un ejemplo de cada expresión y la clasificación de los modelos binarios conseguidos al momento, donde la primera columna representa la clasificación por el experimento 1, la segunda columna representa el experimento 2 de la configuración 1, la tercera columna representa el experimento 2 de la configuración 2 y la cuarta columna representa el experimento 2 de la configuración 3. Con está imagen se puede observar el sobreajuste de los modelos obtenidos del primer experimento, clasificando siempre con expresión de tristeza a los nuevos ejemplos y la generalización los modelos obtenidos del segundo experimento clasificando adecuadamente.

4.1.2. Resultados del método de reconocimiento de micro-expresiones faciales

A continuación se muestran los resultados del experimento realizado mediante el método para el reconocimiento de micro-expresiones empleando la metodología propuesta en la Sección 3.1.2 usando las herramienta de MatLab para la preparación de las imágenes de cada secuencia, la librería MatConvNet en la extracción de características y WEKA para la clasificación.

4.1.2.1. Experimento 3: Reconocimiento de Micro-expresiones Faciales

El próposito de este experimento es evaluar el método propuesto en la Sección 3.1.2 para el reconocimiento de micro-expresiones faciales. Para este experimento se utilizan tres formas diferentes para representar una secuencia de imágenes, que es la media, desviación estándar y fusión temprana.

Figura 4.2: *Ejemplos de la clasificación de los modelos binarios obtenidos del primero y segundo experimento*

Se realizó una revisión en el estado del arte, identificando bases de datos con ejemplos de micro-expresiones faciales. Como resultado se obtuvieron dos bases de datos las cuales se describen a continuación:

- **The Chinese Academy of Sciences Micro-expression II (CASMEII):** cuenta con 297 micro-expresiones de 26 participantes en segmentos de video y secuencia de imágenes seleccionadas. Las micro-expresiones fueron etiquetadas por Unidades de Acción (AUs). Su formato tiene una resolución de 200 fps y el área del rostro aproximadamente de 280x340 pixeles. Contiene cinco clases que son alegría, disgusto, sorpresa, represión y otros [3].

- **Spontaneous Micro-expression Database HS (SMIC-HS):** contiene de 16 sujetos un número de 164 micro-expresiones en total en segmentos de video con una cá-

mara de alta velocidad de 100 pfs. Los segmentos de video se encuentran etiquetas por tres clases que son las micro-expresiones negativas, positivas y sorpresa. Incluye también, un conjunto de 164 segmentos de videos donde no contiene micro-expresiones (non-micro) con el fin de utilizarlos como ejemplos negativos [40].

La Tabla 4.14 contiene el número de ejemplos obtenidos de ámbas bases de datos, así también la distribución del número de ejemplos para cada micro-expresión facial. Cabe mencionar que cada ejemplo es representado por una secuencia de imágenes que contiene una micro-expresión facial. Con la finalidad de utilizar el mayor número de ejemplos y aumentar un poco la diversidad, se combinaron las bases de datos para la formación de los conjuntos de entrenamiento de manera balanceda, con excepción del conjunto de entrenamiento para reconocer una micro-expresión que utiliza 290 ejemplos positivos y 164 ejemplos negativos.

Tabla 4.14: *Distribución de ejemplos de cada base de datos*

Expresión	**CASMEII**	**SMIC-HS**	**TOTAL**
Negativas	66	70	136
Positivas	34	51	85
Sorpresa	26	43	69
Non-Micro	0	164	164

Tabla 4.15: *Resultados de exactitud en el reconocimiento de micro-expresiones faciales*

Expresión	**Media**		**Desv. estándar**		**Fusión temprana**	
Micro	83.08 %	$\pm 1{,}01^{-2}$	84.93 %	$\pm 1{,}03^{-2}$	85.37 %	$\pm 6{,}72^{-3}$
Negativos	73.45 %	$\pm 5{,}45^{-3}$	71.10 %	$\pm 4{,}50^{-3}$	74.62 %	$\pm 1{,}26^{-2}$
Positivos	74.71 %	$\pm 1{,}38^{-2}$	64.71 %	$\pm 2{,}88^{-2}$	75.65 %	$\pm 1{,}15^{-2}$
Sorpresa	80.72 %	$\pm 1{,}10^{-2}$	75.65 %	$\pm 2{,}44^{-2}$	79.57 %	$\pm 2{,}14^{-2}$
Promedio	77.99 %	$\pm 1{,}01^{-2}$	74.10 %	$\pm 1{,}70^{-2}$	**78.80 %**	$\pm 1{,}30^{-2}$

Tabla 4.16: *Resultados de precisión en el reconocimiento de micro-expresiones faciales*

Expresión	**Media**		**Desv. estándar**		**Fusión temprana**	
Micro	0.836	$\pm 9{,}86^{-3}$	0.856	$\pm 9{,}81^{-3}$	0.855	$\pm 4{,}85^{-3}$
Negativos	0.735	$\pm 5{,}73^{-3}$	0.714	$\pm 7{,}57^{-3}$	0.747	$\pm 1{,}25^{-2}$
Positivos	0.747	$\pm 1{,}40^{-2}$	0.651	$\pm 3{,}04^{-2}$	0.757	$\pm 1{,}21^{-2}$
Sorpresa	0.808	$\pm 1{,}12^{-2}$	0.757	$\pm 2{,}45^{-2}$	0.797	$\pm 2{,}12^{-2}$
Promedio	0.782	$\pm 1{,}02^{-2}$	0.744	$\pm 1{,}81^{-2}$	**0.789**	$\pm 1{,}26^{-2}$

Dentro de la clasificación se utilizó el algoritmo SVM LibLinear, donde se ejecutó cinco veces cada conjunto de entrenamiento modificando la semilla aleatoria, evaluando con la validación cruzada con 10 pliegues, para obtener cinco resultados diferentes en exactitud, precisión, recuerdo y medida F.

La Tabla 4.15 contiene los promedios de los resultados obtenidos de exactitud, los cuales indican que el mejor resultado se obtuvó realizando una fusión temprana que

Tabla 4.17: *Resultados de recuerdo en el reconocimiento de micro-expresiones faciales*

Expresión	**Media**		**Desv. estándar**		**Fusión temprana**	
Micro	0.831	$\pm 1{,}01^{-2}$	0.851	$\pm 8{,}71^{-3}$	0.852	$\pm 5{,}24^{-3}$
Negativos	0.734	$\pm 5{,}22^{-3}$	0.711	$\pm 4{,}80^{-3}$	0.746	$\pm 1{,}28^{-2}$
Positivos	0.747	$\pm 1{,}41^{-2}$	0.647	$\pm 2{,}88^{-2}$	0.757	$\pm 1{,}21^{-2}$
Sorpresa	0.807	$\pm 1{,}11^{-2}$	0.757	$\pm 2{,}43^{-2}$	0.796	$\pm 2{,}14^{-2}$
Promedio	0.780	$\pm 1{,}01^{-2}$	0.742	$\pm 1{,}66^{-2}$	**0.788**	$\pm 1{,}29^{-2}$

Tabla 4.18: *Resultados de medida F en el reconocimiento de micro-expresiones faciales*

Expresión	**Media**		**Desv. estándar**		**Fusión temprana**	
Micro	0.833	$\pm 9{,}91^{-3}$	0.851	$\pm 9{,}98^{-3}$	0.853	$\pm 5{,}24^{-3}$
Negativos	0.735	$\pm 5{,}54^{-3}$	0.711	$\pm 4{,}38^{-3}$	0.746	$\pm 1{,}26^{-2}$
Positivos	0.747	$\pm 1{,}41^{-2}$	0.645	$\pm 2{,}87^{-2}$	0.757	$\pm 1{,}17^{-2}$
Sorpresa	0.807	$\pm 1{,}08^{-2}$	0.756	$\pm 2{,}46^{-2}$	0.796	$\pm 2{,}14^{-2}$
Promedio	0.780	$\pm 1{,}01^{-2}$	0.741	$\pm 1{,}69^{-2}$	**0.788**	$\pm 1{,}27^{-2}$

concatena la representación de la secuencia de imágenes de la media y la desviación estándar, donde el modelo obtenido para la detección de micro-expresiones tiene un resultado de exactitud de 85.37 % y el promedio para el reconocimiento de clase de micro-expresiones obtiene una exactitud del 78.80 %, con una precisión de 0.789 (Tabla 4.16), recuerdo de 0.788 (Tabla 4.17) y por los tanto, una medida F de 0.788 (Tabla 4.18).

Comparando el estado del arte que tienen un rendimiento de exactitud del 60 % al 80 %, el resultado obtenido de los modelos de micro-expresiones faciales se considera aceptable, puesto que el rendimiento obtenido se mantiene dentro del rango. Por lo tanto, se puede considerar la implementación CNN para la extracción de características en el reconocimiento de micro-expresiones faciales como una alternativa para resolver esta tarea.

4.2. Resultados de la etapa de contrucción del prototipo

En esta sección se muestran los resultados obtenidos durante la contrucción del prototipo, y se encuentra organizada de la siguiente manera: 4.2.1 se describe la obtención de los modelos aprendidos compatibles con el prototipo; 4.2.2 muestra la estructura del prototipo e integración de los modelos; 4.2.3 muestra el diseño gráfico a utilizar en el prototipo y la explicación de sus elementos; por último en 4.2.4 se muestra la evidencia y funcionamiento del prototipo.

4.2.1. Obtención de los modelos aprendidos

En esta etapa se decidió utilizar la herramienta Matlab 2016b [55], ya que permite pre-procesar las imágenes, integrar las librerías de MatConvnet V1.0 [49] y LibLinear V2.11 para el aprendizaje de modelos, y desarrollar la interfaz gráfica del usuario (GUI).

4.2.2. Estructura del prototipo e integración de los modelos

La estructura del prototipo tiene como entrada un video que será analizado para obtener como resultado las expresiones y micro-expresiones faciales contenidas en el video. En la Figura 4.3 se observa a detalle las partes de la estructura del prototipo, y que a continuación se describe en los siguientes pasos:

1. Selección del video. Se elige el video a analizar y obtener los siguientes datos: nombre del video, ubicación, número de frames, duración, frames por segundo (fps).

2. Selección el segmento. Permite al usuario elegir cierta parte del video a analizar, introduciendo el periodo de números de frames.

3. Obtención de imagen. Dependiendo del segmento seleccionado, se toma secuencialmente las imágenes del segmento.

4. Procesamiento de la imagen. La imagen pasa por la fase del pre-procesamiento de la imagen y extracción de características, obteniendo como salida el vector de características de la imagen.

5. Modelos de expresiones faciales. El vector de características pasa por los modelos binarios de expresiones faciales: enojo, disgusto, miedo, alegría, tristeza y sorpresa, obteniendo como respuesta si hubo alguna expresión facial en la imagen.

6. Representación del segmento. Al obtener los vectores de características de n imágenes ($n = fps/2$), es representada por la fusión temprana conformada por la media y desviación estándar.

7. Modelo de micro-expresión. La representación del segmento pasa por el modelo binario de micro-expresión. Este modelo sirve como filtro, puesto que si la respuesta dada por el modelo es positiva pasa a reconocer el tipo de micro-expresión.

8. Modelos de micro-expresiones. La representación del segmento pasa por los tres tipos de modelos binarios de micro-expresiones que son: negativo, positivo y sorpresa, obteniendo como respuesta si hubo alguno de los tres tipos de micro-expresión.

9. Lista de las expresiones y micro-expresiones encontradas en el video.

4.2.3. Diseño de la interfaz gráfica

De acuerdo a la estructura del prototipo, se obtuvieron las siguientes interfaces gráficas que se muestran en la Figura 4.4 y se describen a continuación.

El diseño de la interfaz principal que se muestra en la Figura 4.4(a), cuenta con los siguientes elementos:

- **Área para la visiualización del video.** La mayor parte de esta área muestra los *frames* contenidos en el video y en la parte inferior contiene una barra de progreso que visualiza el *frame* actual y el porcentaje de avance del análisis.

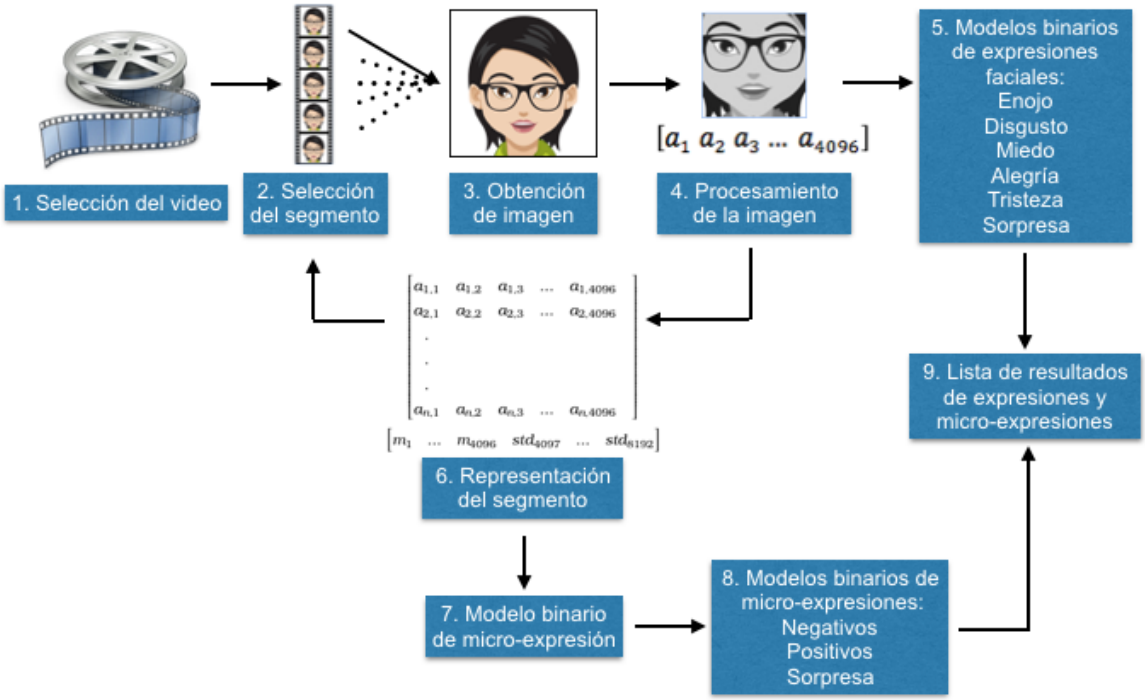


Figura 4.3: *Estructura del prototipo*

- **Área de información del video seleccionado.** Los datos proporcionados del video son: nombre, ruta, duración, número de *frames* y número de *frames* por segundo (fps)

- **Área para la selección del segmento a analizar.** Esta área contiene dos campos de entrada para definir el periodo de *frames* a analizar, introduciendo los números de inicio y final, así también se muestran los botones *"Analyze"* para iniciar el proceso del análisis y *"Results"* para mostrar la lista de las expresiones y micro-expresiones encontradas en el segmento del video elegido.

- **Botón *"Select Video"*.** El botón permite la búsqueda y elección del video a analizar

- **Botón *"Clean Screen"*.** Sirve para inicializar la pantalla para una nueva elección de video.

- **Botón *"Information"*.** Muestra el motivo principal de la funcionalidad del prototipo, autores y contactos para aclarar cualquier duda.

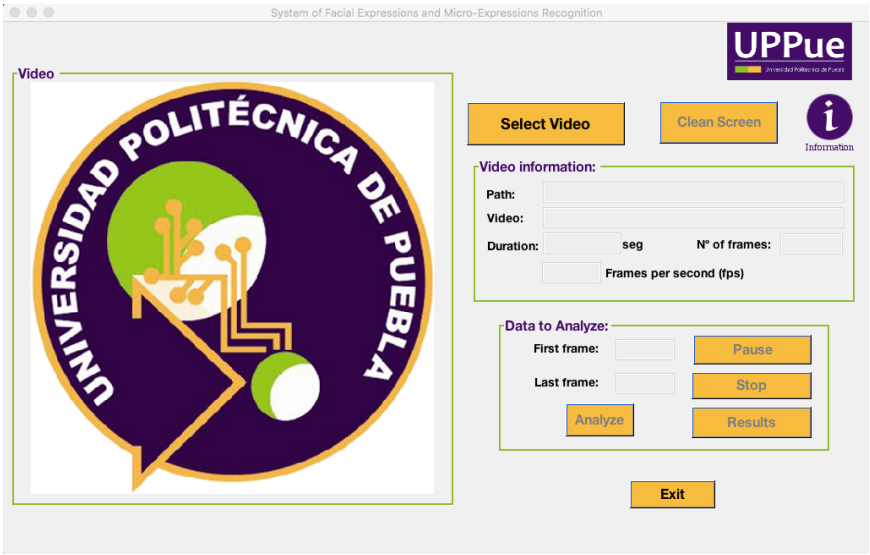
- **Botón *"Exit"*.** Este botón permite salir de la interfaz gráfica.

La ventana de resultados como se observa en la Figura 4.4(b), aparece al momento de presionar el botón *"Results"* que se habilita al terminar el análisis de un segmento. Esta interfaz se divide en tres partes:

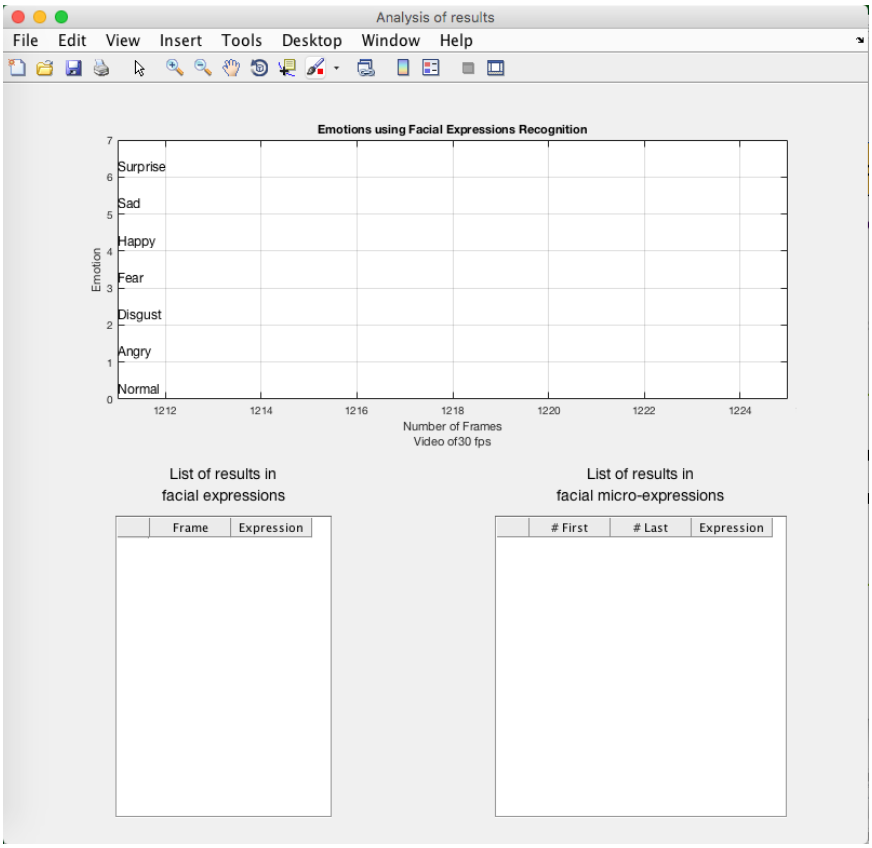
1. La primera parte muestra una gráfica con los resultados de las expresiones faciales encontradas en el segmento analizado, donde el eje *'X'* representa el número de *frames* analizados del segmento y el eje *'Y'* las expresiones.

2. En la segunda parte aparece una tabla que muestra un listado de las expresiones y el *frame* en donde se encontró cada expresión.

3. Por último, en la tercera parte aparece una tabla que contiene un listado de las micro-expresiones faciales y el período de *frames* donde fueron encontradas.



(a) Interfaz principal



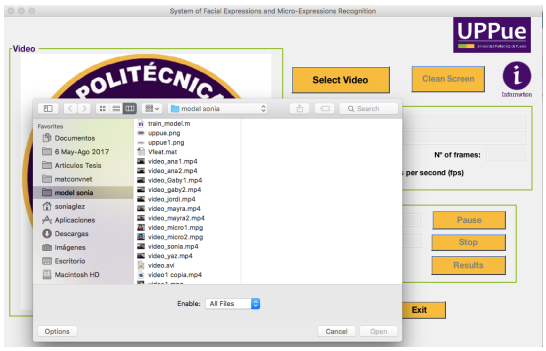
(b) Interfaz de resultados

Figura 4.4: *Interfaces obtenidos para el prototipo*

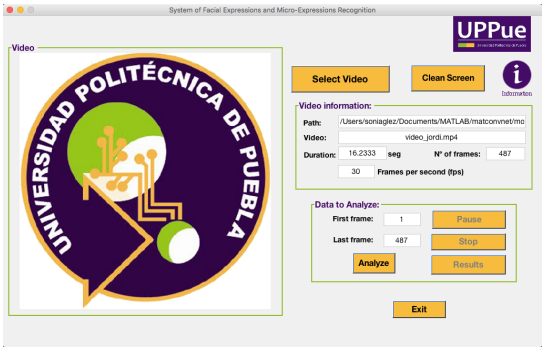
4.2.4. Funcionamiento del prototipo

A continuación, en la figura 4.5 se muestra un ejemplo que a través de imágenes permite mostrar el funcionamiento del prototipo. El ejemplo expuesto comienza seleccionando el video, para ello se presiona el botón *"Select Video"* donde aparece una ventana alterna que permite explorar en la computadora y elegir un video (ver Figura 4.5(a)). Inmediatamente, se muestra la información del video elegido, ruta, nombre, duración, número total de *frames* y el número de *frames* por segundo (fps), así también se completan los campos *First frame* con 1 *frame* y *Last frame* con el último *frames* abarcando todo el video elegido (ver Figura 4.5(b)). Posteriormente, el usuario puede elegir analizar todo el video o seleccionar una parte del video como es el caso del ejemplo, donde se ingresa un período de 221 al 250 *frames* y se presiona el botón *"Analyze"* para inicializar su búsqueda de expresiones y micro-expresiones (ver Figura 4.5(c)). Las Figuras 4.5(d) y 4.5(e) muestran las expresiones encontradas, así también el progreso del segmento en la barra inferior del área del video, y por último al finalizar el análisis del segmento seleccionado se habilita el botón *"Result"* y al presionarse se muestra una ventana alterna con los resultados de expresiones y micro-expresiones encontradas (ver Figura 4.6). En ésta última Figura 4.6 se puede ver que solo fueron encontrados 2 tipos de expresiones que son: sorpresa en los *frames* 221 al 244 y alegría en los *frames* 245 al 250, así también, una micro-expresión positiva durante el segmento del video.

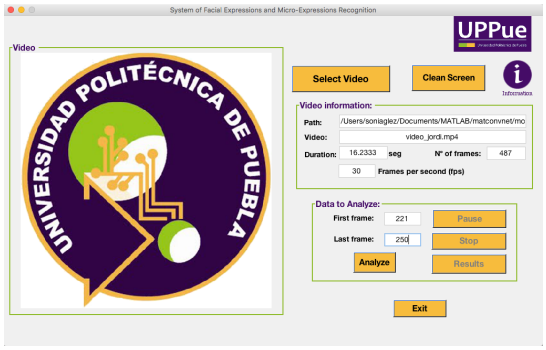
Con motivo de mostrar el funcionamiento de la interfaz y los modelos binarios de las expresiones faciales, se obtuvieron tres nuevos videos de tres personas. En cada video se capturaron las seis expresiones faciales básicas, cabe mencionar que la mayoría no fueron espontáneas, obteniendo las siguientes expresiones como se muestra en la Figura 4.7, Figura 4.8 y Figura 4.9.



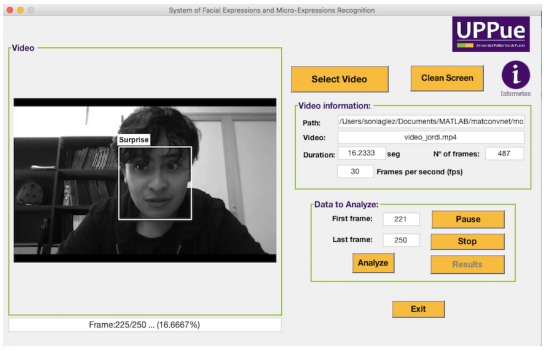
(a) Selección del video



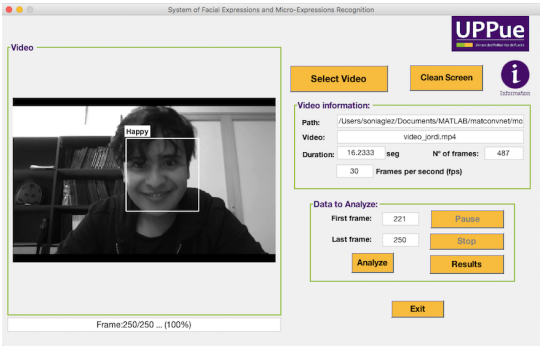
(b) Información del video



(c) Selección del segmento



(d) Inicio del análisis



(e) Final del análisis

Figura 4.5: *Imágenes de ejemplo del funcionamiento del prototipo*

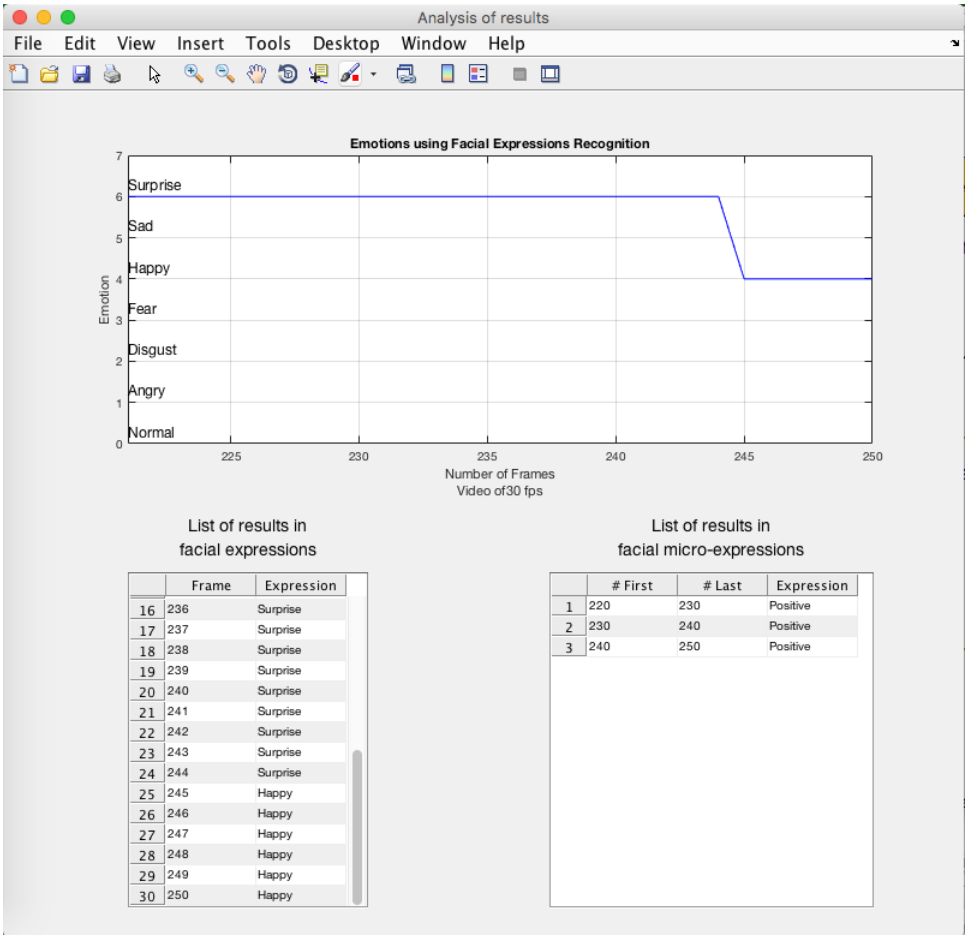


Figura 4.6: *Resultados del segmento*

Figura 4.7: *Expresiones encontradas en el video con la persona 1*

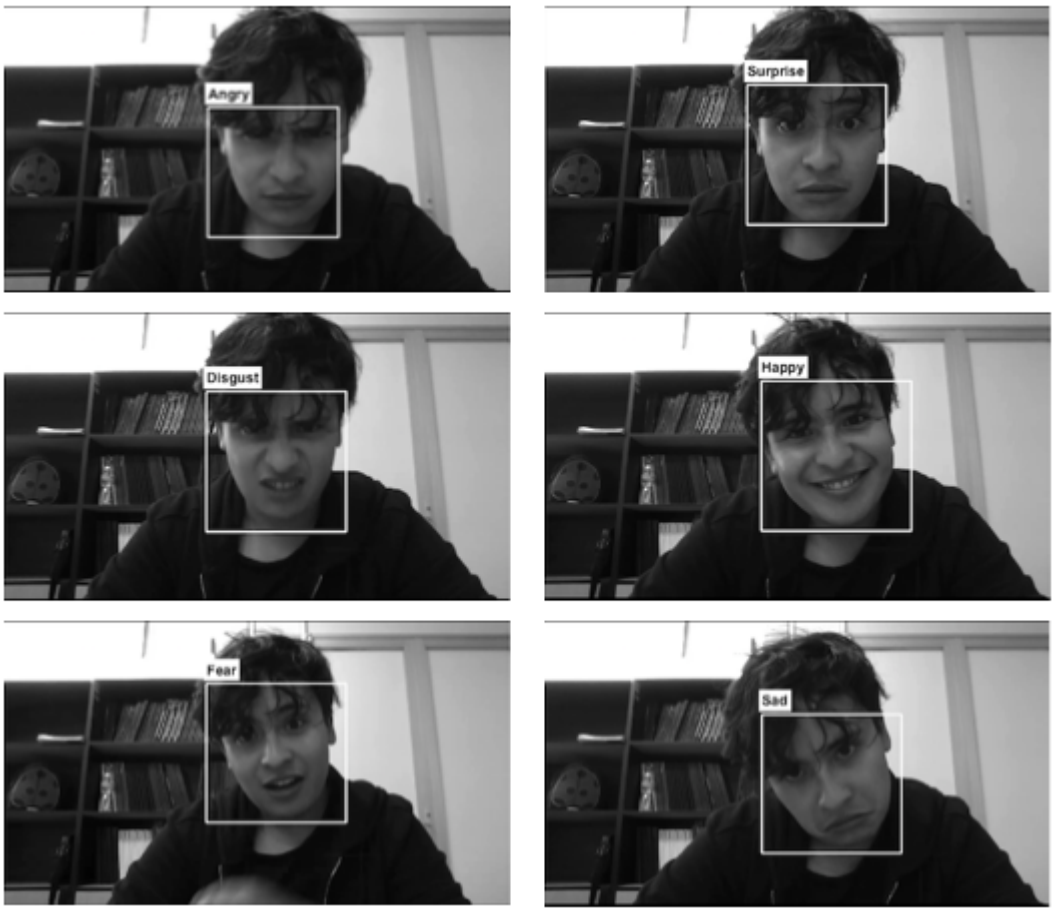


Figura 4.8: *Expresiones encontradas en el video con la persona 2*

Figura 4.9: *Expresiones encontradas en el video con la persona 3*

V

Conclusiones

Capítulo 5: Conclusiones

A continuación, el capítulo describe las principales conclusiones obtenidas del proyecto y se plantea el trabajo a futuro para su mejora.

5.1. Conclusiones

En este proyecto se desarrolló un prototipo capaz de reconocer seis clases de expresiones y tres clases de micro-expresiones faciales mediante el uso del aprendizaje profundo utilizando una CNN como extractor de características.

Por un lado, en la etapa experimental se comprobó que el utilizar una clasificación binaria tiene un mejor desempeño en comparación con una clasificación multi-clase. Así tambien, se realizó una comparación de varios algoritmos de aprendizaje, donde las máquinas de soporte vectorial mostraron el mejor desempeño para predecir la clasificación de nuevos ejemplos. Además, se observó que el aumentar el número de ejemplos en expresiones faciales, reduce el sobreajuste de los modelos prediciendo mejor nuevos ejemplos, ya que incrementa la diversidad del conjunto de entrenamiento. Por lo tanto, se obtuvieron modelos aprendidos para el reconocimiento de expresiones con un resultado promedio de exactitud del 92 %. Mientras que en el reconocimiento de micro-expresiones, tras la comparación de varios métodos, se obtuvó que al utilizar el promedio junto con desviación estándar para representar las caracteristicas extraídas de una secuencia de imágenes, se obtiene un mejor reconocimiento, obteniendo modelos aprendidos con un resultado promedio de exactitud del 77 %, resultado de exactitud comparable con los obtenidos por métodos del estado de arte.

Por otro lado, en la etapa de construcción del prototipo, usando la herramienta MatLab para el desarrollo de la interfaz, se consiguió un prototipo eficaz y fácil de utilizar.

De esta manera, se puede concluir que mediante la metodología propuesta en el Capítulo 3 y los resultados mostrados en el Capítulo 4, se cubren los objetivos planteados en este proyecto de tesis.

5.2. Trabajo futuro

Mediante los resultados y discusiones presentados en el capítulo anterior, el trabajo a futuro se enfoca en las siguientes direcciones:

- En primer lugar los resultados de exactitud de los modelos para el recocimiento de micro-expresiones pueden ser mejorados explorando métodos que permitan combinar los modelos obtenidos con el sistema de codificación facial (FACS *Facial Action Coding System*).

- Construir una nueva base de datos con ejemplos de micro-expresiones faciales, beneficiaría el aumentar la diversidad en los ejemplos y mejorar el desempeño en su reconocimiento.

- Actualmente, el prototipo no es portable ni distribuible, lo cual son aspectos deseables para una aplicación de este tipo. Se podría solucionar esto, importando su código mediante el uso de herramientas y lenguajes de programación libres, así como framework especializados para aplicaciones móviles.
- El prototipo actualmente trabaja en la detección de un solo rostro. Una manera de ampliar sus capacidades es la ubicación de n número de rostros en un video.
- Trabajar la usabilidad del prototipo mediante un caso de uso, es decir, en donde un especialista evalúe su funcionalidad.
- Finalmente, el adoptar enfoques de procesamiento en paralelo, ampliaría la eficiencia de análisis en la detección de rostro y reconocimiento de expresiones y micro-expresiones faciales por parte del prototipo.

Bibliografía

[1] L. E. Sucar and G. Gómez, "Visión computacional."

[2] K. Grauman and B. Leibe, "Visual object recognition," *Synthesis lectures on artificial intelligence and machine learning*, vol. 5, no. 2, pp. 1–181, 2011.

[3] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, "Casme ii: An improved spontaneous micro-expression database and the baseline evaluation," *PloS one*, vol. 9, no. 1, p. e86041, 2014.

[4] E. Alpaydin, *Introduction to Machine Learning*, 2nd ed. The MIT Press, 2010.

[5] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.

[6] K. P. Murphy, *Machine learning: a probabilistic perspective*, 2012.

[7] F. Gravenhorst, A. Muaremi, J. Bardram, A. Grünerbl, O. Mayora, G. Wurzer, M. Frost, V. Osmani, B. Arnrich, P. Lukowicz *et al.*, "Mobile phones as medical devices in mental disorder treatment: an overview," *Personal and Ubiquitous Computing*, vol. 19, no. 2, pp. 335–353, 2015.

[8] G. R. H.-G. C. Campuzano RJC, Hernández-Álvarez A, "Tendencia de la mortalidad por trastornos mentales: un panorama epidemiológico en méxico (1980-2011)," vol. 57, no. 3, pp. 22–30, 2014.

[9] C. Heredia, G. Santaella, and L. Somarriba, "Informe psicológico," Textos de apoyo didáctico, Facultad de Psicología, UNAM, 2012.

[10] P. Ekman, "Lie catching and microexpressions," *The philosophy of deception*, pp. 118–133, 2009.

[11] S. J. Prince, *Computer vision: models, learning, and inference*. Cambridge University Press, 2012.

[12] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[13] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Synthetic data and artificial neural networks for natural scene text recognition," *CoRR*, vol. abs/1406.2227, 2014.

[14] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition." in *BMVC*, vol. 1, no. 3, 2015, p. 6.

[15] B. Huval, T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, P. Rajpurkar, T. Migimatsu, R. Cheng-Yue, F. Mujica, A. Coates, and A. Y. Ng, "An empirical evaluation of deep learning on highway driving," *CoRR*, vol. abs/1504.01716, 2015.

[16] L. Wang and D. Sng, "Deep learning algorithms with applications to video analytics for A smart city: A survey," *CoRR*, vol. abs/1512.03131, 2015.

[17] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., 2006.

[18] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *European conference on computational learning theory*. Springer, 1995, pp. 23–37.

[19] P. Loncomilla. (2016, Nov.) Deep learning: Redes convolucionales. [Online]. Available: https://ccc.inaoep.mx/~pgomez/deep/presentations/2016Loncomilla.pdf

[20] J. J. L.-I. Aliño, M. V. Miyar, A. P. Association *et al.*, *DSM-IV-TR: Manual diagnóstico y estadístico de los trastornos mentales*. American Psychiatric Pub, 2008.

[21] P. Ekman and W. Friesen, "Nonverbal leakage and clues to deception," *Psychiatry*, vol. 32, no. 1, pp. 88–106, 1969.

[22] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 21, no. 10, pp. 974–989, 1999.

[23] C. Cognitiva, "La expresión facial de las emociones: Historia y aplicaciones."

[24] C. Darwin, *La expresión de las emociones en los animales y en el hombre*. Alianza editorial, 1998.

[25] P. Ekman and W. Friesen, "Facial action coding system: A technique for the measurement of facial movement," *Psychiatry*, 1978.

[26] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: A survey," *Pattern recognition*, vol. 25, no. 1, pp. 65–77, 1992.

[27] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[28] H.-B. Deng, L.-W. Jin, L.-X. Zhen, J.-C. Huang *et al.*, "A new facial expression recognition method based on local gabor filter bank and pca plus lda," *International Journal of Information Technology*, vol. 11, no. 11, pp. 86–96, 2005.

[29] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*. IEEE, 2016, pp. 1–10.

[30] D. Matsumoto and H. S. Hwang, "Evidence for training the ability to read microexpressions of emotion," *Motivation and Emotion*, vol. 35, no. 2, pp. 181–191, 2011.

[31] E. A. Haggard and K. S. Isaacs, "Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy," in *Methods of research in psychotherapy*. Springer, 1966, pp. 154–165.

[32] M. G. Frank and P. Ekman, "The ability to detect deceit generalizes across different types of high-stake lies." *Journal of personality and social psychology*, vol. 72, no. 6, p. 1429, 1997.

[33] P. Ekman, "Microexpression training tool (mett)," 2002.

[34] S. Porter and L. Ten Brinke, "Reading between the lies identifying concealed and falsified emotions in universal facial expressions," *Psychological Science*, vol. 19, no. 5, pp. 508–514, 2008.

[35] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE transactions on image processing*, vol. 16, no. 1, pp. 172–187, 2007.

[36] M. Shreve, S. Godavarthy, V. Manohar, D. Goldgof, and S. Sarkar, "Towards macro- and micro-expression spotting in video using strain patterns," in *Applications of Computer Vision (WACV), 2009 Workshop on*. IEEE, 2009, pp. 1–6.

[37] S. Polikovsky, Y. Kameda, and Y. Ohta, "Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor," in *Crime Detection and Prevention (ICDP 2009), 3rd International Conference on*. IET, 2009, pp. 1–6.

[38] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, "Recognising spontaneous facial micro-expressions," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1449–1456.

[39] W.-J. Yan, Q. Wu, Y.-J. Liu, S.-J. Wang, and X. Fu, "Casme database: A dataset of spontaneous micro-expressions collected from neutralized faces," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–7.

[40] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen, "A spontaneous micro-expression database: Inducement, collection and baseline," in *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–6.

[41] J. Kumari, R. Rajesh, and K. Pooja, "Facial expression recognition: A survey," *Procedia Computer Science*, vol. 58, pp. 486–491, 2015.

[42] F. De la Torre, W.-S. Chu, X. Xiong, F. Vicente, X. Ding, and J. F. Cohn, "Intraface," in *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2015.

[43] X. Peng, Z. Xia, L. Li, and X. Feng, "Towards facial expression recognition in the wild: A new database and deep recognition system," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2016 IEEE Conference on*. IEEE, 2016, pp. 1544–1550.

[44] A. Mollahosseini, B. Hasani, M. J. Salvador, H. Abdollahi, D. Chan, and M. H. Mahoor, "Facial expression recognition from world wild web," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 58–65.

[45] W.-J. Yan, S.-J. Wang, Y.-J. Liu, Q. Wu, and X. Fu, "For micro-expression recognition: Database and suggestions," *Neurocomputing*, vol. 136, pp. 82–87, 2014.

[46] S.-J. Wang, H.-L. Chen, W.-J. Yan, Y.-H. Chen, and X. Fu, "Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine," *Neural processing letters*, vol. 39, no. 1, pp. 25–43, 2014.

[47] X. Li, X. Hong, A. Moilanen, X. Huang, T. Pfister, G. Zhao, and M. Pietikäinen, "Reading hidden emotions: spontaneous micro-expression spotting and recognition," *arXiv preprint arXiv:1511.00423*, 2015.

[48] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 529–534.

[49] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 2015, pp. 689–692.

[50] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[51] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*. IEEE, 1998, pp. 200–205.

[52] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005, pp. 5–pp.

[53] N. Aifanti, C. Papachristou, and A. Delopoulos, "The mug facial expression database," in *Image Analysis for Multimedia Interactive Services (WIAMIS), 2010 11th International Workshop on*. IEEE, 2010, pp. 1–4.

[54] P. PatnaikA. (2016) Indian spontaneous expression database (ised). [Online]. Available: https://sites.google.com/site/iseddatabase/download

[55] C. Thompson and L. Shure, *Image Processing Toolbox: For Use with MATLAB;[user's Guide]*. MathWorks, 1995.