



**UAEM** | Universidad Autónoma  
del Estado de México

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE  
MÉXICO

INGENIERÍA EN COMPUTACIÓN

---

**Sistema de Reconocimiento de emociones, expresiones  
faciales y oculares.**

---

Tesis que presenta

**Laura Yadira DOMÍNGUEZ JALILI**

Para obtener el Grado de

**Doctor en Ciencias de la Computación**

Asesor de Tesis:

**Dr. Jair CERVANTES**

Texcoco, Estado de México.

Octubre del 2022





**UAEM** | Universidad Autónoma  
del Estado de México

**AUTONOMOUS UNIVERSITY OF MEXICO STATE**

**COMPUTER SCIENCE DEPARTMENT**

---

**Recognition system of emotions, facial and ocular  
expressions.**

---

Submitted by

**Laura Yadira DOMÍNGUEZ JALILI**

As a fulfillment of the requirement for the degree of

**Doctorate in Computer Science**

Thesis Advisor

**Ph.D. Jair Cervantes**

Texcoco City, Mexico state.

October 2022



# Índice general

<b>Índice general</b>	<b>1</b>
<b>Índice de figuras</b>	<b>5</b>
<b>Índice de tablas</b>	<b>7</b>
<b>1. Introducción</b>	<b>9</b>
1.1. Planteamiento del problema . . . . .	11
1.2. Justificación . . . . .	12
1.3. Objetivos y alcances del proyecto . . . . .	13
1.4. Hipótesis . . . . .	14
1.5. Estado del Arte . . . . .	14
1.6. Metodología . . . . .	20
<b>2. Marco Teórico</b>	<b>23</b>
2.1. Redes Neuronales . . . . .	23
2.1.1. Redes Neuronales Recurrentes . . . . .	24
2.1.2. Red Neuronal Convolutiva . . . . .	27
2.1.3. Máquinas de Soporte Vectorial (SVM) . . . . .	29
Fase de Entrenamiento . . . . .	29
Fase de prueba . . . . .	30
2.2. Sensores . . . . .	31
2.2.1. Sensores Movimiento . . . . .	32
2.2.2. Sensores de entorno . . . . .	32
2.3. Imágenes . . . . .	33
2.3.1. Imágenes de las señales de un EEG . . . . .	33
2.3.2. Binarias . . . . .	34
2.3.3. Imágenes RGB . . . . .	34
2.4. Preprocesamiento de imágenes . . . . .	35
2.4.1. Operaciones grupales . . . . .	35

2.4.2.	Técnicas de Filtrado . . . . .	35
	Filtros pasa bajo . . . . .	36
	Filtro Moda . . . . .	36
	Filtro de la mediana . . . . .	36
	Filtro adaptativo . . . . .	37
2.4.3.	Filtros Pasa Altos . . . . .	37
	Transformaciones de Intensidad . . . . .	37
2.4.4.	Histogramas . . . . .	38
	Expansión del histograma . . . . .	38
	Contracción del histograma . . . . .	39
	Desplazamiento del histograma . . . . .	39
	Ecuilización de Histograma . . . . .	39
	Normalización de Histograma . . . . .	39
	Contraste . . . . .	40
	Brillo . . . . .	40
2.5.	Segmentación y detección de bordes . . . . .	41
2.5.1.	Operador de Sobel . . . . .	41
2.5.2.	Operador de Prewitt . . . . .	42
2.5.3.	Operador de Roberts . . . . .	43
2.5.4.	Método de Otsu . . . . .	44
2.6.	Técnicas de Extracción y Reducción de Características . . . . .	45
2.6.1.	Local Binary Patterns (LBP) . . . . .	45
2.6.2.	Histogramas de gradientes orientados (HOG) . . . . .	46
2.6.3.	Análisis de componentes principales (PCA) . . . . .	48
	Obtención de los CP (Componentes Principales) . . . . .	49
	Eigenfaces . . . . .	50
2.6.4.	Análisis discriminante lineal (LDA) . . . . .	50
2.6.5.	Algoritmos Genéticos . . . . .	51
2.6.6.	Características de Haralick . . . . .	53
2.7.	Técnicas de detección del rostro, microexpresiones y macro expresiones . . . . .	56
2.7.1.	Viola Jones . . . . .	56
	Características de Haar . . . . .	57
	Imagen Integral . . . . .	58
	Clasificador AdaBoost . . . . .	59
2.7.2.	Detector de Objetos SIFT . . . . .	60

---

2.7.3. Detector de Objetos SURF . . . . .	61
2.7.4. Análisis discriminantes lineal (LDA) . . . . .	62
<b>3. Metodología</b>	<b>65</b>
3.0.1. Conjuntos de Datos . . . . .	66
SMIC database . . . . .	66
SAMM database . . . . .	66
3.1. Metodología I . . . . .	67
3.1.1. Selección de zonas de interés . . . . .	68
3.1.2. Extracción de características . . . . .	70
3.1.3. Clasificación . . . . .	71
3.1.4. Validación . . . . .	71
3.2. Metodología II . . . . .	71
3.2.1. Selección de las características con Algoritmo Genético . . . . .	71
3.3. Metodología III . . . . .	73
3.3.1. Selección de zonas de interés . . . . .	74
3.3.2. Extracción de características . . . . .	74
<b>4. Resultados Experimentales</b>	<b>77</b>
4.0.1. Resultados Metodología I . . . . .	77
4.0.2. Resultados Metodología II . . . . .	80
4.0.3. Resultados Metodología III . . . . .	85
<b>5. Conclusiones</b>	<b>87</b>
5.1. Metodología I . . . . .	87
5.2. Metodología II . . . . .	87
5.3. Metodología III . . . . .	88
<b>6. Artículos publicados</b>	<b>91</b>
<b>Bibliografía</b>	<b>95</b>



# Índice de figuras

2.1. Redes Neuronales (RNAs) . . . . .	24
2.2. Conexión de una Red Neuronal Recurrente (RNN) . . . . .	25
2.3. Conexión de una Red Neuronal Recurrente (RNN) Simple con los intervalos de tiempo (timestep). . . . .	26
2.4. Arquitectura de Elman (RNN Elman) . . . . .	26
2.5. Hiperplano Máquinas de Soporte Vectorial (SVM) . . . . .	30
2.6. Filtro Moda . . . . .	36
2.7. Filtro Mediana . . . . .	37
2.8. Posiciones para la Segmentación de Roberts . . . . .	43
2.9. Diagrama general de un algoritmo genético . . . . .	51
2.10. Métodos de cruza de un algoritmo genético . . . . .	53
2.11. Filtros de Haar . . . . .	57
2.12. Imagen Integral . . . . .	58
2.13. Imagen Integral SURF . . . . .	62
3.1. Metodología I. Expresiones faciales (macro expresiones) . . . . .	67
3.2. Extracción de Características . . . . .	70
3.3. Metodología II. Identificación de emociones y reducción de características (Algoritmo Genético) . . . . .	72
3.4. Metodología III. Zonas discriminativas del rostro para identificación de emociones . . . . .	73
4.1. Matrices de confusión SMIC . . . . .	79
4.2. Matrices de confusión SAMM . . . . .	79
6.1. Artículo 1. Emotion recognition by eyes region using textural features, lbp and hog . . . . .	92
6.2. Artículo 2. Emotion Recognition from Facial Expressions Using a Genetic Algorithm to Feature Extraction . . . . .	93



# Índice de tablas

2.1. Tipos de ecualización . . . . .	40
3.1. Tabla de errores. Prueba del Modelo entrenado . . . . .	69
4.1. Resultados del clasificador SVM para conjunto de datos SAMM . . . . .	77
4.2. Resultados del clasificador SVM para conjunto de datos SMIC . . . . .	78
4.3. Precisión de clasificadores para cada conjunto . . . . .	80
4.4. Vector de Características . . . . .	81
4.5. Total de características SAMM . . . . .	82
4.6. Total de características SMIC . . . . .	83
4.7. SAMM - Algoritmo Genético aplicado . . . . .	83
4.8. SAMM - Desempeño sin reducción de características . . . . .	83
4.9. SMIC - Algoritmo Genético aplicado . . . . .	84
4.10. SMIC - Desempeño sin reducción de características . . . . .	84
4.11. SAMM - Desempeño de los clasificadores . . . . .	85
4.12. SMIC - Desempeño de los clasificadores . . . . .	86



# Capítulo 1

## Introducción

En la actualidad el uso del reconocimiento de emociones en sistemas de visión se ha incrementado debido a su utilidad en diversas áreas. Las emociones son reflejadas en el habla, expresiones faciales, dilatación de la pupila, movimientos oculares y movimientos corporales. Estas características permiten a los sistemas identificar y clasificar las emociones verbales y no verbales del ser humano. Las emociones son estados complejos de los sentimientos, basados en experiencias constantes de forma consciente, que abastecen de cambios físicos y psicológicos al ser humano y esto influye en el comportamiento corporal y ocular al transmitir emociones.

Conocer las emociones de una persona posibilita saber que ocurre en su estado emocional, que necesita o que posible conducta puede presentar por lo que tener un sistema que identifique las emociones podría ser utilizado para enriquecer nuevas investigaciones. Sin embargo, este no es un reto fácil debido a múltiples problemas, como: cambios en la iluminación, oclusión, pérdida de información por baja resolución, etc. Todo esto en conjunto ocasiona pérdida de información y se ve reflejado en un sistema deficiente.

Aunado a lo anterior, al detectar emociones influyen aspectos como: la personalidad, el temperamento, el estado de ánimo, la motivación, la disposición, los sonidos en el entorno y otros [80]. Estos aspectos pueden crear confusión para identificar las microexpresiones y determinar emociones e identificar un posible comportamiento. Por ejemplo, una persona con una personalidad fría no representa de forma notoria sus emociones.

El estudio de las expresiones faciales para la identificación de emociones ha sido un tema de investigación de tiempo atrás.

En 1962 el psicólogo estadounidense Silvan Tomkins propuso que la retroalimentación sensorial llevada a cabo por los músculos del rostro, y la representación de las sensaciones en la piel, pueden generar una experiencia o estado emocional sin necesidad de un proceso cognitivo [26].

En 1988, los psicólogos Fritz Strack, Leonard L. Martin y Sabine Stepper realizaron un estudio llamado el paradigma del bolígrafo sostenido. En este estudio explican la primera teoría de las emociones creada por Silvan Tomkins. En esta, se presenta la hipótesis de que la actividad facial influye en las respuestas afectivas. Los autores realizaron un estudio, en donde las personas debían observar una serie de caricaturas divertidas, mientras que la mitad de las personas debían sostener un bolígrafo con sus labios y los demás debían sostener el bolígrafo con los dientes. De manera que la postura facial que se realiza al tener entre los dientes el bolígrafo contraía el músculo cigomático mayor, que usamos para sonreír. Lo cual facilitaría la expresión sonriente, por el contrario, el bolígrafo en los labios contrae el músculo orbicular lo que inhibe el movimiento necesario para sonreír. Con esta actividad se medía la actividad facial para la sonrisa y la implicancia de la hipótesis del psicólogo Silvan Tomkins.

El resultado fue que las personas que sostuvieron el bolígrafo con los dientes reportaron que las caricaturas eran más divertidas que aquellas personas que sostuvieron el bolígrafo con los labios. Por lo tanto, las expresiones faciales asociadas con alguna emoción efectivamente pueden transformar la experiencia subjetiva de dicha emoción, incluso cuando las personas no están totalmente conscientes de los gestos faciales que está llevando a cabo [70].

En 2016, el psicólogo Wagenmakers replicó el experimento del bolígrafo sostenido. En sus resultados obtenidos considera que no encontraron evidencia suficiente que sostuviera el efecto de la retroalimentación facial. En respuesta Fritz Strack explicó que el experimento de Wagenmakers se había realizado contemplando una variable que no estuvo presente en el estudio original. Indicando que había afectado y determinado nuevos resultados, dicha variable es el conocimiento de saber que está siendo observado, al no ocultar la cámara de video que grababa la actividad de los participantes. De acuerdo con la investigación del psicólogo Fritz la experiencia de sentirse observado causa significativas modificaciones en el efecto de la retroalimentación facial.

Lo que nos indica que el reconocimiento de emociones tendrá mejores resultados si las personas desconocen que están siendo observadas, esto es bueno para sistemas de reconocimiento en tiempo real donde la persona no identifique que el uso de la cámara es directamente para analizar sus emociones o sus posibles acciones.

Los sistemas de reconocimiento de emociones están actualmente presentes en varias áreas, algunos ejemplos de desarrollo son: En psicología se han realizado trabajos de investigación que determinan aspectos nuevos que interfieren para la identificación de emociones; en [66] los autores realizan un estudio que concluye que características

como la raza y la edad pueden influir en el reconocimiento de emociones. Otro estudio como el realizado en [41] muestra como interfieren las emociones en la toma de decisiones, en donde se concluye que nuestros estados de ánimo influyen en nuestros gustos. En [48] los autores desarrollan un sistema inteligente para reconocimiento de emociones, para monitorear los cambios emocionales de una persona a partir de señales de electroencefalografía (EEG). En [65] los autores realizan el análisis del cuerpo humano para la identificación de estados emocionales a partir de patrones de movimiento de cuerpo completo para personas sin habla.

También se tienen investigaciones que hacen uso de sensores para la clasificación de emociones, como en [42] los autores extraen la información combinando la información de los sensores en el cuerpo, en el ambiente y de ubicación, esta investigación permite la identificación de emociones con una mayor portabilidad, pero es una técnica invasiva. El audio también es un influyente en las emociones, al ser una estimulación visible a partir de señales EEG, y se identifica el impacto de la comprensión del lenguaje en las emociones [36]. Y desde luego se tienen investigaciones que hacen uso de imágenes para el reconocimiento de emociones donde se destacan las emociones negativas y positivas para posibles suicidios. Ya que al realizar un análisis de los estados de ánimo de la persona, se puede obtener el patrón de comportamiento ya que el suicidio es una alteración de los estados de ánimo con tiempo atrás [31].

En esta Tesis, se desarrolla un sistema de reconocimiento de emociones basado en las diferentes zonas del rostro, analizando las emociones básicas las cuales son la ira, el asco, la felicidad, la tristeza, la sorpresa y el miedo.

## 1.1. Planteamiento del problema

En la actualidad las investigaciones para la detección de las emociones se han incrementado y se han expandido los sistemas de reconocimiento de emociones, teniendo gran avance en este tema. Sin embargo, debido a las restricciones generadas por la pandemia SARS-COV2, se ha forzado a la población a utilizar una mascarilla de forma constante. Esta situación refleja un problema en los sistemas de visión dando como resultado un problema para los sistemas de reconocimiento de emociones, ya que utilizan el análisis del rostro completo; el uso de una mascarilla resulta un inconveniente enorme para estos sistemas porque su precisión cae significativamente debido a la falta de información.

Por lo que el utilizar el rostro completo para la identificación de una emoción hoy en día es inconcebible debido a los cambios realizados por la nueva normalidad derivados de la pandemia del SARS-COV2, por tal motivo los sistemas actuales necesitan hacer frente al reto que representa tener solo visible una parte del rostro. Aunado a esto desde antes, el realizar la identificación de una emoción era todo un reto al implementarse en tiempo real, algunos de los problemas que se presentan son la oclusión, movimientos constantes y también la distancia que hace perder precisión de las microexpresiones, en algunos casos el maquillaje y accesorios que solo permiten observar algunas partes del rostro. Otros aspectos que afectan son las variaciones de iluminación o poca iluminación, que no permite apreciar con precisión las microexpresiones faciales de la persona, ni los movimientos oculares. Debido a estos aspectos los nuevos sistemas deben ser capaces de identificar las emociones a partir de solo unas partes del rostro, para la mejora continua de la precisión y eficiencia en la detección de las emociones en tiempo real.

## 1.2. Justificación

En la actualidad los sistemas de reconocimiento de emociones presentan algunos retos como se ha mencionado en el planteamiento del problema, además en la mayoría de los sistemas se considera trabajar con el rostro completo para la identificación de las emociones, pero eso conlleva a más costo computacional y tiempo que si solo se tomara una parte. Los sistemas de reconocimiento de emociones varían su precisión por tres factores críticos como son el preprocesamiento, la extracción de características y el diseño de un clasificador, por lo que se propone la aplicación de métodos combinados, analizando las diferentes áreas del rostro para identificar las zonas de mayor precisión, además de trabajar con características de microexpresiones faciales y características de expresiones oculares para la mejora continua de la precisión.

En esta tesis se aborda un análisis completo de diferentes áreas del rostro para determinar su poder discriminativo en la identificación de una emoción, por lo que los resultados propuestos en esta tesis permitirán identificar las emociones de la persona sin tener información de la totalidad del rostro y obteniendo solo información de algunas partes del rostro. No solo eso, sino también un análisis exhaustivo de las diferencias entre los distintos métodos permitirá obtener una idea clara de como influyen las áreas del rostro en la identificación de emociones

## 1.3. **Objetivos y alcances del proyecto**

### Objetivo General

El desarrollo de un sistema para el reconocimiento e identificación de las emociones básicas en tiempo real, realizando un análisis de las distintas áreas del rostro con cada emoción, con el objetivo de identificar la zona del rostro más característico para la expresión de las emociones, además de trabajar con movimientos oculares y microexpresiones para la obtención de precisiones óptimas en la identificación de las emociones.

### Objetivos Específicos

1. Realizar el análisis del problema para identificar los retos y alcances para el desarrollo del sistema.
2. Identificar los movimientos de las expresiones faciales para cada emoción.
3. Identificar los movimientos oculares presentes para cada emoción.
4. Recolección de las imágenes o búsqueda de dataset para el proyecto.
5. Definir una metodología principal para el desarrollo del proyecto
6. Investigar y determinar las técnicas necesarias para identificar las emociones en tiempo real
7. Identificar y aplicar técnicas para separación de frames en video para el análisis en tiempo real.
8. Investigar las posibles técnicas para la extracción de características y definir las que se utilizaran en el proyecto.
9. Investigar y seleccionar las técnicas de clasificación
10. Identificar los cambios de iluminación que crean ruido en nuestra clasificación y como evitar ese problema.
11. Identificar la tasa de error de cada clasificador para determinar el algoritmo de clasificación óptimo en nuestro sistema.
12. Realizar la validación de nuestros algoritmos
13. Realizar las pruebas finales del sistema

## 1.4. Hipótesis

Se puede mejorar la eficiencia de los sistemas de reconocimiento de emociones en tiempo real, analizando la zona del rostro más representativa de las emociones, utilizando las microexpresiones y los movimientos oculares.

## 1.5. Estado del Arte

El reconocimiento de las emociones se presenta como una tarea desafiante debido a la complejidad y la ambigüedad de los aspectos que intervienen en la representación de una emoción. En la actualidad el área de reconocimiento de emociones tiene una amplia gama de aplicaciones e investigaciones.

En [41] los autores afirman que se pueden lograr mejoras de rendimiento a través de redes neuronales de aprendizaje profundo para el análisis de los sentimientos. Sus resultados mostraron que tanto las redes neuronales recurrentes como el aprendizaje por transferencia superan constantemente el aprendizaje automático tradicional. En sus resultados, las mejoras de rendimiento pueden variar hasta un 23,2 % en la puntuación F1 para la clasificación y un 11,6 % en MSE para la regresión. Los autores proponen *sent2affect*, como una estrategia personalizada de transferencia de aprendizaje que se basa en las diferentes tareas de análisis de sentimientos.

Un enfoque de aprendizaje profundo para la clasificación de emociones a través de señales de sensores de diferentes modalidades es propuesto en [36]. Según los autores su dinámica fue fusionar la interacción local de tres modalidades del sensor: en el cuerpo, ambiental y de ubicación. Los autores emplean una serie de algoritmos de aprendizaje que incluyen un enfoque híbrido que utiliza la red neuronal convolucional y la red neuronal recurrente de memoria a corto plazo (CNN-LSTM) en los datos brutos del sensor, de esta forma eliminan la necesidad de la extracción manual de características. Los resultados muestran que la adopción de enfoques de aprendizaje profundo es efectiva en la clasificación de las emociones cuando se utiliza un gran número de sensores de entrada (precisión promedio 95 % y F-Measure = % 95) y los modelos híbridos superan a la red neuronal profunda totalmente conectada tradicional (precisión media 73 % y F-Measure = 73 %). Además, los modelos híbridos superan a los algoritmos Ensemble

que desarrollaron previamente y utilizan ingeniería de características para entrenar la precisión promedio del modelo 83 % y F-Measure = 82 %).

En [42] se estudia el reconocimiento de emociones en el análisis de imágenes EEG. Los autores comentan que de forma general es difícil inducir la emoción deseada en sí mismo y ser igual con la de otras personas, ya que varios individuos reaccionan de manera diferente a los estímulos externos (audio, video, etc.). Ellos utilizan señales EEG utilizando técnicas de clasificación de series de tiempo basadas en la distancia para la identificación de las emociones, que involucran a diferentes personas expuestas a estímulos de audio. En esta investigación tuvieron el caso de que algunos de los participantes en el experimento no entendían el lenguaje de los estímulos, por lo que también investigaron el impacto de la comprensión del lenguaje en la percepción de las emociones.

En [31] se propone una nueva red neuronal convolucional para predecir la emoción en una imagen. Los autores proponen un modelo en dos partes: una red binaria de clasificación de emociones positivas o negativas y una red profunda para el reconocimiento específico de emociones. Durante el entrenamiento de la red, los autores introducen una estrategia de aprendizaje asistido para aumentar el rendimiento del reconocimiento. En sus resultados experimentales demuestran que la red propuesta es capaz de extraer características de nivel activo y logra ganancias significativas en la precisión del reconocimiento de emociones de 81.7 %.

En [48] los autores proponen un sistema inteligente de reconocimiento de emociones que proporciona un método flexible para monitorear los cambios emocionales en la vida diaria y enviar información de advertencia cuando ocurren estados emocionales inusuales y considerados como poco saludables debido a su frecuencia. Los autores proponen un sistema de reconocimiento de emociones para decodificar estados emocionales a partir de señales de electroencefalografía (EEG).

En [66] los autores hacen diversos experimentos con el objetivo de identificar si la raza influye en las emociones, enfocando su estudio en niños. Los autores mencionan que algunos estudios han encontrado que las características para el reconocimiento de las emociones expresadas por los rostros de diferentes razas varían para la mejora en la precisión de la identificación de emociones. La comunicación no verbal como los movimientos corporales, la expresión facial, los gestos y los movimientos oculares se utilizan para reconocer las emociones [65].

En [62] los autores realizan el reconocimiento de emociones, utilizando una clasificación de género y clasificación de voz. En sus experimentos los autores utilizan una

combinación de características faciales y del habla para una mayor eficiencia en el reconocimiento de emociones. Uno de los algoritmos más comunes es el Viola Jones para detectar regiones del rostro y MSER (Regiones extremas máximamente estables), MSER segmenta la imagen a partir de diferentes fronteras, agrupando los componentes conectados en secuencia y son tomados como los conjuntos de todas las regiones extremas. Para el manejo de la voz utiliza las técnicas de MFCC (Coeficientes Ceptrales de Frecuencia de Mel) este algoritmo realiza la extracción de características de voz utilizando los periodos de pitch para la frecuencia fundamental de cada periodo y toma el método de Cepstrum para obtener el pitch. Para realizar la clasificación de genero se utiliza el algoritmo de clasificación AdaBoost, y el algoritmo SVM para la clasificación de emociones en las expresiones faciales y de voz.

En [18] los autores realizan la identificación de cuatro emociones: feliz, calma, tristeza y miedo. Para esto los autores realizan la extracción de características desde señales de electroencefalograma (EGG), aplican la transformada de Hilbert-Huang (HHT) para la descomposición de la señal, HHT realiza la descomposición en modo empírico, de tal forma que la señal queda en varias ondas de coseno aproximadas y se analiza sus periodos y amplitudes. Este método suprime el ruido, en combinación de la entropía de aproximación (EMD) para finalmente realizar la extracción de características (E-ApEn). Se utilizó el método DBN, es un método que extrae las características profundas de los datos, y un modelo probabilístico con una estructura profunda, el DBN se compone de capas de máquinas de Boltzmann restringidas (RBM), se utilizó las SVM en combinación de DBN para la construcción del modelo de reconocimiento de las emociones. Se realiza un aprendizaje no supervisado con DBN y SVM como clasificador.

En [67] se propone un sistema de reconocimiento de emociones, basándose en el internet de las cosas para obtener imágenes faciales y señales del habla. El preprocesamiento de la señal la llevan a cabo en las nubes de borde, para posteriormente transmitir las en la nube central. En la nube central aplican un modelo pre-entrenado con una red neuronal convolucional (CNN) para la extracción de las características aprendidas en profundidad de las señales de las imágenes y del habla. Para la clasificación de las emociones se utiliza una máquina de soporte vectorial. Las técnicas utilizadas para la señal de la voz son, las ventanas de Hamming para el ventaneo, una transformada de Fourier aplicada para obtener un espectrograma de la señal, Filtros de paso de banda espaciados por Mel y un logaritmo para obtener un espectrograma de Mel. Se utilizó un modelo pre-entrenado AlexNet CNN para la señal de la voz, modificando solo la capa de salida para agregar las seis neuronas de las emociones. Para las imágenes del

video se aplicó la técnica de Viola Jones para la detección de rostros en las imágenes.

Los autores en [83] trabajan con señales de la respiración (RSP) para identificar la capacidad de la actividad psicológica, proponen un marco de aprendizaje profundo para extraer y reconocer la información emocional de la respiración. Toman como partida la teoría de la excitación-valencia (Circumplex de Russel) que ayuda a reconocer las emociones, esta técnica mapea las emociones en un espacio de dos dimensiones. Utilizan un codificador automático (SAE) con capas ocultas para extraer características relacionadas con las emociones y dos regresiones logísticas, para la clasificación de excitación y la otra para clasificación de valencia. La teoría de Russel indica que los estados emocionales se distribuyen en un espacio circular bidimensional con dimensiones de excitación y valencia. La excitación mide la intensidad de la activación emocional y la valencia describe la mente negativa o positiva.

En [16] los autores presentan un enfoque relacional difuso para el reconocimiento de emociones a partir de las expresiones faciales. Las expresiones faciales se analizan segmentando y localizando las regiones de interés del rostro, se difuminan y se mapean con apoyo de modelos relacionales de tipo Mamdani. Para la segmentación de la región de la boca se trabajó con un algoritmo de segmentación difusa, un algoritmo de agrupación Fuzzy C Means sensible al color; Para las regiones oculares y la región del cabello se trabajaron con un método de umbral tradicional. Sus parámetros de los codificadores difusos los ajustaron mediante un algoritmo de aprendizaje supervisado (algoritmo de retropropagación), donde generaban la emoción deseada a partir de las mediciones dadas del extracto facial.

En [5] se realiza el reconocimiento de expresiones faciales a partir de secuencias de video en 3D, la propuesta de esta investigación está en basarse por las curvas radiales que representan las caras en 3D, aplican un análisis de Riemann para cuantificar las deformaciones inducidas por las expresiones faciales y un análisis discriminante lineal (LDA) sobre las características. Utilizan una extracción de movimiento tridimensional con modelo de Markov oculto temporal (HMM) sobre las características para entrenamiento y un bosque aleatorio para captura de deformación para calcular las deformaciones medias. Se realiza un preprocesamiento donde la malla 3D en cada cuadro se alinea con la anterior, posteriormente se recorta. Cada superficie facial es representada por una colección de curvas radiales y utilizan un marco de referencia de Riemannian para estudiar las formas de esas curvas. Las deformaciones entre cuadros sucesivos de la malla del rostro se obtenían con extracción densa de las deformaciones en las características (DFS), de esta forma se cuantificaba el movimiento de los puntos de la cara

a lo largo de las curvas radiales y así capturaban los cambios en la geometría de la superficie facial.

Los autores en [82] proponen el reconocimiento de la expresión facial, capturando la variación dinámica de la estructura física facial en los videos, utilizando una red neuronal recidivante bidireccional jerárquica basada en partes (PHRNN) para analizar la información de expresión facial de secuencias temporales. Modelan las variaciones morfológicas faciales con el algoritmo PHRNN y obtienen las características constantes que son temporales, las cuales son basadas en puntos de referencia del rostro. Por último proponen una red neuronal convolucional con multiples señales (MSCNN) para extraer las características espaciales de tramas fijas, obtienen la información geométrica y de imagen dinámica con las redes PHRNN y MSCNN para obtener el reconocimiento de las expresiones faciales.

En [64] los autores proponen un modelo para el reconocimiento de expresión de rostros (FER) con redes neuronales convoluciones. La arquitectura de la red propuesta consta de cuatro capas aprendidas, tres convoluciones y una completamente conectada, la red se programa para el entrenamiento con la biblioteca TFLearn, TensorFlow Python, ya que se tiene la ventaja de la retroalimentación en tiempo real sobre el proceso y la precisión del entrenamiento. Utilizaron una unidad lineal rectificadora como función de activación, ya que esta función no requiere de normalización a la entrada para evitar saturación, la normalización se aplica después de la ReLu. Se clasifican siete emociones (neutral, sorpresa, triste, feliz, disgustado, temeroso y enojado).

Los autores en [78] proponen transferir el conocimiento de fuentes externas como imágenes y datos de tipo texto para el reconocimiento de emociones en video, utilizando un algoritmo de codificación de transferencia de imagen auxiliar (ITE) para agregar características de nivel de cuadro. Construyen un espacio vectorial semántico donde identifican las relaciones semánticas tanto visuales como textuales de las emociones. Se utilizo una arquitectura profunda de red neuronal convolucional (CNN), entrenada con AlexNet y una función de activación fc7 para cada cuadro.

En [33] los autores presentan un enfoque para el reconocimiento de emociones en imágenes con redes neuronales convolucionales profundas (DNN) con bloques residuales profundos. Se clasifica la emoción en triste, enojado, feliz, sorpresa, miedo, asco y neutral. Aplicaron una normalización gaussiana y la desviación estándar para el preprocesamiento de las imágenes. La red neuronal que utilizaron fue seis capas de convolución, dos bloques de aprendizaje residual profundo y una capa de agrupación

máxima después de cada capa de convolución, además de dos capas totalmente conectadas (FC) con una función de activación ReLu.

En [77] proponen la identificación de emociones, realizaron una base de datos con expresiones espontáneas de más de 100 personas, grabadas por una cámara térmica visible e infrarroja con iluminación desde tres direcciones, realizaron el reconocimiento utilizando análisis de componentes principales (PCA), PCA + análisis discriminante lineal (LDA), Modelo de apariencia activa (AAM) y el LDA basado en AAM. También utilizaron PCA y PCA + LDA para reconocer expresiones de imágenes térmicas infrarrojas.

En [32] los autores realizan el reconocimiento de las emociones utilizando características del habla. Utilizan un modelo de mezcla gaussiana (GMM), que trabaja con características espectrales. Realizan el entrenamiento de un GMM para cada emoción utilizando los coeficientes cepstrales de frecuencia de Mel (MFCC) y los coeficientes cepstrales predictivos lineales (LPCC), finalmente concatenando los GMM crean un super vector. El super vector GMM se usa como característica de entrada en el clasificador SVM con un kernel de divergencia GMM KL.

En [44] los autores proponen un método de extracción de características llamado incrustación discriminante localmente lineal (LLDE) para el reconocimiento de las emociones. El método se basa en la construcción de un modelo de traducción de vector y un cambio en la escala de distancia para la mejora de la capacidad del reconocimiento LLE, considerando que la función de costo de incrustación es invariante a traslación y reescalado, también consideran la transformación para maximizar el criterio de margen de maximización modificado (MMMC). Les realizan un preprocesamiento y recorte sobre las imágenes para la zona del rostro, extraen características con el método de incrustación para cada clase y aplicaron un clasificador de vecinos cercanos KNN con una métrica euclidiana.

En [54] los autores realizan un análisis de las técnicas basadas en funciones y técnicas basadas en modelos. Concluyen que el mejor método está determinado por factores como la iluminación, la oclusión y la falta de resolución. El uso de algoritmos basados en características locales como descriptores de superficies locales a múltiples escalas que separan peculiaridades únicas alrededor de puntos clave localizados son mejores que un modelo de máscara para el reconocimiento de las emociones si se presentan esos problemas ya mencionados. Determinaron un enfoque +óptimo para el método de extracción de características basada en Curvelet en el conjunto de datos FRGC v2 arroja una mayor precisión del 97,83 %.

Los autores proponen en [72] un modelo de reconocimiento de emociones mediante características faciales. Este modelo es propuesto para el sistema de educación a distancia, para identificar los estados de ánimo de los alumnos en sus clases bajo ese plan de trabajo. Localizan los puntos característicos del rostro, utilizando el algoritmo de modelos de apariencia activa (AAM) con un estándar MPEG-4, que es un estándar para la selección de puntos característicos. Y los algoritmos FAGS para derivar parámetros de definición facial (FDP) y parámetros de animación facial (FAP), del cual toman 66 parámetros para la descripción del movimiento del rostro. Aplican un sistema de reconocimiento de las emociones en tiempo real basado en máquinas de soporte vectorial en función del kernel con base radial con método uno contra uno, les arroja una precisión de 84,55 %.

## 1.6. Metodología

1. Investigar las técnicas para determinar el inicio y fin de los frames dentro de un video para la identificación en tiempo real
2. Investigar y estudiar las técnicas de preprocesamiento y reconocimiento facial que muestren mejores resultados.
3. Investigar y analizar técnicas para extraer los vectores de características para la identificación de las expresiones faciales y oculares.
4. Implementar las técnicas de preprocesamiento en las imágenes de ser necesario y determinar que técnicas se requieren
5. Emplear los algoritmos de extracción de características para cada emoción
6. Recopilar los resultados de la extracción de las características de forma individual para expresiones faciales y expresiones oculares para determinar las mejoras de precisión en forma individual y combinada.
7. Analizar los algoritmos de clasificación y probar su eficiencia para nuestro sistema.
8. Realizar las pruebas pertinentes del funcionamiento del sistema.
9. Realizar la validación de los algoritmos

10. Analizar y corregir los resultados.
11. Presentar los resultados finales y conclusiones.



# Capítulo 2

## Marco Teórico

### 2.1. Redes Neuronales

Uno de los estudios más destacados en el campo de la inteligencia artificial, son las redes neuronales artificiales (RNAs). Las RNAs son modelos que tratan de emular las posibles respuestas del cerebro humano, son una imitación de la estructura neuronal de nuestro cerebro. El aprender y memorizar significa hacer asociaciones de los eventos con cambios en las neuronas y sus contactos con otras neuronas en red, el cerebro tiene unos cien millones de neuronas que se conectan entre si mediante sinapsis, que es el proceso de transmitir la información de unas a otras neuronas. Las RNAs se componen de un conjunto de entradas externas, un conjunto de capas de procesamiento y una función de activación para cada capa de procesamiento. Las redes neuronales que aprenden mediante el ejemplo, programadas para tareas específicas como el reconocimiento de patrones, o clasificación. Las RNAs tienen diferentes modelos, con distinta arquitectura y también se tienen diversas funciones de activación. La salida de las capas de una red neuronal está definida por la función de activación. Se prefieren funciones con una derivada simple, ya que tienen un menor costo computacional, debido a la menor cantidad de cálculos matemáticos [28].

En la Figura 2.1 se muestra el funcionamiento básico de una red neuronal, cada capa contribuye a la entrada de las capas sucesivas a las que esté conectada. La Figura muestra un ejemplo de red neuronal con cuatro entradas  $X_1, X_2, X_3$  y  $X_4$ , una salida  $L$  y  $\Theta_k$ , donde  $\Theta$  nos indica el valor del bias sobre esa capa [68].

La salida se representa con la siguiente ecuación:

$$L = \sum_j w_{jk}(t)x_j(t) + \Theta_k(t) \quad (2.1)$$

Para actualizar los pesos  $w_i$ , se utiliza la siguiente ecuación:

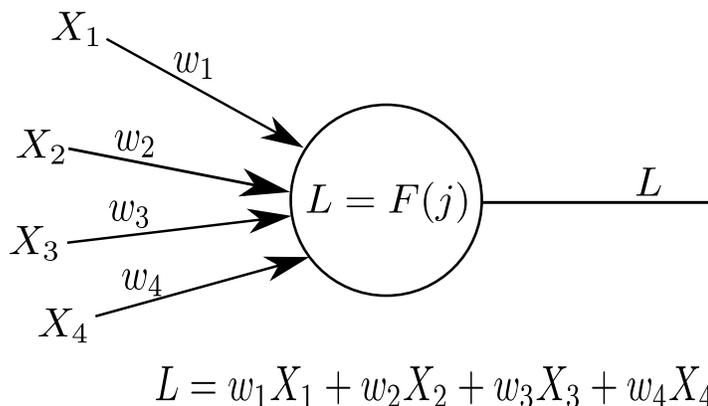


FIGURA 2.1: Redes Neuronales (RNAs)

$$w_i = w_i(\text{old}) + \alpha(t - y_j)x_j \quad (2.2)$$

Para la actualización del bias:

$$\Theta_i = \Theta_i(\text{old}) + \alpha(t - y_j) \quad (2.3)$$

En cada modelo de RNN es necesario un parámetro de aprendizaje definido por  $\alpha$ , valor que se asigna aleatoriamente bajo un intervalo de 0 a 1, y un valor de  $t$  target que representa nuestra salida deseada cuando nuestro problema es de clasificación.

### 2.1.1. Redes Neuronales Recurrentes

Las redes neuronales recurrentes (RNN) son aquellas redes cíclicas, con conexiones de retroalimentación, lo que permite a la red obtener una temporalidad de los datos y se pueda considerar que la red tenga memoria. Las RNN están diseñadas para aprender patrones secuenciales o de tiempo variable lo cual permite que esta red sea potente para problemas relacionados con análisis de secuencias temporales con eventos. Algunos estudios recientes en RNN destacan para el análisis de textos, sonido o video, incluso seguimiento predictivo del movimiento de la cabeza y predicción financiera. Existen diferentes tipos de arquitecturas de RNN de acuerdo con el número de capas ocultas y la forma de hacer retro propagación (ciclos). Las arquitecturas pueden partir con las capas completamente interconectadas unas a otras o parcialmente conectadas, incluso se pueden tener las capas con una retroalimentación a la misma capa [34]. El trabajar

con RNN es posible continuar propagando valores de activación de forma infinita ya que es un ciclo, debemos establecer una función de paro llamada atractor, los valores de activación se actualizan hasta que se alcanza ese atractor y posteriormente actualizar los pesos.

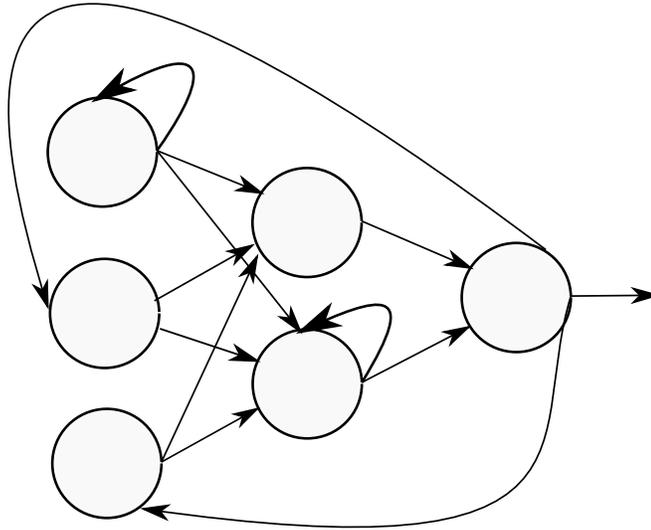


FIGURA 2.2: Conexión de una Red Neuronal Recurrente (RNN)

Se tienen diferentes conexiones de RNN, tales como:

1. RNN Simple, es la arquitectura más simple compuesta por una sola neurona que recibe una entrada y produce una salida, enviando esa salida a si misma creando un ciclo. Para cada paso (timestep) la neurona recibe como entrada el resultado de la capa anterior o del paso anterior para generar su salida.

$$Y_t = f(wx_t + Uy_{t-1} + w_0) \quad (2.4)$$

Donde  $X = (X_1, X_2, \dots, X_t)$  indica el vector de entradas  $X$  de la capa anterior,  $w$  son los pesos y el bias es representado por  $w_0$ . Para el intervalo de tiempo tenemos a  $U$  es la matriz de pesos que opera sobre la red en el instante de tiempo anterior  $y_{t-1}$  y el entrenamiento se realiza con el algoritmo Backpropagation [74].

2. RNN de Elman, es la arquitectura base en donde se implementa la retroalimentación con la cual se crea la temporalidad, lo que nos permite agregar a las RNN la cualidad de memoria. En esta red se introduce un conjunto de unidades, como unidades de entrada adicionales y sus valores de activación se retroalimentan de las capas ocultas. Se caracteriza porque sus unidades ocultas son las que se retroalimentan y sus entradas adicionales no tienen auto conexiones [14].

El algoritmo de aprendizaje se realiza de la siguiente manera:

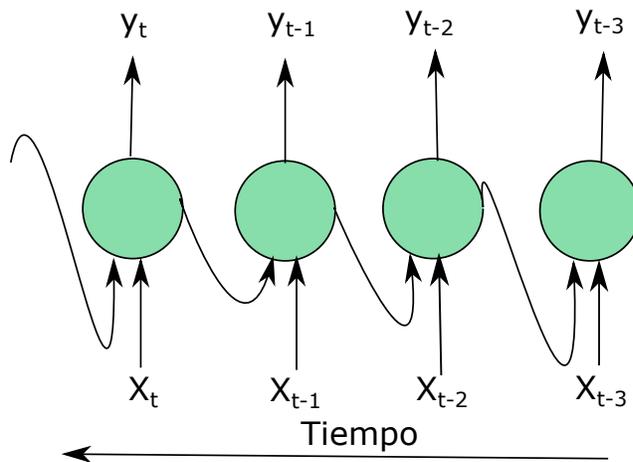


FIGURA 2.3: Conexión de una Red Neuronal Recurrente (RNN) Simple con los intervalos de tiempo (timestep).

1. Las unidades de contexto se establecen en  $0; t = 1$ .
2. Se sujeta el patrón  $x^t$ , se realizan los cálculos hacia adelante una sola vez.
3. Se aplica la propagación hacia atrás.
4.  $t = t + 1$ ; vaya a 2. Las unidades de contexto en el paso  $t$  siempre tienen el valor de activación de las unidades ocultas en el paso  $t - 1$ .

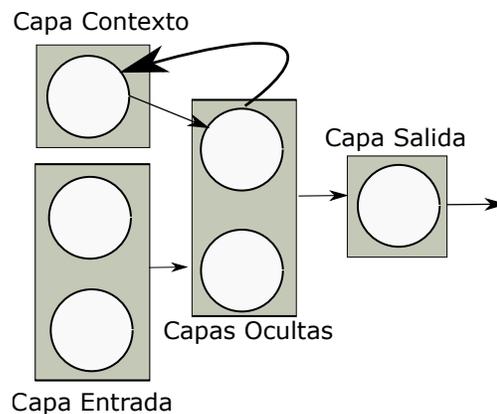


FIGURA 2.4: Arquitectura de Elman (RNN Elman)

3. La red de Hopfield consta de un conjunto de neuronas interconectadas, que se actualizan de forma asincrónica y de forma independiente a las demás neuronas. Cada una de las neuronas funcionan como neurona de entrada y salida, tomando en cuenta que

los valores de activación son binarios (1 y 0). Sin embargo, el utilizar (+1 y -1) genera más ventajas en la simetría de los estados de la red.

El estado del sistema está dado por valores de activación  $y = (y_k)$ , de una entrada  $s_k(t + 1)$  de una neurona  $k$  en el ciclo  $t + 1$ .

$$s_k(t + 1) = \sum_{j \neq k} y_j(t) w_{jk} + \Theta_k \quad (2.5)$$

Se aplica una función de umbral simple a la entrada para obtener el valor nuevo de activación  $y_i(t + 1)$  en el tiempo  $t + 1$ .

$$y_k(t + 1) = \begin{cases} 1 & s_k(t + 1) > U_k \\ -1 & s_k(t + 1) < U_k \\ y_k(t) & \text{otro} \end{cases} \quad (2.6)$$

$y_k(t + 1) = \text{sgn}(s_k(t + 1))$  por simplicidad se toma  $U_k = 0$ , una neurona se considera estable cuando si  $t$  es acorde a la sig. ecuación.

$$y_k = \text{sgn}(s_k(t - 1)) \quad (2.7)$$

Un estado  $\alpha$  es llamado estable, por lo tanto cuando la red está en un estado  $\alpha$ , todas las neuronas están estables. Una aplicación principal de la red Hopfield es la memoria asociativa, los pesos deben establecerse de tal forma que los patrones que se almacenarán en la red sean estables, cuando la red recibe un patrón ruidoso o incompleto genera los datos incorrectos o faltantes al iterar a un estado estable. Para almacenar patrones utilizamos la regla de Hebb:

$$w_{jk} = \begin{cases} \sum_{p=1}^P x_j^p x_k^p & j \neq k \\ 0 & \text{otros} \end{cases} \quad (2.8)$$

### 2.1.2. Red Neuronal Convolutiva

Las redes neuronales convolucionales (CNN) son muy parecidas a las redes neuronales multicapas; Cada parte de las CNN es entrenada para realizar una tarea, de esta forma se reducen de forma significativa el número de capas ocultas, creando un entrenamiento en menor tiempo. Las redes neuronales convolucionales tienen un aprendizaje supervisado y están construidas con tres tipos de capas:

- Capas de convolución.
- Capas de reducción para características únicas.
- Una capa clasificadora.

Estas características de las CNN les dan la cualidad de ser muy potentes para el análisis de imágenes, debido a su reducción de características. Las CNN contienen capas ocultas especializadas, que van de lo específico a lo general, por ejemplo, las primeras capas identifican bordes y curvas, posteriormente las capas van reconociendo formas complejas de interés. Esta red debe aprender por sí sola a reconocer un objeto, aprende las características únicas de cada objeto de interés y lo generaliza, de tal forma que ese objeto lo reconozca sin importar la oclusión; Es por ello por lo que requiere de grandes cantidades de imágenes del objeto de interés, ya que realiza un estudio exhaustivo de tal objeto.

El funcionamiento de una red neuronal convolucional se realiza de la siguiente forma:

1. Determinamos en cuantos canales de color se trabajará. Si es escala de grises tomaremos cada uno de los píxeles como una neurona de entrada. Para los tres canales de color tendríamos tres neuronas de diferente valor de intensidad para cada píxel, por lo que sería lo triple de neuronas a comparación de una escala de grises.
2. Se realiza una normalización de los valores de entrada, entre 0 y 1. píxel/255.
3. Se realiza la convolución con un kernel, se genera una matriz resultante que serán los valores de la siguiente capa de neuronas. El kernel tomara valores aleatorios de entrada, se ajustarán mediante backpropagation.
4. Se aplica una función de activación, la función de activación más utilizada en estas redes es ReLu, obteniendo un mapa de detección de características.
5. Posterior a una capa de convolución sigue una capa de reducción, conocida como subsampling, en este proceso se filtrarán las características más importantes de cada kernel. El subsampling más utilizado es Max Pooling. Max Pooling preserva el valor más grande de cada espacio de la imagen, de esta forma se reduce la imagen y se obtienen mejor cantidad de neuronas para la siguiente capa de convolución.

### 2.1.3. Máquinas de Soporte Vectorial (SVM)

Técnica que se introdujo en el paradigma de redes neuronales para la clasificación de patrones, inicialmente en la clasificación binaria, contemplando el proyectar los datos en espacios de más dimensiones que los originales y así lograr una mayor separabilidad de las clases, las SVM realizan sus procesos en dos fases: entrenamiento y decisión.

Las SVM están constituidas por algoritmos de aprendizaje supervisado, estas aparecieron en los años noventa, anteriormente ya se tenían avances que tomaron como ideas para su desarrollo, como fue el uso de los kernels, su interpretación geométrica y la construcción de un hiperplano de separación óptimo en un contexto no paramétrico [43] [38].

#### Fase de Entrenamiento

Basada en la observación de un conjunto  $X$  de  $n$  muestras, las salidas del sistema son dos valores simbólicos  $y \in \{+1, -1\}$  de forma que el conjunto de entrenamiento está formado por los pares  $(x_i, y_i), i = 1 \dots n$  donde cada vector  $x_i$  se corresponde con un vector de entrenamiento y los valores  $y_i \in \{+1, -1\}$  indican la clase a la que pertenece cada vector. El objetivo del proceso de entrenamiento consiste en encontrar una función de decisión capaz de separar las dos clases, en caso de que las clases no sean separables los vectores de entrenamiento se proyectan a un espacio de dimensión superior mediante el uso de funciones de transformación no lineales, de ser así, la función de decisión se sitúa en el hiperplano de esa dimensión, la función de decisión se define con la siguiente ecuación ([1]):

$$F(x) = \sum_{i=1}^n \alpha_i y_i H(x_i, x) - b \quad (2.9)$$

donde  $b$  es una constante,  $\alpha_i, i = 1, \dots, n$ .

Los patrones  $x_i$  asociados con valores de  $\alpha_i$  distintos de cero se denominan vectores de soporte, determinados los vectores de soporte la función de decisión se muestra:

$$f(x) = \sum_{\text{support vectors}} \alpha_i y_i H(x_i, x) - b \quad (2.10)$$

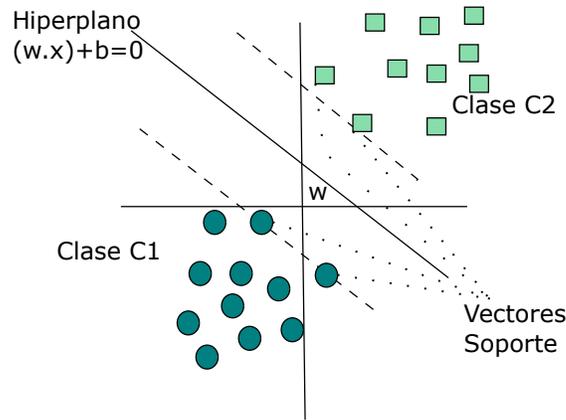


FIGURA 2.5: Hiperplano Máquinas de Soporte Vectorial (SVM)

Existen diferentes kernels para mapear el conjunto de datos de entrada:

1. Lineal

$$H(x, y) = x^t y \quad (2.11)$$

2. Funciones de base radial Gaussiana

$$H(x, y) = \exp\left\{-\frac{\|x - y\|^2}{2\sigma^2}\right\} \quad (2.12)$$

3. Exponencial

$$H(x, y) = \exp\left\{-\frac{\|x - y\|}{2\sigma^2}\right\} \quad (2.13)$$

4. Polinomios con grado  $d$ :

$$H(x, y) = (1 + \langle x, y \rangle)^d \quad (2.14)$$

5. Función sigmoide

$$H(x, y) = \tanh(p \langle x, y \rangle + \gamma) \quad (2.15)$$

donde  $\langle x, y \rangle$  se refiere al producto interno con  $\rho$  y  $\gamma$ , como parámetros de ajuste.

### Fase de prueba

Se determina la clase a la que pertenece de acuerdo con el signo (polaridad) de  $f(x)$ . La magnitud puede considerarse como una medida de certidumbre sobre la decisión [47].

Los vectores de soporte son los datos más representativos de todos, se llega a la misma solución que con todos los patrones.

La distancia mínima desde el hiperplano que separa las clases al patrón más cercano se denomina margen  $\tau$ . Un hiperplano de separación se denomina óptimo si el margen es máximo; la distancia entre el hiperplano de separación y un patrón  $z$  es  $y_k |f(x)| / \|w\|$   $w$  está dado por:

$$w = \sum_{i=1}^n \alpha_i y_i x_i \equiv \sum_{\text{vectores soporte}} \alpha_i y_i x_i \quad (2.16)$$

la constante  $b$  se obtiene:

$$\alpha_i \{y_i [(w \cdot x_i) + b] - 1\} = 0 \quad (2.17)$$

Suponiendo que existe un margen  $\tau$  todas las muestras de entrenamiento obedecen la desigualdad:

$$\frac{y_k f(x_k)}{\|w\|} \geq \tau, k = 1, \dots, n \quad (2.18)$$

donde  $y_k = \{+1, -1\}$ .

Para disminuir el número infinito de soluciones que difieren sólo en el escalado de  $w$ , se debe fijar la escala según el producto de  $\tau$  y la norma de  $w$ .  $\tau \|w\| = 1$ . Maximizar el margen  $\tau$  es equivalente a minimizar la norma de  $w$ .

## 2.2. Sensores

Los sensores han sido utilizados para la automatización de los sistemas de clasificación y reconocimiento, permitiendo estos la recolección de información de variables físicas tales como la longitud, ángulo de giro, temperatura, presión, movimientos, etc. Los sensores trabajan convirtiendo las variables físicas en señales eléctricas o en términos de presión, el funcionamiento de los sensores puede ser de contacto físico o sin contacto. En el reconocimiento de emociones se han utilizado sensores de movimiento y en proceso de evaluación de siluetas de visión en 3D [25].

Se tienen un conjunto de características que definen el comportamiento de un sensor:

- Rango, indica los valores máximos y mínimos que alcanza el sensor.

- Exactitud, indica el mayor valor de error esperando entre señales medidas e ideales.
- Repetitividad, capacidad del sensor para reproducir una lectura con una precisión.
- Reproducibilidad, participa como la repetitividad, pero se utiliza cuando se toman medidas bajo condiciones diferentes.
- Resolución, es la cantidad de medida mínima
- Error, diferencia entre el valor medido y los reales.
- Sensibilidad, razón de cambio de la salida frente a cambios en la entrada.
- Excitación, corriente o voltaje necesaria para el funcionamiento del sensor.
- Estabilidad, mide la posibilidad de que el sensor tenga la misma salida constante a una misma entrada sin cambios.

### 2.2.1. Sensores Movimiento

Los sensores analógicos son sensores que convierten una magnitud física en una señal analógica. Los sensores analógicos de movimiento permiten obtener datos de movimiento lineal y rotativo. Los sensores de movimiento en los sistemas de reconocimiento de emociones son utilizados para el movimiento del cuerpo, tomando como base los movimientos característicos para los estados de ánimo. Los sensores miden la magnitud con la que se desplaza los objetos, acompañados de un sensor de posición que indican la posición del objeto, tomando estos datos a un punto de referencia de cada movimiento [25].

### 2.2.2. Sensores de entorno

Los sensores de entorno captan estímulos de su entorno y traduce la información en impulso eléctrico. Para los sistemas de reconocimiento de emociones han utilizado diferentes sensores de entorno, se enfocan en extraer información ambiental e información de los ruidos externos tales como música. Los sensores para extraer los ruidos externos funcionan como un micrófono donde el sensor lo que hace es clasificar ruidos agudos, graves entre otros para determinar cómo influye en el estado de ánimo de la persona.

Los ruidos intervienen en nuestra persona de diferente forma, pero por ejemplo una discusión se manifestaría de forma negativa en una persona [30].

## 2.3. Imágenes

Una imagen es la representación visual de un objeto, que pueda ser captada bajo cierta frecuencia [24].

Una imagen digital es una matriz de píxeles, el píxel es la mínima unidad de una imagen. La imagen se forma bajo ciertos parámetros como el color de la luz, el material de la superficie y la sensibilidad de la cámara. La intensidad del color se relaciona con la luz de la toma de la imagen, está dada por:

$$I(p) = (R + G + B) \frac{255}{\max(R_i + G_i + B_i)} \quad (2.19)$$

El color es una característica de la luz, el ojo humano percibe un espectro visible entre 400nm a 700nm, cada energía es representada por una onda de longitud, su frecuencia es dada de acuerdo con el periodo de tiempo con la que se repite la señal. El color de la luz es la función  $I$ . El material de la superficie es el que determina que longitud de onda es la que se refleja y las demás las absorbe; El porcentaje de la luz que refleja el material es dado por la función  $S$  [53]. La luz reflejada está dada por  $I(\lambda) * S(\lambda)$ . Las cámaras tienen tres tipos de sensores que integran una longitud de onda (R,G,B), abarcando todo el espectro visible.

$$I(p) = \int (I(\lambda)S(\lambda)R(\lambda))d\lambda, \int (I(\lambda)S(\lambda)G(\lambda))d\lambda, \int (I(\lambda)S(\lambda)B(\lambda))d\lambda \quad (2.20)$$

### 2.3.1. Imágenes de las señales de un EEG

Las imágenes EEG están dadas por un electroencefalograma, el cual es una colección de señales de la actividad cerebral registrada en el cuero cabelludo, se obtienen al medir las corrientes eléctricas en movimiento del cerebro. Estas señales permiten a los sistemas de IA reconstruir imágenes que han sido observadas por la persona [7].

El electroencefalograma fue descubierta por Hans Berger en 1924, consiste en obtener una señal de pulsaciones eléctricas del cerebro. La encefalografía se clasifica en dos grupos: la forma invasiva, consiste en implantar electrodos dentro del cráneo, tiene la ventaja de focalizar la señal, donde se puede distinguir una parte específica del cerebro.

La técnica no invasiva, adquieren las pulsaciones desde el cuero cabelludo, a través de pares de electrodos conductores de plata, que permiten la lectura de las señales eléctricas. La actividad eléctrica es producida debido a la comunicación entre las neuronas, lo cual es llamado actividad cerebral [73].

Los electroencefalogramas obtienen cinco tipos de ondas con diferentes características, clasificadas como:

- *Delta*: Ondas que van de 0,5 a 4 Hz. Son las ondas más lentas y están presentes cuando una persona duerme, el que se manifieste esta onda en estado de vigilia se determina a defectos físicos en el cerebro.
- *Theta*: Ondas entre 4 y 7,5 Hz. Vinculadas a la ineficiencia y el soñar despierto, suelen relacionarse con el acceso a material inconsciente del cerebro y estados de profunda meditación.
- *Alfa*: Ondas entre 8 a 13 Hz. Ondas lentas y asociadas con la relajación y desconexión. Estado relajado de consciencia, sin atención o concentración.
- *Beta*: Ondas de 14 y 26 Hz. Ondas pequeñas y rápidas asociadas a concentración enfocada o bien un estado de pánico.
- *Gamma*: Ondas mayores a 30 Hz. La amplitud es muy pequeña y su ocurrencia es rara, por lo que se relacionan a enfermedades del cerebro. Refleja el mecanismo de la conciencia, la unión de las ondas beta y gamma se asocian a la atención, percepción y cognición.

### 2.3.2. Binarias

Cada una de las intensidades de los píxeles se encuentran normalizados entre un valor de 0 ó 255 determinados por un umbral T, de forma arbitraria se determina el color del píxel en negro o blanco, de acuerdo con el valor del umbral.

$$f(x, y) = \begin{cases} 0 & \text{si } p \leq p1 \\ 255 & \text{si } p > p1 \end{cases} \quad (2.21)$$

### 2.3.3. Imágenes RGB

Imagen compuesta por tres canales de intensidades, una matriz de rojo, matriz de verde y la tercera matriz de azul. Cada uno de los píxeles esta dado por tres valores de las tres dimensiones de color, definidos como:

$$R = \frac{R}{R+G+B}; G = \frac{G}{R+G+B}; B = \frac{B}{R+G+B} \quad (2.22)$$

## 2.4. Preprocesamiento de imágenes

### 2.4.1. Operaciones grupales

Las operaciones grupales se denominan operaciones de vecindad, nos permiten mejorar el contraste de una imagen, para trabajar con la vecindad de un píxel se utilizan mascarar. Donde I es la máscara y W son los pesos de los vecinos del píxel que se está analizando.

$$\begin{bmatrix} W_{1,1} & W_{1,2} & W_{1,3} \\ W_{2,1} & W_{2,2} & W_{2,3} \\ W_{3,1} & W_{3,2} & W_{3,3} \end{bmatrix} \begin{bmatrix} I_1 & I_2 & I_3 \\ I_4 & I_5 & I_6 \\ I_7 & I_8 & I_9 \end{bmatrix} \quad (2.23)$$

El cálculo de un nuevo píxel  $R_{ij}$  es definido mediante  $R$

$$R_{ij} = W_{1,1}I_{i-1,j-1} + W_{1,2}I_{i,j-1} + W_{1,3}I_{i+1,j-1} + W_{2,1}I_{i-1,j} + \dots + W_{3,3}I_{i+1,j+1} \quad (2.24)$$

### 2.4.2. Técnicas de Filtrado

Las técnicas de filtrado son utilizadas para resaltar o suprimir, de forma selecta las intensidades de las imágenes. Los filtros más utilizados son los de paso bajo para suavizar la imagen y los pasa alto para aumentar contraste. Los filtros direccionales detectan en la imagen estructuras con misma dirección y los filtros de detección de bordes que permiten identificar los objetos y aislarlos de acuerdo con sus propiedades homogéneas [53]. El filtrado consiste en la aplicación de una matriz de filtro sobre cada conjunto de píxeles, generando un nuevo valor. El filtrar una imagen (f) consisten en aplicar una transformación (T) para obtener una nueva imagen (S), se define una técnica de filtrado con la siguiente ecuación:

$$S(x, y) = T[f(x, y)] \quad (2.25)$$

## Filtros pasa bajo

Los filtros pasa bajo son aquellos que dejan pasar las bajas frecuencias y atenúan o eliminan las altas frecuencias, los pasa bajo nos entregan un suavizado de la imagen, eliminando el ruido de las imágenes. Las técnicas de filtrado son técnicas que permiten resaltar o suprimir de forma selectiva la información obtenida en una imagen a diferentes escalas espaciales. El filtrado tiene la ventaja de ocultar o destacar elementos en una imagen, incluso omitir valores.

## Filtro Moda

El filtro de la moda es un operador de tipo grupal, realiza el suavizado del píxel central de acuerdo con los píxeles cercanos. Se aplica una máscara de 3x3 (píxel central y 8 vecinos próximos), se realiza el recorrido de la máscara por toda la imagen, tomando cada píxel como píxel central. Se toma el valor más frecuente de los 9 píxeles que intervienen en cada iteración, al encontrar más de un valor frecuente, se toma cualquiera de los valores frecuentes para realizar el suavizado [60].

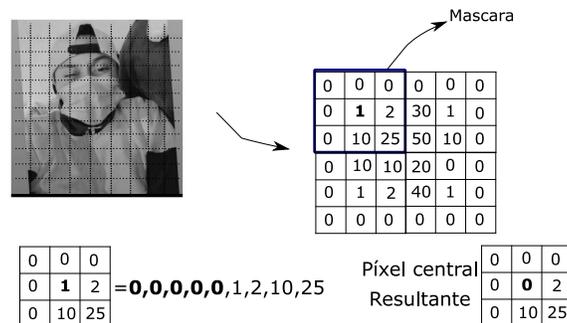


FIGURA 2.6: Filtro Moda

## Filtro de la mediana

El filtro de la mediana es utilizado para realizar un suavizado en las imágenes, creando tonalidades homogéneas de acuerdo con los píxeles que le rodean. Es un filtro útil para eliminar ruido de tipo sal y pimienta. Se basa en acomodar los píxeles de la máscara de menor a mayor e identificar el valor central. Ese nuevo valor central será el que reemplazara nuestro píxel central de la máscara de nuestra imagen original [60].

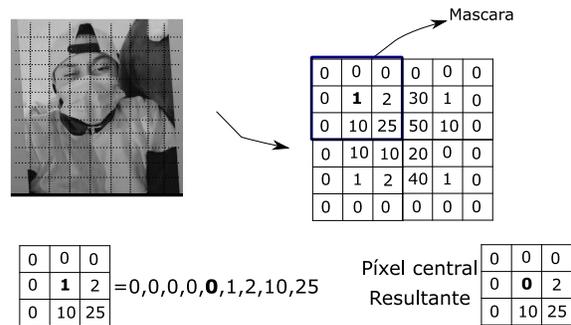


FIGURA 2.7: Filtro Mediana

### Filtro adaptativo

Consiste en calcular la varianza de la imagen y la varianza de la imagen de forma local (mascara). Elimina las tonalidades que desentonan en las tonalidades de mayor ponderación en la imagen. Es un filtro muy bueno para eliminar el ruido [60].

$$Adaptativo = f(x, y) - \frac{\sigma_n^2 - m_l}{\sigma_l^2} [f(x, y) - m_l] \quad (2.26)$$

Donde  $\sigma_n^2$  es la varianza del ruido relativa de toda la imagen,  $\sigma_l^2$  es la varianza local de la ventana y  $m_l$  es la media local de la ventana.

$$Desviación = \frac{(x - \bar{x})^2}{n - 1} \quad (2.27)$$

Donde  $\bar{x}$  es la media de la imagen o ventana;  $x$  es el valor del píxel y  $n$  es el tamaño de la imagen o ventana.

### 2.4.3. Filtros Pasa Altos

Los filtros se utilizan para detectar cambios de luminosidad, para la detección de patrones como bordes o para resaltar detalles finos de una imagen.

### Transformaciones de Intensidad

Una transformación de intensidad consiste en modificar los valores de intensidad de cada píxel a otros valores de acuerdo con las técnicas de mejora de la imagen. Las

técnicas se clasifican de acuerdo con el operador que utilizan, pueden ser de tipo puntual o grupal.

#### 2.4.4. Histogramas

El histograma de una imagen es una función que representa los niveles de intensidad de la imagen  $g$ . La probabilidad  $P(g)$  de repetición de un determinado nivel  $g$  se define como:

$$P(g) = \frac{N(g)}{T} \quad (2.28)$$

Donde  $T$  es el número de píxeles en la imagen y  $N(g)$  es el número de píxeles en el nivel de intensidad  $g$ . Un Histograma contiene el número de píxeles que tienen el mismo nivel de intensidad. Se representa como un gráfico de barras en el que las abscisas son los distintos colores de la imagen y las ordenadas la frecuencia con la que cada color aparece en la imagen. El histograma proporciona información sobre el brillo y el contraste de la imagen [52].

#### Expansión del histograma

Esta operación consiste en distribuir las tonalidades de la imagen, expande los píxeles en todo el ancho del histograma, ya que el valor de la intensidad más baja se posiciona en cero y la intensidad más alta se coloca en el valor máximo del histograma. Con la expansión del histograma se puede aumentar el contraste, si una imagen es demasiado oscura se volverá más visible [23].

La expansión de un histograma se define de la siguiente manera:

Donde  $f(i, j)$  es el nivel de gris;  $f(i, j)_{MAX}$  es la intensidad máxima de la imagen  $f$ .  $f(i, j)_{MIN}$  es el menor valor de la intensidad en la imagen  $f$ .  $MAX$  y  $MIN$  corresponden al máximo y mínimo posible de los niveles de gris (para una imagen de 8 bits sería 0 y 255 respectivamente) [52].

$$G(i, j) = \left[ \frac{f(i, j) - f(i, j)_{MIN}}{f(i, j)_{MAX} - f(i, j)_{MIN}} \right] [MAX - MIN] + MIN \quad (2.29)$$

### Contracción del histograma

Esta técnica realiza una disminución del contraste de la imagen.

Donde  $f(i, j)$  es el valor del píxel de la imagen de entrada,  $f(i, j)_{MAX}$  es la intensidad más grande en valor del píxel de toma la imagen  $f$ .  $f(i, j)_{MIN}$  es la intensidad mínima de la imagen  $f$ .  $C_{MAX}$  y  $C_{MIN}$  corresponden al máximo y mínimo valores posibles de los niveles de gris (0 y 255) [52].

$$G(i, j) = \left[ \frac{C_{MAX} - C_{MIN}}{f(i, j)_{MAX} - f(i, j)_{MIN}} \right] [f(i, j) - f(i, j)_{MIN}] + C_{MIN} \quad (2.30)$$

### Desplazamiento del histograma

El desplazamiento del histograma se usa para aclarar y oscurecer una imagen pero manteniendo la relación entre los valores de los niveles de intensidad. Para oscurecer se suma un valor constante a todos los píxeles de la imagen. Para aclarar la imagen se resta un valor constante en todos los píxeles. Donde  $DES$  es un valor constante para aplicar el desplazamiento del histograma [52].

$$G(i, j) = f(i, j) + DES \quad G(i, j) = f(i, j) - DES \quad (2.31)$$

### Ecualización de Histograma

Mediante la ecualización se realiza la distribución adecuada en el nivel de luminosidad de la imagen, este proceso modifica las intensidades, colocándolas en la misma frecuencia distribuyendo los valores de intensidad a lo largo del espectro [55]. donde  $g_{max}$  y  $g_{min}$  es la intensidad mínima y máxima posibles en una imagen (0 y 255). Con  $P_g(g) = \sum_{g=0}^g p(g)$  y  $\alpha$  un parámetro a definir [52].

### Normalización de Histograma

Para evitar que los valores de un histograma sean muy dispares, se puede normalizar dicho histograma. Discretizando todas las intensidades de la imagen.

Donde  $N_{max}$  es el nuevo máximo valor y  $N_{min}$  es el nuevo mínimo valor, para discretizar en ese intervalo [52].

TABLA 2.1: Tipos de ecualización

Ecualización	Expresión Matemática
Uniforme	$F(g) = [g_{\text{máx}} - g_{\text{mín}}] P_g(g) + g_{\text{mín}}$
Exponencial	$F(g) = g_{\text{mín}} - \frac{1}{\alpha} \ln [1 - P_g(g)]$
Reyleigh	$F(g) = g_{\text{mín}} + \left[ 2\alpha^2 \ln \left( \frac{1}{1 - p_g(g)} \right) \right]^{\frac{1}{2}}$
Hipercúbica	$F(g) = \left( [\sqrt[3]{g_{\text{máx}}} - \sqrt[3]{g_{\text{mín}}}] P_g(g) + \sqrt[3]{g_{\text{mín}}} \right)^3$
Logaritmo hiperbólica	$F(g) = g_{\text{mín}} \left[ \frac{g_{\text{máx}}}{g_{\text{mín}}} \right] P_g(g)$

$$N_{x,y} = \frac{N_{\text{max}} - N_{\text{min}}}{O_{\text{máx}} - O_{\text{mín}}} \times (O_{x,y} - O_{\text{mín}}) + N_{\text{mín}} \quad (2.32)$$

### Contraste

La diferencia entre tonalidades de acuerdo con la intensidad de la luz en tonos claros. Un contraste alto oscurece las sombras y las tonalidades oscuras y se destacan las intensidades claras; mientras que un contraste bajo disminuye la intensidad general de la imagen. El contraste se define como las variaciones entre las intensidades claras y oscuras, si el intervalo entre esas intensidades es menor tendremos menor contraste y si el intervalo es amplio el contraste es más grande [52] [63].

El ajuste del contraste se define por la siguiente ecuación:

$$f(x, y) = C * (f(x, y) - 128) + 128 \quad (2.33)$$

Donde C es el contraste y 128 es el valor promedio entre posibles intensidades.

### Brillo

Modificar el brillo de una imagen consiste en sumar o restar una constante a cada píxel, dentro del intervalo límite 0 y 255 [4].

$$R(x, y) = I(x, y) + KR(x, y) = I(x, y) - K \quad (2.34)$$

## 2.5. Segmentación y detección de bordes

La segmentación son técnicas que se basan en separar las intensidades, separando un objeto en específico. La segmentación es una técnica útil para realizar procesos de extracción de bordes y regiones; se puede utilizar para separar objetos con intensidades diferentes, esa diferencia de intensidades permite marcar una frontera de tonalidades y encontrar los límites de estos objetos dentro de una imagen [52].

### 2.5.1. Operador de Sobel

Se tiene el gradiente en la posición del eje  $x$  y el gradiente en la posición del eje  $y$ , para cada técnica de segmentación se tiene preestablecidas las máscaras de convolución de acuerdo con la dirección de los gradientes en eje  $x$  y  $y$  [52].

Los bordes se obtienen por convolución de la imagen con las siguientes máscaras:

Máscara para gradiente en  $G_x$ :

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad (2.35)$$

Máscara para obtener el gradiente  $G_y$ :

$$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (2.36)$$

La siguiente matriz muestra las posiciones para aplicar las máscaras de convolución y obtener las derivadas.

$$\begin{bmatrix} z_1 & z_2 & z_3 \\ z_4 & z_5 & z_6 \\ z_7 & z_8 & z_9 \end{bmatrix} \quad (2.37)$$

De acuerdo con las posiciones y al aplicar la máscara se obtienen las siguientes ecuaciones:

$$G_x = (z_1 + 2z_4 + z_7) - (z_3 + 2z_6 + z_9) \quad (2.38)$$

$$Gy = (z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3) \quad (2.39)$$

donde  $z$  son los valores de los píxeles ubicados en las posiciones de la máscara, estos van cambiando conforme al barrido de la máscara por toda la imagen original.

Una vez que tenemos las nuevas matrices, una matriz por cada magnitud. Se aplica la siguiente ecuación:

$$G(F(x, y)) = |Gx(x, y)| + |Gy(x, y)| \quad (2.40)$$

Y aplicamos el siguiente umbral, donde  $U$  es un valor arbitrario, entero (dentro del intervalo 0-255) y positivo.

$$f(x, y) = \left\{ \begin{array}{ll} 0 & G(f(x, y)) < U \\ 255 & G(f(x, y)) \geq U \end{array} \right\} \quad (2.41)$$

## 2.5.2. Operador de Prewitt

El operador de Prewitt trabaja igual que Sobel, con diferentes máscaras [71]:

1. Máscara usada para obtener  $Gx$  en el punto central de una región de dimensión  $3 \times 3$

$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \quad (2.42)$$

2. Máscara usada para obtener  $Gy$ :

$$\begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad (2.43)$$

Una vez que tenemos las nuevas matrices, una matriz por cada magnitud. Se aplica la siguiente ecuación:

$$G(F(x, y)) = |Gx(x, y)| + |Gy(x, y)| \quad (2.44)$$

Y aplicamos el siguiente umbral, donde  $U$  es un valor arbitrario, entero (dentro del intervalo 0-255) y positivo.

$$f(x, y) = \begin{cases} 0 & G(f(x, y)) < U \\ 255 & G(f(x, y)) \geq U \end{cases} \quad (2.45)$$

### 2.5.3. Operador de Roberts

El operador de Roberts no contempla la dirección de las magnitudes, +únicamente opera mediante los vecinos próximos en diagonales del pixel central [11] [53]:

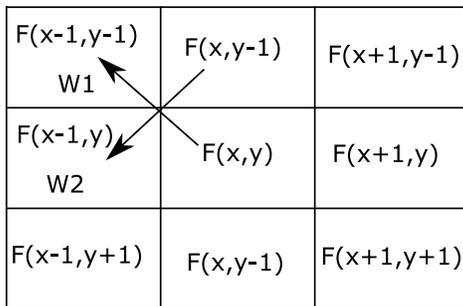


FIGURA 2.8: Posiciones para la Segmentación de Roberts

En la 2.8 se muestra las posiciones de los píxeles a tomar para aplicar el operador de Roberts e identificar los posibles bordes, en la imagen se visualiza las variables  $W1$  y  $W2$ , se definen por las siguientes ecuaciones:

$$W1 = f(x, y) - f(x - 1, y - 1) \quad (2.46)$$

$$W2 = f(x, y - 1) - f(x - 1, y) \quad (2.47)$$

Siguiendo el mismo procedimiento que el operador Prewitt y Sobel. Una vez que se tiene las matrices de  $W1$  y  $W2$  se realiza la suma de sus valores absolutos.

$$R = |W1| + |W2| \quad (2.48)$$

Y por último se aplica el umbral para la definición de 0 o 255 en cada pixel central, para obtener la imagen resultante.

$$f(x, y) = \begin{cases} 0 & G(f(x, y)) < U \\ 255 & G(f(x, y)) \geq U \end{cases} \quad (2.49)$$

#### 2.5.4. Método de Otsu

La segmentación de Otsu trabaja identificando una posible frontera de acuerdo con los valores de los pixeles promedio. Se tiene una imagen con  $N$  niveles de intensidad y el umbral optimo se define con  $T$  [17] [53]. La dinámica de Otsu es realizar recorridos de las intensidades, desde la intensidad mínima hasta la intensidad máxima y posteriormente de la intensidad máxima a la intensidad mínima. Primero se debe obtener a  $w1$  y  $w2$  que son la frecuencia de repetición de cada intensidad en la imagen de entrada [52].

$$w_1(t) = \sum_{z=1}^T P(z) \quad (2.50)$$

$$w_2(t) = \sum_{z=T+1}^L P(z) \quad (2.51)$$

Se calculan las medias y a partir de ellas las varianzas de intensidad mínima a máxima y la segunda varianza de la máxima intensidad a la mínima.

$$\mu_1(t) = \sum_{z=1}^T zP(z) \quad (2.52)$$

$$\mu_2(t) = \sum_{z=T+1}^L zP(z) \quad (2.53)$$

$$\sigma_1^2(t) = \sum_{z=1}^T (z - \mu_1(t))^2 \frac{P(z)}{w_1(t)} \quad (2.54)$$

$$\sigma_2^2(t) = \sum_{z=T+1}^L (z - \mu_2(t))^2 \frac{P(z)}{w_2(t)} \quad (2.55)$$

Finalmente se obtiene la varianza ponderada, para seleccionar las intensidades promedio con las que se hará la frontera optima.

$$\sigma_w^2(t) = w_1(t)\sigma_1^2(t) + w_2(t)\sigma_2^2(t) \quad (2.56)$$

Se selecciona el valor o valores de T, donde se tenga el valor mínimo en la varianza ponderada, donde T será la intensidad que se tomara como umbral. Se define la imagen bajo la siguiente ecuación:

$$f(x, y) = \begin{cases} 0 & G(f(x, y)) < U \\ 1 & G(f(x, y)) \geq U \end{cases} \quad (2.57)$$

## 2.6. Técnicas de Extracción y Reducción de Características

### 2.6.1. Local Binary Patterns (LBP)

Es un descriptor muy utilizado en sistemas de visión para la detección de objetos. LBP se considera un algoritmo invariante a los cambios de iluminación, ya que es

capaz de trabajar con cambios monotónicos en las intensidades. También es invariante a la traslación, esto se debe a que trabaja posición a posición tomando en cuenta los vecinos cercanos del píxel. El algoritmo LBP inicialmente se consideraba un descriptor de textura, pero actualmente se utiliza para variados aspectos en la visión por computador, la combinación de LBP con el algoritmo HOG ha sido de gran utilidad para la identificación de las formas [2].

Algoritmo básico LBP:

- 1.- Se trabaja con imagen en escala de grises, se coloca una máscara de tamaño 3x3 para la extracción del píxel central y sus vecinos cercanos.
- 2.- Se realiza el recorrido de una ventana, respetando la posición del píxel central en la imagen resultante de los LBP.
- 3.- Se toman los vecinos cercanos y se ordenan de forma arbitraria, ese orden se respeta para todos los píxeles de la imagen.
- 4.- Se coloca 1 o 0 según sea el caso aplicando la ecuación (2.58) donde  $p_c$  es el píxel central y  $xv$  los píxeles vecinos. , y se colocan conforme el orden ya asignado, para posteriormente indicar su valor de binario a decimal de acuerdo a la posición de un byte y se suman para obtener el patrón LBP [58].

$$J(p_c) = \left\{ \begin{array}{ll} 1 & x \geq C \\ 0 & xv < p_c \end{array} \right\} \quad (2.58)$$

- 5.- Con la imagen resultante LBP, se obtiene su histograma normalizado.

### 2.6.2. Histogramas de gradientes orientados (HOG)

HOG es un descriptor utilizado para la obtención de la estructura de las formas, extrae el contorno de los objetos indicando la magnitud y dirección de cada píxel. Divide las frecuencias de las magnitudes en intervalos, se consideran bloques de dirección y va creando un histograma del gradiente magnitud, donde cada gradiente es un cambio de tonalidad de la imagen en una cierta dirección. Para cada píxel se define la dirección donde el cambio de intensidad es máximo y la magnitud del cambio en esa dirección o orientación [21] [3].

Algoritmo básico HOG:

1.- Se calculan los gradientes en el eje x y y para cada píxel.

$$G_y = I(x, y + 1) - I(x, y - 1) \quad (2.59)$$

$$G_x = I(x + 1, y) - I(x - 1, y) \quad (2.60)$$

2.- Se tienen dos matrices, una por cada gradiente. Se calcula la magnitud y la orientación.

$$G = \sqrt{G_x^2 + G_y^2} \quad (2.61)$$

$$\phi(x, y) = \arctan \frac{G_y}{G_x} \quad (2.62)$$

3.- Se agrupan los píxeles de la imagen, de acuerdo con el tamaño de los bloques. Comúnmente estos bloques son de 6,8,16 o 32 píxeles que es la forma de dividir la imagen. Cada píxel tiene dos matrices, una matriz de magnitud y otra matriz de dirección.

4.- Para cada bloque se obtiene un histograma, donde  $x$  es la magnitud y  $y$  la dirección. Se define de forma arbitraria el número de barras del histograma, cada barra contiene un intervalo de la dirección. Un valor común de barras es de 9 y se determina si se contemplan solo valores de la dirección de  $0^\circ$  a  $180^\circ$ , o de  $0^\circ$  a  $360^\circ$ , por lo tanto, se divide 180 entre 9 barras o 360 si es el caso. Si se consideran solo  $180^\circ$ , cada barra tendría un intervalo de  $20^\circ$  y si se tuviera  $360^\circ$  cada barra tendría  $40^\circ$  [56].

Mediante la siguiente ecuación se asigna las magnitudes de acuerdo con el valor de la dirección:

$$h(k) = \sum w_k(x, y)G(x, y) \quad (2.63)$$

$$w_k(x, y) = \left\{ \begin{array}{ll} 1 & \text{si } (k - 1)(\phi(x, y)) \leq \phi(x, y) < k(\phi(x, y)) \\ 0 & \text{caso\_contrario} \end{array} \right\} \quad (2.64)$$

5.- Normalización del histograma:

$$N' = \frac{N}{\sqrt{\|N\|_2^2 + \varepsilon}} \quad (2.65)$$

$$\|N\|_2 = \sum x_i^2 \quad (2.66)$$

6.- Se concatenan todos los histogramas de los bloques para crear un histograma final normalizado [49].

### 2.6.3. Análisis de componentes principales (PCA)

Una imagen puede ser expresada solo por un vector, el cual contiene los componentes principales. Esta técnica fue principalmente desarrollada para el campo de la estadística, pero con el tiempo se descubrió que puede ser utilizada en el campo de visión por computador. Esta técnica extrae los rasgos principales y característicos, y reduce la información necesaria para identificar la variable que se está analizando, y todos estos rasgos principales son convertidos en un vector. Con este vector se crea la EingeFace, la cual es la representación del vector que contiene los rasgos principales.

PCA nos permite realizar un análisis de los componentes principales en conjuntos de características, el cual se encarga de obtener datos similares, en donde la pérdida de la información sea mínima. Los objetos pueden tener  $n$  características, y por lo tanto tienen  $n$  dimensiones, en las cuales se pueden representar. Cuando tenemos dos características es fácil realizar el análisis e identificarla, cuando se va aumentando el número, a un punto de tener  $n$  características, realizar el análisis lleva más tiempo y es más tedioso. En el análisis de componentes principales se encarga de recoger el mayor número de porcentaje de variabilidad y obtener la mayor precisión [79]. Esto lo hace mediante vectores que recorren el conjunto de datos y obtienen los componentes, buscando distancia entre puntos y grupos de similitud, para representar la dirección que tiene el vector, al momento de representar sea lo más exacta a la original. En donde la representación es representada por el cuadrado de las variables de entrada y las componentes del vector [51]. PCA es un método con un enfoque "Based Learning" que, aplicado en rostros, nos permite resultados enfocados en las vistas frontales y verticales, se ha caracterizado por reducir el número de características en conjuntos amplios de datos, creando un modelo basado en distribución de los patrones de la cara y un conjunto de parámetros de distancia [12].

Los autores Turk y Pentland en [76] desarrollaron el método de eigenfaces, y lo implementaron en PCA. El trabajo consistía en extraer los eigenfaces del rostro de

una persona y posteriormente utilizar los componentes principales para reconstruir el rostro a partir del vector de eigenfaces. 40 eigenfaces son suficientes para obtener una buena descripción del rostro de la persona con un mínimo de errores [40].

Análisis de Componentes Principales (PCA) es una técnica tradicional de proyección para reconocimiento de rostros, útil para la reducción de características. El método de eigenfaces resulta ser ampliamente útil y con resultados muy concisos, además de ser una técnica sencilla de utilizar y aplicar.

Se realiza preprocesamiento en la imagen  $I$ , posteriormente se le resta a cada una de las imágenes de entrenamiento, obteniendo el conjunto de datos mostrado en la ecuación. Las imágenes obtenidas se normalizan en base de la alineación de la boca y ojos del individuo.

$$i_1, i_2, \dots, i_n \in I - I^T \quad (2.67)$$

Con la siguiente ecuación se calculan los componentes principales:

$$A = R^T (XX^T) R \quad (2.68)$$

Donde:

$A$  = Matriz diagonal de valores propios.  $R$  = Matriz de vectores propios ortonormales.  $XX^T$  es la matriz de la covarianza de las muestras de entrenamiento.

ICA (Independent Components Analysis): Es una herramienta de análisis cuyo objetivo es descomponer la imagen del rostro, en una combinación lineal a partir de combinaciones estas.

El número de observaciones  $M(1 \leq i \leq N)$  debe ser mayor o igual al número de fuentes originales  $M(1 \leq j \leq M)$ , En general se utiliza  $N = M$ , asumiendo cada  $X$ , es una combinación desconocida y diferente de vectores originales, ICA expande cada señal  $X_j$ , en una suma ponderada de vectores fuente.

### Obtención de los CP (Componentes Principales)

1.-Buscar las combinación lineal de las variables.

2.-Buscar el subespacio que mejor se ajuste, minimizando la distancia euclidiana de los puntos en el subespacio.

### Eigenfaces

Los eigenfaces son utilizados en el análisis de componentes principales (PCA) para el análisis de las imágenes. Cuando se proyecta la imagen de un rostro sobre el subespacio creado con PCA, la imagen no cambia mucho; pero cuando se proyectan imágenes de no-rostros, estas varían considerablemente. Para la identificación de los rostros se calcula la región a evaluar y el subespacio del encuadre de la cara. La distancia es menor para regiones con rostros, que aquellas en las que no se encuentre el rostro.

#### 2.6.4. Análisis discriminante lineal (LDA)

Este algoritmo lleva el subespacio de rostros a una dimensión menor en donde aumenta la separabilidad de las clases. El objetivo de este algoritmo es encontrar un sub-espacio que obtenga los discriminantes de las diferentes clases. En donde se calcula la matriz de dispersión entre las clases que son distintas (intergrupales) y la matriz de la misma clase (intragrupal).

Sean  $\bar{x}_j$  y  $S_j$  los vectores de medias y las matrices de covarianzas de cada uno de los  $g$  grupos y sea  $\bar{x}$  el vector de medias global del conjunto de entrenamiento.

Usando estos elementos se pueden definir dos matrices  $B$  y  $M$  que representan la variabilidad entre los grupos y la variabilidad dentro de los grupos respectivamente y que están dadas por:

$$B = \sum_{j=1}^g n_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})' \quad (2.69)$$

$$W = \sum_{j=1}^g (n_j - 1) S_j \quad (2.70)$$

### 2.6.5. Algoritmos Genéticos

Los algoritmos genéticos es una técnica basada en mecanismos de la evolución natural con la idea base de que sobrevive el que mejor este adaptado a su entorno. Combinan la supervivencia de los individuos más aptos de una población con operadores genéticos tomados de la naturaleza, o bien características para formar mecanismos robustos para posibles soluciones a un problema [39].

Sus etapas se dividen en explotación y exploración y su eficiencia depende de la combinación entre estas mismas. La selección de las características y la cruce pertenecen a la etapa de explotación y la mutación a la etapa de exploración [13].

En el siguiente diagrama se muestra el funcionamiento de un algoritmo genético:

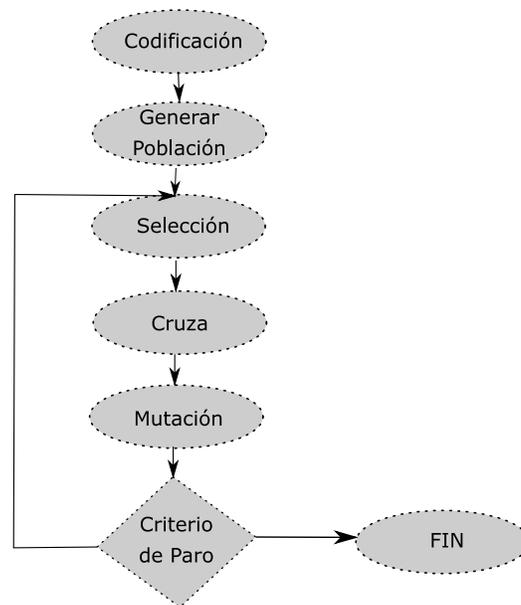


FIGURA 2.9: Diagrama general de un algoritmo genético

Partiendo de la codificación, primero identificamos la longitud de nuestras características o nuestros posibles datos en binario, como referencia nos apoyaremos en la siguiente ecuación:

$$2^{\beta} - 1 \leq m \leq 2^{\beta+1} \quad (2.71)$$

y para la decodificación utilizamos la siguiente ecuación:

$$x_i = x_i^l + \frac{x_i^A - x_i^l}{2^\beta - 1} \sum_{j=0}^{\beta} \sigma_j 2^j \quad (2.72)$$

- 1.- Para la generación de la población, tenemos n cantidad de datos sin importar el orden de los datos (son representados en binario-codificados).
- 2.- Para la selección se debe evaluar la actitud.
- 2.1- Primero debemos decodificar los datos, representados en decimal y aplicar la ecuación (2.72).
- 2.2- Aptitud, los valores resultantes del paso anterior los elevamos al cuadrado y evaluación en nuestra función. i.e.  $X^2$  y le restamos cada uno de los valores elevados al cuadrado, posteriormente realizamos la sumatoria de todos nuestros resultados para calcular los valores esperados.
- 2.3- Valor esperado, cada uno de los resultados posterior a la resta del mayor dato con las actitudes los estaremos dividiendo por el valor esperado.
- 3.- Selección, se tienen diferentes técnicas para la selección. Una técnica muy utilizada es la ruleta, esta se forma a partir de los resultantes del paso 2.3, colocando intervalos que inician desde cero al valor obtenido, tomando en cuenta que el intervalo siguiente es la suma del anterior y debe comenzar con una décima más para comenzar al intervalo. i.e. valor resultante del paso 2,3 es 0,090, 0,37, 0,32 y 0,21 el primer intervalo de la ruleta va de 0 – 0,090, 0,091 – 0,46, aquí el 0,46 ha salido de sumar 0,090+0,37, 0,461 – 0,78 y 0,781 – 1. Por lo tanto, nuestra ruleta se divide en 4 partes y el tamaño corresponde al intervalo establecido. Posteriormente generamos valores aleatorios para selección identificando en que parte de la ruleta será asignado, el siguiente valor aleatorio es acumulativo, se suma con el anterior y nuevamente identificamos en que parte de la ruleta se asigna.
- 4.- Cruza, existen diferentes métodos como: cruce de 1 punto, cruce de 2 puntos, cruce aleatoria, etc. En la siguiente Figura se representaron los diferentes métodos de cruce mencionados:

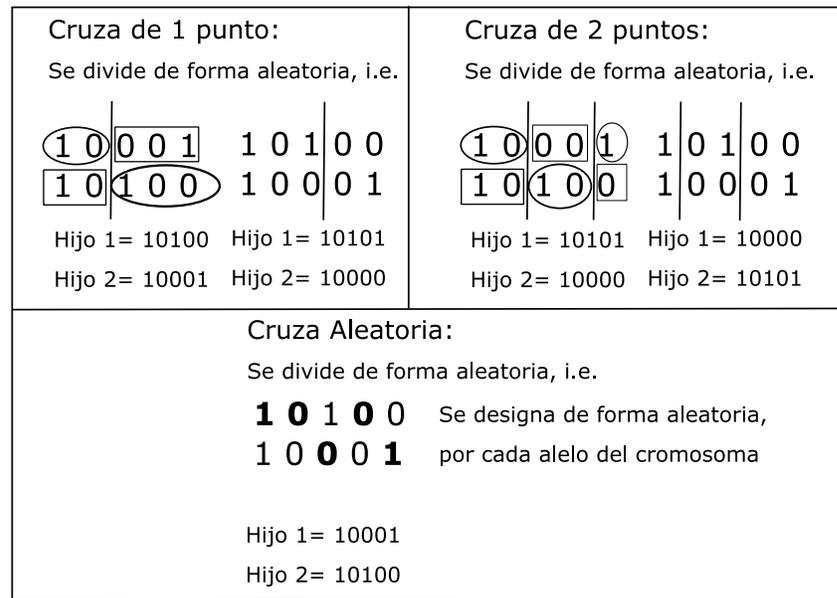


FIGURA 2.10: Métodos de cruce de un algoritmo genético

- 5.- Mutación, la mutación se coloca con un valor muy pequeño generalmente de 0,1. Esto dice que el 1% dirá que sí y el 99% dirá que no, si cae en si se cambia el alelo en la posición presente.

Para características el genético busca maximizar la precisión y toma las características de precisión más alta, por ello un algoritmo genético es de utilidad para la reducción de características, prevalecen las características más discriminantes [15] [35].

### 2.6.6. Características de Haralick

El cálculo para la identificación de textura en una imagen es uno de los procesos de extracción de características muy importante, debido a que muchos de los objetos son diferentes, únicamente por la textura. Por ejemplo, si realizamos un sistema que clasifique toronjas y mandarinas no solo podemos contemplar las características geométricas y de color, sino también de textura para una correcta identificación. De forma intuitiva podemos identificar texturas como suavidad, rugosidad, regularidad, etc. Algunos autores definen a la textura como la variación entre píxeles en una vecindad pequeña. Se considera una vecindad pequeña para identificar esos pequeños detalles que formar una textura como lo vemos en

una foto de un pasto (cambios ligeros por la acumulación de mucho pasto junto) [61].

El descriptor de Haralick utiliza como punto de partida la obtención de la matriz de coocurrencia de toda la imagen en escala de grises, debido a que todas las técnicas del descriptor Haralick para obtener texturas se basan en la matriz de coocurrencia.

La matriz de coocurrencia se obtiene de la siguiente forma:

- 1.- En escala de grises sabemos que el intervalo de posibles intensidades es de 0 a 255 sin embargo debemos revisar cual es la intensidad mínima y la intensidad máxima de la imagen. Por ejemplo, tenemos una imagen de  $N \times M$  con un intervalo de intensidad de 0 a 120 nuestra matriz de coocurrencia será de tamaño  $120 \times 120$ .
- 2.- Identificamos las posiciones de nuestra matriz nueva de referencia del paso anterior. Realizamos un barrido de nuestra imagen hacia la derecha y regresamos el barrido a la izquierda. Por ejemplo, primer posición es (1, 1) por lo tanto en la matriz de la imagen buscaremos cuantas veces se repite ese acomodo, buscamos 11 en la primera fila con una revisión corrida hacia la derecha, termina la fila y regresamos el barrido sobre la misma fila 1 en busca del 1 con adyacente 1. Realizamos el barrido sobre toda la imagen y obtenemos el total de veces que se repite ese acomodo que se obtuvo de la posición.
- 3.- Una vez obtenido todas las repeticiones tendremos una nueva matriz con esas repeticiones en la casilla de cada posición que se buscó. Esa sera nuestra matriz de coocurrencia.

Se define con la siguiente ecuación:

$$f(x, y) = P(i, j) \left\{ \begin{array}{l} L_x * L_y \\ L_y * L_x \end{array} \right\} \quad (2.73)$$

A partir de la matriz de coocurrencia podemos partir a otras técnicas como F1, F2, F3, contraste, correlación, entropía, varianza, coeficientes de correlación, etc.

Contraste:

$$\sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2 \quad (2.74)$$

Disimilitud:

$$\sum_{i,j=0}^{N-1} P_{i,j}|i-j| \quad (2.75)$$

Homogeneidad

$$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \quad (2.76)$$

Segundo momento angular (ASM) :

$$\sum_{i,j=0}^{N-1} P_{i,j}^2 \quad (2.77)$$

Energía:

$$E = \sqrt{ASM} \quad (2.78)$$

Máxima probabilidad:

$$MAX_{i,j}^N(P_{i,j}) \quad (2.79)$$

Entropía:

$$\sum_{i,j=0}^{N-1} P_{i,j}(-\ln P_{i,j})^2 \quad (2.80)$$

Matriz de coocurrencia Estadística Media:

$$\mu_i = \sum_{i,j=0}^{N-1} i(P_{i,j}) \quad (2.81)$$

$$\mu_j = \sum_{i,j=0}^{N-1} j(P_{i,j}) \quad (2.82)$$

Matriz de correlación con varianza:

$$\sigma_i^2 = \sum_{i,j=0}^{N-1} P_{i,j}(i - \mu_i)^2 \quad (2.83)$$

$$\sigma_j^2 = \sum_{i,j=0}^{N-1} P_{i,j}(j - \mu_j)^2 \quad (2.84)$$

Desviación estándar:

$$\sigma_i \sqrt{\sigma_i^2} \quad (2.85)$$

$$\sigma_j \sqrt{\sigma_j^2} \quad (2.86)$$

## 2.7. Técnicas de detección del rostro, microexpresiones y macro expresiones

La detección de rostros ha adquirido mucha popularidad en conjunto con otras técnicas para potentes algoritmos de inteligencia artificial para el desarrollo de grandes sistemas. A pesar de los diferentes algoritmos creados para la detección de rostros, el algoritmo Viola Jones sigue siendo uno de los algoritmos más rápidos y capaces para la detección facial [86].

### 2.7.1. Viola Jones

Viola Jones es un algoritmo dedicado a la identificación y reconocimiento del rostro dentro de una imagen. Está basado en técnicas de características de Haar, imagen integral y el clasificador AdaBoost en cascada. Utilizan las características de haar para la identificación de los cambios de tonalidad en los bordes de la imagen; Al tener grandes cantidades de estas características utilizamos la imagen integral

para realizar la transformación de cada píxel tomando en cuenta los píxeles de la izquierda y la derecha. Una vez que se tiene la información de la imagen, se clasifica con el clasificador adaboost para identificar si es un rostro [75].

### Características de Haar

Las características de Haar se obtienen a partir de aplicar sobre toda la imagen los filtros de Haar. Existen diferentes filtros de Haar y de diferentes dimensiones, su selección es arbitraria y se pueden clasificar para obtener información de las intensidades de los píxeles en horizontal y en vertical [6].

En la siguiente imagen se muestran los filtros básicos de Haar:

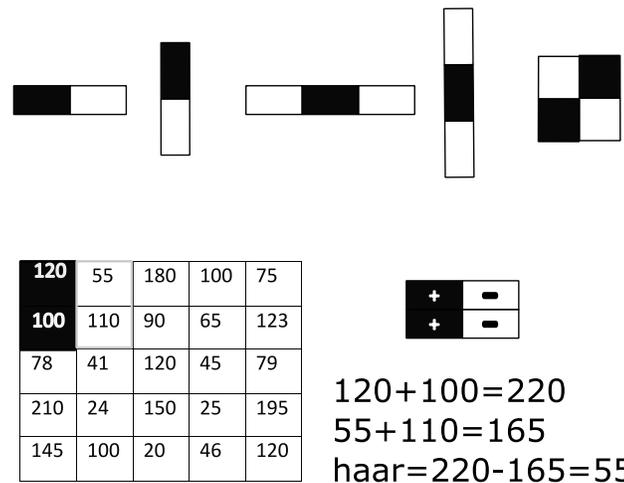


FIGURA 2.11: Filtros de Haar

Los filtros de Haar es la concatenación de varios rectángulos del mismo tamaño tanto para el área blanca y negra, y los filtros pueden ser de diferentes tamaños. Comienzan desde los más pequeños a los filtros más grandes. Los rectángulos en negro representan las zonas de contribución positiva y la parte blanca indica una contribución negativa. Los filtros de Haar se obtienen con la diferencia en la suma de los valores de la zona negra y la zona blanca como se muestra en la Figura 2.11 y cada uno de los resultados se va concatenando en un vector, el cual será el vector de características [59] [37].

## Imagen Integral

La cantidad de características que se obtienen a partir de los filtros de Haar es muy grande, debido a que tenemos filtros de diferente tamaño y posiciones de acuerdo con el tamaño de las imágenes, por ello es necesario aplicar la imagen integral. La imagen integral es una transformación de la imagen original de tamaño  $N \times M$  y nuestro resultado es una imagen del mismo tamaño, donde cada píxel es la suma de todos los píxeles en las posiciones izquierda y arriba de la imagen encerrando los datos en rectángulos, lo cual hace alusión a los filtros de Haar entre otros descriptores [20].

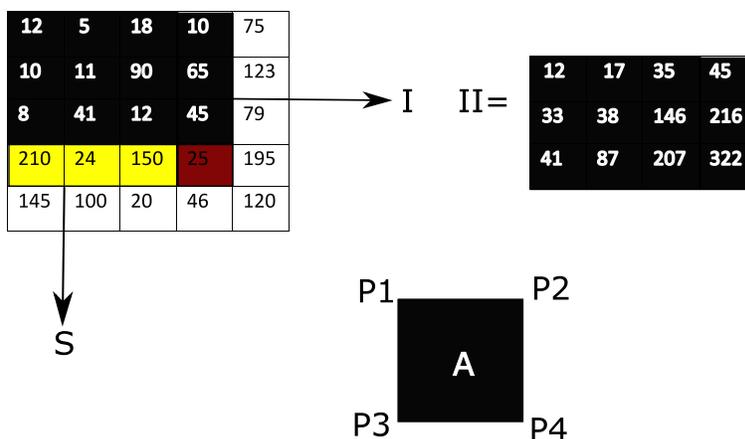


FIGURA 2.12: Imagen Integral

La imagen integral toma como valores acumulativos todo lo antecedente al píxel de interés, en la Figura 2.13 se realiza el cálculo realizando la sumatoria de las columnas del recuadro antecedente del píxel y posteriormente realizamos la sumatoria acumulativa en filas, por ello el recorrido se considera primero en horizontal y por último en vertical, las siguientes ecuaciones definen el cálculo de la imagen integral:

$$II(x, y) = II(x, y - 1) + S(x, y) \quad (2.87)$$

$$S(x, y) = \sum_{x' \leq x} I(x', y) = S(x - 1, y) + I(x, y) \quad (2.88)$$

La siguiente ecuación hace referencia a la Figura 2.13 donde se tienen los recuadros antecedentes a la zona de cálculo A, siendo cada recuadro ya un cálculo previo a la imagen integral y la concatenación de  $P1, P2, P3$  y  $P4$  en conjunto con A [27].

$$A = II(P4) - II(P2) - II(P3) + II(P1) \quad (2.89)$$

Podemos despejar si no tenemos el cálculo de una imagen integral P, y A si.

### Clasificador AdaBoost

Es un detector que trabaja en cascada, está basado en generar grandes conjuntos de características de Haar por toda la imagen. Adaboost es considerado eficiente para el proceso de aprendizaje con grandes cantidades de características, en comparación con otros clasificadores Adaboost no coloca su frontera de clases con una función paramétrica, si no como resultado de combinar un conjunto de clasificadores simples. Cada uno de sus clasificadores aprenden dando un peso diferente en cada ejemplo, para los datos mal clasificados se les asigna un peso mayor y para los datos bien clasificados se les da un peso menor como entrada a los clasificadores posteriores [84] [85].

Todas las fronteras que define cada clasificador se combinan de forma general, por lo tanto tendremos un trazo de fronteras que están dadas por los clasificadores que las van creando de acuerdo con los datos de mayor peso y de esta forma el clasificador Adaboost es capaz de encontrar una frontera que no es lineal, pero permite separar perfectamente los datos, los umbrales se van trazando de acuerdo con los pesos bajo la siguiente ecuación [81], donde  $\alpha = -1; +1$ :

$$h(x) = \{ \alpha f(x) < \theta \alpha f(x) \geq \theta \} \quad (2.90)$$

Actualización de pesos:

$$w_i(t+1) = \left\{ \begin{array}{ll} \frac{1}{\epsilon_t} \frac{w_i(t)}{2} & h(x_i) \neq y_i \\ \frac{1}{1-\epsilon_t} \frac{w_i(t)}{2} & h(x_i) = y_i \end{array} \right\} \quad (2.91)$$

$$\epsilon_t < 0,5 = \left\{ \begin{array}{l} \frac{1}{2\epsilon_t} > 1 \\ \frac{1}{2(1-\epsilon_t)} < 1 \end{array} \right\} \quad (2.92)$$

### 2.7.2. Detector de Objetos SIFT

Scale Invariant Feature Transform (SIFT), es un detector de puntos de interés. El algoritmo sift es invariante a escalas de las imágenes, lo que lo hace capaz de identificar objetos de diferente tamaño dentro de la imagen, con apoyo de un algoritmo de clasificación se puede obtener excelentes resultados para identificar los objetos y clasificarlos de acuerdo con la clase [50] [10].

SIFT trabaja consiste en aplicar filtro gaussiano, con un valor de  $\sigma$  en incremento para diferentes escalas y resoluciones. La siguiente ecuación define la función gaussiana:

$$g(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2.93)$$

Con el filtro gaussiano se realiza un suavizado, este suavizado se da con una combinación del pixel central con sus vecinos y la ponderación es de acuerdo con la distancia de cada vecino al punto central. Al realizar un suavizado se pierde la pureza de los píxeles, el suavizado es dado por el valor de  $\sigma$ , mientras más se incrementa el suavizado es mayor y por lo tanto mayor pérdida de los bordes [29].

Para el desarrollo del algoritmo SIFT se debe aplicar un incremento constante en  $\sigma$  para cada imagen, posteriormente se aplica la diferencia de las gaussianas conforme se obtienen  $A, B, C, D$  por ejemplo si tuviéramos esas imágenes cada una con el incremento de  $\sigma$ , realizamos la diferencia de  $A1 = A - B$  y  $A2 = C - D$ . Cuando hemos obtenido las diferencias se grafican los resultantes para la identificación de los máximos y mínimos a diferentes escalas, para cada una de las imágenes aplicamos realizamos una pirámide de imágenes [69].

En los máximos y mínimos las zonas más claras son todos los mínimos y las zonas oscuras los máximos. Para la selección de estos, debemos identificar cual fue

el mínimo total en todas las imágenes o el máximo de todas las imágenes re escaladas. Una vez identificado debemos comparar ese punto de interés con una imagen antes y una posterior para validar que si es un punto más alto o bajo de las tres en comparación.

El uso de las pirámides de las imágenes es para identificar objetos en diferentes tamaños, siendo esta la característica del algoritmo SIFT que sea invariante al tamaño de los objetos.

El algoritmo SIFT se describe a continuación:

- 1.- Aplicamos una convolución con filtro gaussiano en cada imagen, y aplicamos un incremento en el suavizado con  $\sigma$
- 2.- Realizamos la diferencia de gaussianas (imágenes con el incremento en  $\sigma$ )
- 3.- Detectamos los puntos de interés (máximos y mínimos locales) en las diferentes escalas.
- 4.- En ciclo con el punto 3 realizamos un re-escalado para cada imagen a la mitad en el eje x y en el eje y. Hasta el nivel mínimo posible.

### 2.7.3. Detector de Objetos SURF

Speeded Up Robust Features (SURF), es un detector de puntos de interés que trabaja de una forma muy similar a SIFT. A diferencia de SIFT, SURF aplica un filtro de una segunda derivada gaussiana, localizando los puntos de interés a partir de una matriz Hessiana, se define:

$$H(x, \sigma) = \begin{pmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{pmatrix} \quad (2.94)$$

La siguiente ecuación hace referencia a la imagen filtrada por la segunda derivada de una gaussiana de tamaño  $\sigma$ :

$$L_{xy}(x, \sigma) = I(x) \frac{\partial^2}{\partial xy} g(\sigma) \quad (2.95)$$

Para la identificación de los puntos de interés, SURF utiliza el determinante de la matriz Hessiana para identificar cuando se den cambios locales alrededor de

un punto  $x$ . Seleccionamos los puntos de interés que presenten un mayor determinante, dado que eso representa un cambio local mayor respecto a los demás cambios dentro de la imagen, esta interpretación en SIFT es la identificación de los mínimos y máximos [9].

Previo a la aplicación de la segunda derivada a las imágenes se recomienda pasar a binaria para un incremento en la definición de los determinantes. En SURF posterior a la aplicación de la Hessiana aplicación el uso de las imágenes integrales, donde el valor de un píxel en la posición  $(x,y)$  es la suma de todos los valores antecedentes en el rectángulo definido por el origen  $O(x)$ , como se muestra en la siguiente Figura:

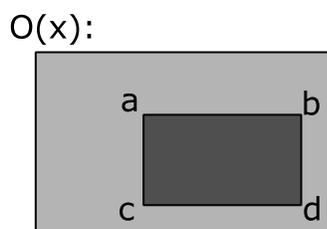


FIGURA 2.13: Imagen Integral SURF

Se aplica la ecuación (2.89) en posiciones de las nuevas variables. El descriptor SURF es invariante a la rotación y a los cambios de escala, trabaja con la técnica de pirámide de imágenes para dar solución a las diferentes escalas de los objetos. Para lograr ser invariante a la rotación, SURF busca la orientación de los puntos de interés, utilizando la técnica de ondícula de Haar para los ejes  $x,y$  dentro de una vecindad en forma de círculo, de esta forma se asegura de manipular el entorno de todo alrededor del punto de interés [8].

#### 2.7.4. Análisis discriminantes lineal (LDA)

Análisis Discriminante Linear (LDA), esta técnica es una mejora de la técnica de PCA, la cual analiza la información, con el objetivo de reducirla la dimensión de

las características, aplicando LDA, donde la matriz de proyección sea máxima, la diferencia que tiene LDA con PCA, es que arroja mejores resultados en condiciones de de iluminación y gestos.

Otro método es Análisis de Componentes Principales LDA (Linear Discriminant Analysis), es una aproximación estadística, la cual maximiza entre usuarios y minimiza la varianza de cada usuario. Es una técnica de aprendizaje supervisado para clasificar datos. La idea central de LDA es obtener una proyección de los datos es un espacio de menor (o incluso mayor) dimensión que los datos entrantes, con el fin de la separabilidad sea la mejor posible.

Utiliza una aproximación estadística para diferenciar las muestras conocidas de los que no lo son. Sean  $x_{ij}$  si los vectores de medias y las matrices de covarianza de cada uno de los grupos  $g$  y sea  $x$  el vector de medias global del conjunto de entrenamiento.

El análisis discriminante es una técnica multivariante cuya finalidad es analizar si existen diferencias significativas entre grupos respecto a un conjunto de variables sobre lo mismo, para el caso en que existan. Se considera como un análisis de regresiones donde la variable dependiente es categórica y tiene como categorías la etiqueta de cada uno de los grupos, mientras que las variables independientes son continuas y determinan a qué grupos pertenecen.

El objetivo de la técnica LDA es encontrar un vector  $a \in \mathbb{R}^p$  de tal manera que se maximice el cociente  $\Lambda$  definida en la siguiente ecuación.

$$\Lambda = \frac{a' B a}{A' W a} \quad (2.96)$$

Así se encuentra un hiperplano que genera la máxima diferencia entre la variabilidad.

Sean  $\bar{x}$  y  $S_j$  los vectores de medidas y las matrices de covarianza de cada uno de los grupos y sea  $\bar{x}$  el vector de medidas global del conjunto de entrenamiento. Usando estos elementos se pueden definir dos matrices  $B$  y  $W$  que representan variabilidad entre los grupos y la variabilidad dentro de los grupos respectivamente.

$$b = \sum_{j=1}^g n_j (\bar{x}_j - x)(\bar{x}_j - \bar{x})' \quad (2.97)$$

El objetivo de la técnica de LDA es encontrar un vector  $a \in \mathbb{R}^p$  de tal manera que no se maximice el cociente  $\Lambda$  definido por:

$$\Lambda = \frac{a^T B a}{A^T W a} \quad (2.98)$$

## Capítulo 3

# Metodología

La identificación de una emoción en la vida diaria se considera ser sencillo cuando tenemos aspectos discriminativos del rostro de una persona, de forma visual identificamos un posible estado de ánimo, generalmente estos estados de ánimo se clasifican en siete emociones básicas, de acuerdo con investigaciones del psicólogo Ekman, el definió las siguientes emociones: asco, desprecio, miedo, sorpresa, alegría, ira y tristeza. Debido a que si tenemos una imagen estática de una persona podemos identificar de forma intuitiva la emoción. Sin embargo a pesar de que se considere posible, se puede tener un margen de error alto, debido a que el ser humano representa una emoción con varias microexpresiones, que son pequeños cambios en el rostro con una duración de un cuarto de segundo, estas microexpresiones son cruciales para identificar a una emoción con un margen de error mínimo, aunado a ello las emociones se representan con algún cambio de acuerdo a la persona, todos estos cambios están predispuestos al lugar de donde vive esta persona, la cultura que tiene y creencias. Por ejemplo, una persona de Japón demuestra menos una emoción que un latino y todo esto debido a la cultura, el aprendizaje, entre otros aspectos.

En este Capítulo se muestran las diferentes metodologías que se han trabajado para el reconocimiento de las emociones, explicando las técnicas que se han utilizado, y cuáles de ellas nos han sido de mayor utilidad para obtener un margen de error menor que otras. En base a todas las pruebas que se han realizado, se explica el desarrollo de las metodologías provechosas a las que se han llegado para esta investigación.

### 3.0.1. Conjuntos de Datos

Uno de nuestros principales retos fue el obtener los conjuntos de datos para nuestra investigación, se pretendió crear un conjunto de datos propio, pero debido a problemas de pandemia se interrumpió este proceso. Por ello hemos decidido trabajar con los siguientes conjuntos de datos:

#### SMIC database

SMIC (SMIC-Base de datos de microexpresiones espontáneas) incluye microexpresiones espontáneas provocadas por clips de películas emocionales; Se mostraron a los participantes clips que pudieran inducir a reacciones emocionales fuertes. SMIC contiene 164 microexpresiones espontáneas de 16 personas, los datos se registraron con una cámara de alta velocidad (HS) de 100 fps. La Universidad de OULU en conjunto con el Centro de Visión Artificial y Análisis de Señales (CMVS) crearon el conjunto de datos de microexpresiones SMIC, este conjunto de datos maneja cinco emociones (feliz, sorpresa, disgusto, tristeza y asco) con imágenes tamaño  $186 \times 227$  píxeles [46] [57] [45].

#### SAMM database

El segundo dataset es SAMM (Acciones espontáneas y micro movimientos) el conjunto de datos contiene 159 movimientos micro faciales espontáneos obtenidos a través de inducción emocional, se recopilaron micro movimientos de 32 participantes de un grupo demográfico diverso que incluye 13 etnias diferentes y una edad media entre 24 y 33 años con 17 hombres y 16 mujeres. La Universidad Metropolitana de Manchester (MMU), que en conjunto con la Academia de Inteligencia Emocional realizaron el conjunto de datos de microexpresiones SAMM, este conjunto de datos maneja seis emociones (feliz, sorpresa, asco, disgusto, tristeza y miedo) con imágenes tamaño  $960 \times 650$  píxeles [22] [19] [45].

Estos conjuntos de datos se utilizaron debido a que son enfocados a microexpresiones para el reconocimiento de las emociones, lo cual nos permitieron realizar una investigación amplia respecto a expresiones faciales con macro expresiones, microexpresiones y movimientos oculares, este último debido a que puede ser identificado a una emoción debido a los clips de microexpresiones.

### 3.1. Metodología I

El reconocimiento de emociones básicas, las cuales son alegría, ira, sorpresa, tristeza, asco, miedo y desprecio. Se tiene como primer trabajo, el uso de expresiones faciales para identificar las emociones básicas, basando nuestra investigación en ambos conjuntos de datos SMIC y SAMM.

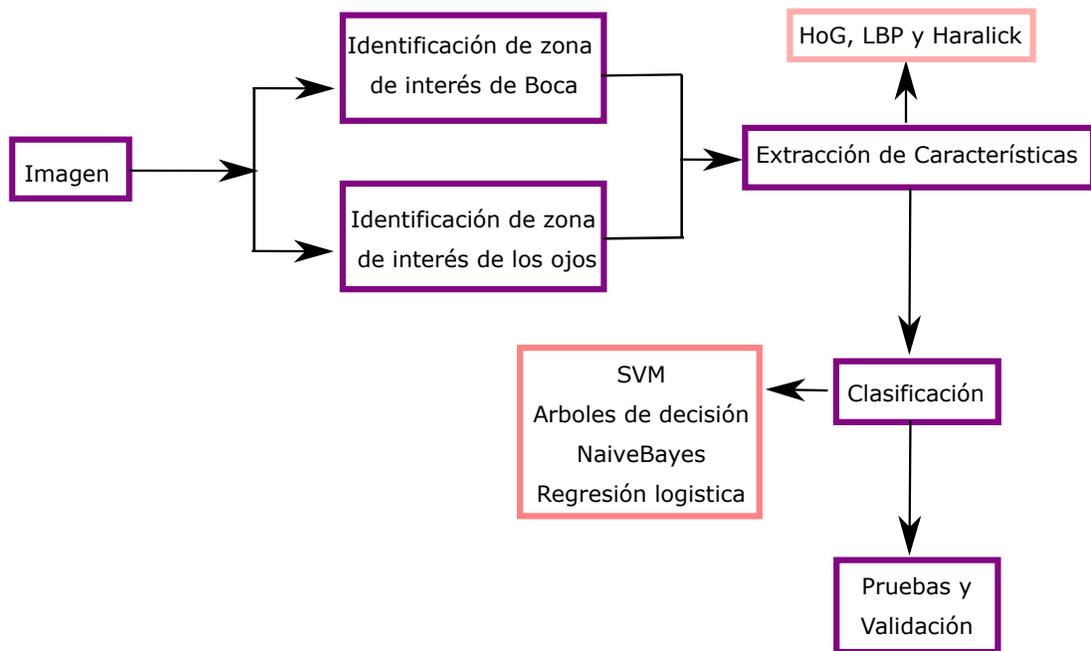


FIGURA 3.1: Metodología I. Expresiones faciales (macro expresiones)

En la Figura 3.1 se muestra un diagrama de bloques de la primera metodología, explicando de forma general en esta metodología trabajamos con imágenes en escala de grises, tomamos como zonas de interés la boca y los ojos, de las zonas de interés se realizó la extracción de características texturales de Haralick, Histogramas de gradientes ordenados (HoG) y Local binary patterns (LBP). Para la clasificación utilizamos una máquina de soporte vectorial (SVM), arboles de decisión, clasificación bayesiana (Naive Bayes) y regresión logística de los cuales obtuvimos muy buenas precisiones en SVM y regresión logística.

### 3.1.1. Selección de zonas de interés

Las emociones se muestran principalmente en el rostro, a pesar de que se complementan con la voz y los movimientos corporales. Con la imagen del rostro podemos determinar que emoción se está presentando en la persona. El rostro representa una forma de expresión con las microexpresiones ante cualquier situación, incluso todos podemos afirmar que las expresiones de nuestro rostro suelen ser involuntarias, debido a que las microexpresiones son ligeros movimientos no conscientes. Nosotros de entrada siempre manifestamos las emociones en el rostro antes que con el cuerpo y estos sentimientos se determinan por los movimientos que hacemos con los ojos, cejas, nariz y boca. Debido a que la zona de la nariz es dependiente a los movimientos que hacemos con la boca y la zona de las cejas no nos arroja la misma información discriminativa que los ojos, además que podemos identificar si hay movimientos en las cejas con los movimientos que se tengan en la zona de los ojos. Por lo que hemos determinado la zona de los ojos y de la boca como las zonas más discriminativas para identificar una emoción en el rostro.

Se realizaron algunas pruebas para la correcta identificación de estas zonas de interés:

Prueba 1: Se realizó el entrenamiento de un clasificador en cascada con imágenes positivas de las zonas de interés de boca y ojos, también se recolectaron imágenes negativas en contraste a estas zonas como fue cabello, orejas, frente, mejillas y barba. No fue necesario implementar imágenes negativas de otros objetos debido a que al ingresar una imagen nueva se realizaba la detección del rostro con el algoritmo Viola Jones. Esta prueba se realizó con 1000 imágenes positivas y 1000 negativas, con el 20 % en escala de grises y las restantes a color.

TABLA 3.1: Tabla de errores. Prueba del Modelo entrenado

Etapas Cascada/Error	001	000001	0000000000000001	000000000009
10	14	14	14	14
50	13	6	12	12
100	14	15	15	15
150	8	8	8	10
200	13	9	9	12
250	15	13	13	13
300	15	13	13	15
500	13	13	13	13
5000	6	6	6	6
10000	6	6	6	6

En el cuadro 3.1 se muestra el comportamiento del modelo basándonos en la cantidad de errores de acuerdo con la cantidad de etapas de entrenamiento del clasificador en cascada y el mínimo margen de error aceptado. Se calcularon con 22 imágenes de prueba, en escala de grises. En el modelo obtuvimos problemas debido a que en la zona de los ojos algunas veces encuadraba los orificios de la nariz, problemas para enfocar los ojos cuando utilizan lentes con armazón grueso, personas con lentes de armazón delgado no se tuvo problemas. Y finalmente se tenía mejor clasificación con imágenes de color que en escala de grises.

El desempeño impredecible del modelo nos ayudó a comprender que deficiencias se debían mejorar, incrementando el número de positivos y negativos para el entrenamiento y muy importante enfocarnos a una escala de grises debido a que ambos de los dataset con los que trabajamos son en escala de grises.

Finalmente se realizó el entrenamiento de un clasificador en cascada, utilizando el algoritmo de Viola Jones para el encuadre del rostro con las características de Haar, se realizó el entrenamiento del clasificador con imágenes positivas y negativas, posteriormente la extracción de características de HoG y LBP para la mejora en la identificación de las zonas de interés.

### 3.1.2. Extracción de características

Para los conjuntos de datos SAMM y SMIC con los que hemos trabajado en esta investigación no fue necesario realizar un preprocesamiento previo, debido a que las imágenes vienen en escala de grises y con buena resolución. Los descriptores que se utilizaron fueron histogramas de gradientes ordenados (HoG), local binary Patterns (LBP) y características texturales de Haralick.

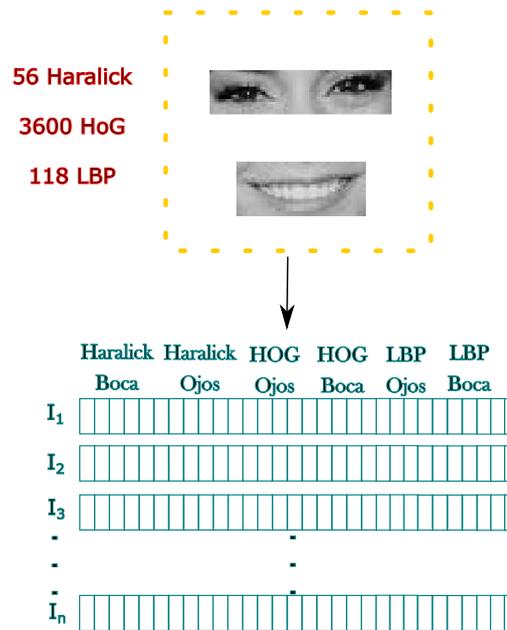


FIGURA 3.2: Extracción de Características

En la Figura 3.2 se muestra la forma en cómo se realizó la extracción de las características con los descriptores HoG, LBP y Haralick. Realizamos la identificación de las zonas de interés en cada una de las imágenes y aplicamos un recorte en estas zonas de interés para realizar la extracción de características. Se puede observar en la Figura 3.2 que, por cada una de las imágenes del rostro, obtuvimos dos imágenes que pertenecen a las zonas de interés, una imagen de los ojos y otra de la boca. A cada una de las imágenes de la boca y de los ojos se realizó la extracción de características por lo que obtuvimos un vector de 3775 columnas, que nos indican el total de las características, se puede observar que el vector está dado por 56 características de Haralick, 3600 de HoG y 118 de LBP, este vector es la concatenación de las características de ambas zonas de interés por cada una de las imágenes de los dataset.

La explicación del uso de los descriptores ya se ha realizado en el Capítulo 2.

### 3.1.3. Clasificación

Para la clasificación se aplicaron cuatro tipos de clasificadores, uno de ellos es Máquina de Soporte Vectorial (SVM) que es muy utilizado en las investigaciones que se tienen de identificación de emociones. Los otros clasificadores son clasificación bayesiana (NaiveBayes), arboles de decisión y regresión logística.

Para encontrar los parámetros óptimos de los clasificadores, se implementó la técnica de búsqueda de malla, se acotó la búsqueda de los valores mediante una división del rango de valores. Una vez que se encontraron los mejores valores, se realiza una nueva subdivisión en el mejor rango encontrado, hasta que ya no exista una variación significativa en la precisión del clasificador.

La explicación de cada uno de los clasificadores se ha realizado en el Capítulo 2.

### 3.1.4. Validación

Para validar los resultados de los clasificadores se utilizó una validación cruzada con el parámetro  $k=10$ . En el capítulo 4 se analizan los resultados obtenidos de acuerdo con cada métrica de validación.

## 3.2. Metodología II

En esta metodología a diferencia de la primera se utiliza un algoritmo genético para la reducción de características, eliminando las menos discriminativas. Se trabajó con la zona de interés de los ojos y la boca, utilizando las técnicas de extracción de características de HoG, LBP y texturales de Haralick. En la siguiente figura se muestra un diagrama de bloques de la metodología:

### 3.2.1. Selección de las características con Algoritmo Genético

Realizamos la extracción de las características de las zonas de interés y obtuvimos nuestro vector de características definido en una matriz de tamaño  $M \times 3824$

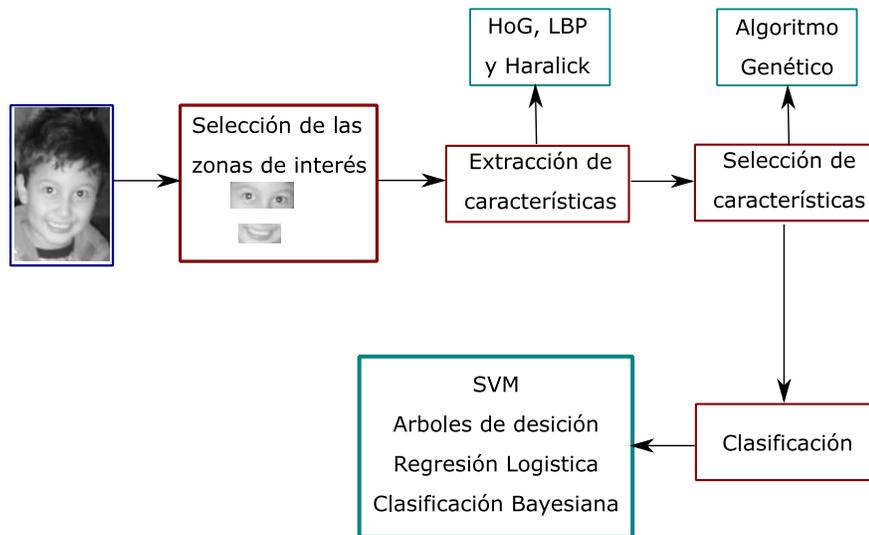


FIGURA 3.3: Metodología II. Identificación de emociones y reducción de características (Algoritmo Genético)

con  $M$  cantidad de imágenes y 3824 características. En cada clasificador seleccionamos las mejores características de acuerdo con la precisión. Nuestro vector de características contiene los datos de entrada para el algoritmo genético, cabe mencionar que todas las características obtenidas se normalizaron con una media cero y una desviación estándar igual a 1. El algoritmo genético parte de una población aleatoria de subconjuntos de características (llamados cromosomas). Cada uno de los subconjuntos o cromosomas se evalúa midiendo su capacidad de predecir las etiquetas, lo cual se representa en función de la precisión de cada conjunto. El funcionamiento del genético es ir seleccionando las mejores características de acuerdo con su precisión, por lo tanto, reemplaza la población tomando las más altas y combina con otras características de otro subconjunto con mayor precisión en clasificación. Se detiene hasta alcanzar la precisión más alta. En cada iteración se crea una nueva población mejorada y recombina mediante la selección, mutación y la cruce. Como se explica en el Capítulo 2.

Para la codificación utilizamos una codificación binaria, que simboliza la ausencia o presencia de una característica. Con un tamaño de 30 en cada subconjunto de la población, se realizó un total de 100 generaciones, para la selección utilizamos la técnica de la ruleta y en la cruce utilizamos la técnica de 2 puntos con una probabilidad de 0,7 y un en la mutación manejamos una probabilidad del 0.02 para el cambio de cada bit en los cromosomas.

### 3.3. Metodología III

En las anteriores metodologías se trabajó con las zonas de interés de los ojos y la boca, debido a que representan gran cantidad de información para identificar con buena precisión las emociones. Sin embargo, en esta metodología se realizó un análisis con todas las zonas discriminativas por separado y el rostro completo, con el fin de dar a conocer cuál es la precisión de cada zona, y que zonas de rostro representan más información para la identificación de las emociones. En el siguiente diagrama de bloques se muestra la metodología utilizada:

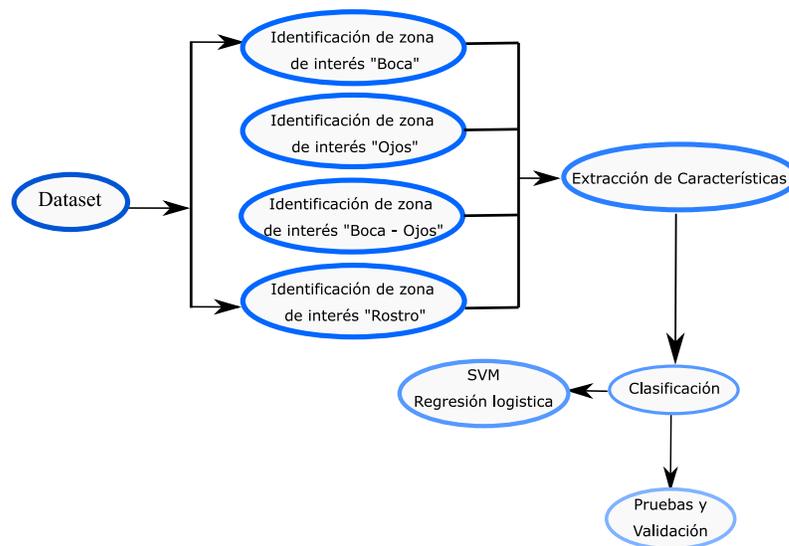


FIGURA 3.4: Metodología III. Zonas discriminativas del rostro para identificación de emociones

En la Figura 6.2 se muestra un diagrama de bloques de la tercer metodología, donde se representa la clasificación de las zonas discriminativas del rostro, las cuales son: la zona de la boca, la zona de los ojos-cejas, boca-ojos y por último el rostro limitado por el encuadre del mismo. Para cada zona se realiza la extracción de características con los descriptores LBP, HoG y las texturales de Haralick.

Posteriormente se aplican las técnicas de clasificación SVM, Regresión Logística, arboles de decisión y clasificación bayesiana, finalmente se realizó la validación de estos.

### 3.3.1. Selección de zonas de interés

El análisis de las diferentes partes del rostro fue nuestro objetivo de estudio, debido al interés de determinar que zonas eran más discriminativas para un mejor precisión y menor tiempo para la identificación de las emociones. Tomando al tiempo como una variable debido a que entre más zonas del rostro analicemos más tiempo nos consume. La selección de las diferentes zonas del rostro lo determinamos con el sistema de codificación de acción facial (FACS), determinamos que las emociones se muestran principalmente en la zona de los ojos, boca, cejas y nariz. Sin embargo, contemplamos también la zona del rostro completo para realizar una comparativa, para la obtención de la zona del rostro completo se utilizó el algoritmo de Viola Jones y para las zonas de los ojos se utilizó la ecuación (3.1) y para la boca se obtuvo con la ecuación (3.2).

$$In = I(f/(f/100), c/(c/100), 250, 100) \quad (3.1)$$

$$In = I(f/(f/100), (c/(c/100)) * 3, 250, 100) \quad (3.2)$$

### 3.3.2. Extracción de características

La extracción de las características se obtuvieron como en la metodología I, con las técnicas de LBP, HoG y Haralick. Por cada zona de interés se realizó la extracción de las características de textura y geométricas. Las características geométricas se obtuvieron con HoG y LBP y las Texturales con Haralick. Se obtuvo un vector de características por cada imagen con un tamaño de 1887 características y una clase. El orden del vector de características es primero 28 características de Haralick, HoG con 1800 características y 59 de LBP, para HoG se tomaron 9 barras de 20 x 20 con un desplazamiento de 10.

Cada una de las pruebas por zona de interés se realizó de forma independiente, se obtuvo un vector de características para la zona de la boca, otro vector para

la zona de los ojos, otro vector para la zona del encuadre del rostro delimitado y finalmente otro vector de características de la combinación de los ojos y la boca.



# Capítulo 4

## Resultados Experimentales

En este Capítulo se muestran los resultados obtenidos con los diferentes algoritmos propuestos.

### 4.0.1. Resultados Metodología I

El desarrollo del proyecto se realizó utilizando el lenguaje de Matlab en apoyo del software Weka para estimar las precisiones de los clasificadores utilizados.

En la Tabla 4.1 se muestran los desempeños obtenidos con SVM para cada una de las emociones. Los parámetros utilizados fueron Kernel Gaussiano con  $C=1000$  y  $G=0.005$ . En todos los experimentos se utilizó validación cruzada y búsqueda de malla para obtener los parámetros óptimos para cada clasificador. La Tabla muestra que todas las clases son identificadas con una alta precisión.

TABLA 4.1: Resultados del clasificador SVM para conjunto de datos SAMM

Clase	TPR	Precisión	Recall	F-measure	MCC	AUC-ROC
Alegría	0.993	0.994	0.994	0.992	0.998	0.993
Sorpresa	0.997	0.995	0.996	0.994	0.998	0.993
Ira	1.000	0.999	0.999	0.999	1.000	0.999
Disgusto	0.998	1.000	0.999	0.999	1.000	0.999
Tristeza	1.000	1.000	1.000	1.000	1.000	1.000
Asco	0.993	0.995	0.994	0.994	0.999	0.992

Los resultados obtenidos por el clasificador SVM para el conjunto de datos SAMM con imágenes con un tamaño de  $960 \times 650$  nos muestran una mayor precisión en

las emociones de disgusto, tristeza y ira siendo las más simples de diferenciar debido a sus diferencias y la precisión más baja fue la emoción de la alegría debido a su semejanza con las emociones de sorpresa. En comparativa con la Tabla 4.2 del conjunto de datos SMIC con un tamaño de  $186 \times 227$  muestran que la mejor precisión son para las emociones tristeza y disgusto al igual que el conjunto SAMM pero aquí nos arroja mejor precisión para la emoción alegría en comparación de la emoción asco esto se refleja debido a la reducción de las imágenes para este conjunto donde se aprecia más el simple hecho de sonreír que los minuciosos movimientos de la boca para la emoción del asco que representa movimientos en la área de los ojos que se puede perder con la emoción de sorpresa, felicidad y disgusto.

TABLA 4.2: Resultados del clasificador SVM para conjunto de datos SMIC

Clase	TPR	Precisión	Recall	F-measure	MCC	AUC-ROC
Alegría	0.980	0.972	0.976	0.968	0.990	0.963
Sorpresa	0.981	0.984	0.982	0.975	0.994	0.977
Disgusto	0.971	0.993	0.982	0.979	0.996	0.978
Tristeza	0.991	0.996	0.994	0.993	0.998	0.991
Asco	0.978	0.965	0.972	0.965	0.990	0.953

Matrices de confusión del conjunto de datos SMIC, resaltando la precisión por cada clase (emoción):

En la Figura 4.1 se muestra una relación entre las emociones, reflejándose en las malas clasificaciones con tendencia a otra emoción por su semejanza. Realizando un análisis contemplando los dos mejores clasificadores que son las SVM y la regresión logística. Para la emoción de la felicidad se tiene una semejanza con la emoción de la sorpresa y un ligero rasgo con la emoción de tristeza y el asco, para la emoción de la sorpresa tiene semejanza con la emoción la felicidad y un poco con la emoción del asco, para la emoción del disgusto tiene características semejantes con el asco y ligeramente con la felicidad, la emoción de la tristeza tiene semejanza con la sorpresa y finalmente el asco tiene semejanza con la sorpresa y ligeramente con la felicidad y el disgusto. Las semejanzas entre las emociones nos indica que comparten características entre ellas, mismas que representan una mínima tasa de error en la precisión de los clasificadores.

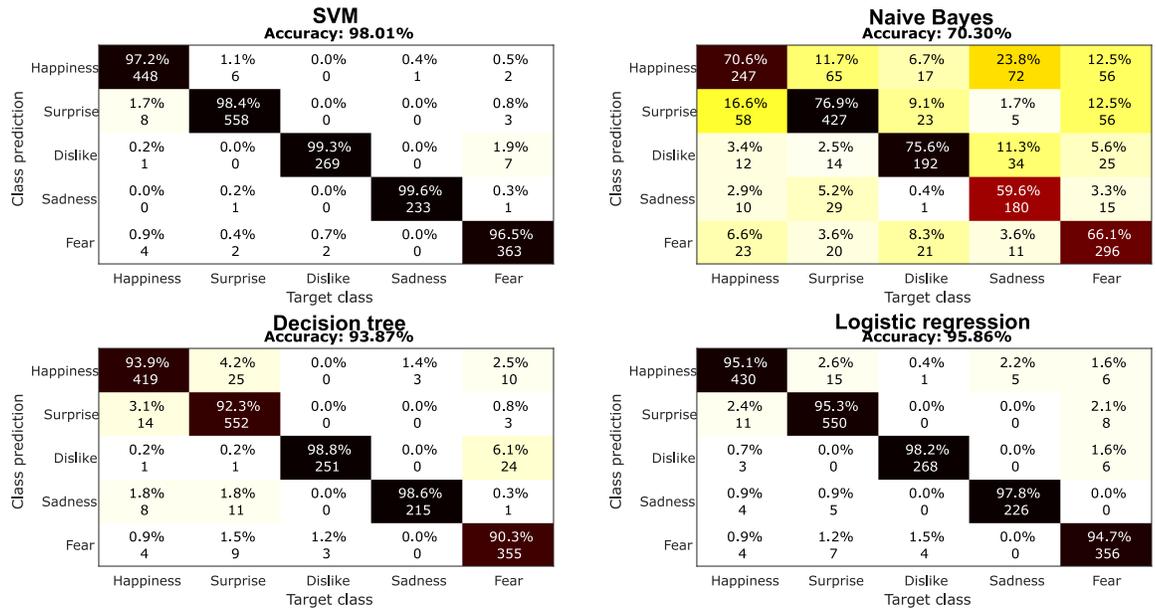


FIGURA 4.1: Matrices de confusión SMIC

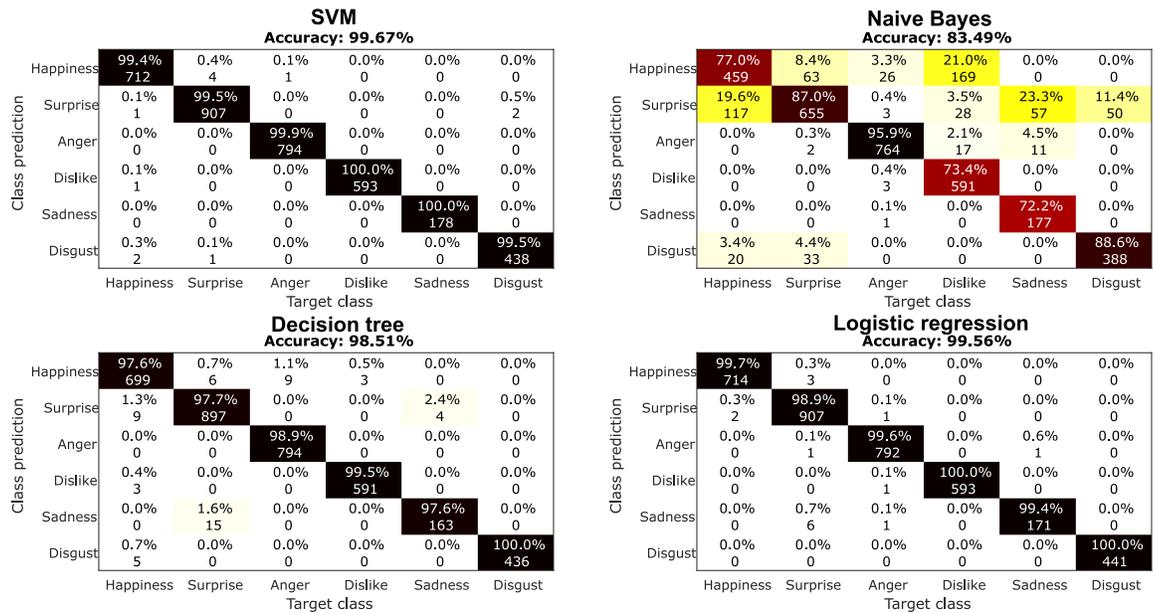


FIGURA 4.2: Matrices de confusión SAMM

Matrices de confusión del conjunto de datos SAMM:

En la Figura 4.1 se refleja los aciertos y las fallas de nuestros clasificadores para la clasificación de cada emoción, las tasas de errores son muy mínimas sin embargo, nos muestran que las emociones al ser muy semejantes pueden causar ruido al momento de identificarlas, por ello se deben contemplar las características más representativas de cada emoción. Nuestros clasificadores arrojaron precisiones muy favorables para todas las emociones pero se destaca aquí la resolución de las imágenes debido a que nuestros dos conjuntos de datos tienen tamaños muy distintos, lo que hace que nuestras resoluciones para el rostro sea con mayor representación en el conjunto de datos SAMM.

Rescatando los resultados de los clasificadores que nos arrojaron la mejor precisión, SVM y regresión logística. Nos muestran una menor cantidad en la tasa de error, por lo tanto tenemos muy pocas emociones mal clasificadas. Indica una ligera semejanza entre la emoción de la felicidad con la sorpresa; al igual que el conjunto de SMIC, con la emoción sorpresa se muestra una semejanza con felicidad y enojo, y el disgusto con la felicidad y el enojo; esto debido a la semejanza en las misuras de los movimientos de la boca.

#### 4.0.2. Resultados Metodología II

En la Tabla 4.3 se muestran las precisiones obtenidas para cada conjunto de datos, para cada clasificador se utilizó la métrica de precisión para evaluar el desempeño y mostrar una comparativa entre los clasificadores. Esta métrica se calcula con el total de las correctas clasificaciones entre el total de datos del dataset. Y se utilizó una validación cruzada con  $k=10$  para representar la precisión de nuestros clasificadores.

TABLA 4.3: Precisión de clasificadores para cada conjunto

SAMM		SMIC	
Clasificador	Precisión	Clasificador	Precisión
SVM	99.56 %	SVM	98.14 %
Arboles Decisión	98.87 %	Arboles Decisión	93.62 %
Naive Bayes	86.08 %	Naive Bayes	78.29 %
Regresión Logística	99.31 %	Regresión Logística	95.33 %

En las siguientes tablas se indica en que características se realizó la reducción utilizando el algoritmo genético y cuál fue la mejora de usarlo:

SMIC con 1990 instancias y SAMM con 3634 instancias con un total de características de 3824. La Tabla 4.4 se muestra el acomodo de las características obtenidas, siendo este un acomodo arbitrario, pero hemos querido compartir como esta construido el vector de las características.

TABLA 4.4: Vector de Características

Nuevo vector de características	
Total de Características	Tipo de Característica
18	Geométrica Boca
18	Geométrica Ojos
7	Cromáticas Boca
7	Cromáticas Ojos
28	Haralick Boca
28	Haralick Ojos
1800	HoG Ojos
1800	HoG Boca
59	LBP Ojos
59	LBP Boca

En la Tabla 4.5 se muestra en cantidades como el algoritmo genético reduce las características eliminando las más redundantes y menos discriminativas. Se puede apreciar para este conjunto de datos SAMM, una reducción del 85,71 % para las características cromáticas en ojos y un 78,57 % para las características texturales de Haralick en la boca; siendo estas dos las características con mayor índice de eliminación y en un 50 % se eliminaron las características geométricas de los ojos y las texturales de haralick en los ojos. De igual forma con un 53 % se redujeron las características de HoG en los ojos y en la boca.

Se concluye que las características más representativas con un mínimo de reducción son geométricas boca, cromáticas boca y LBP ojos, sin embargo, las características de HoG a pesar de tener una reducción a la mitad son muy importantes para la buena clasificación. Y se destaca también que la reducción que se tuvo utilizando un algoritmo genético fue del 46,24 %.

TABLA 4.5: Total de características SAMM

Tipo de Característica	SAMM	
	Características SIN Genético	Características CON Genético
Geométrica Boca	18	13
Geométrica Ojos	18	9
Cromáticas Boca	7	4
Cromáticas Ojos	7	1
Haralick Boca	28	6
Haralick Ojos	28	14
HoG Ojos	1800	833
HoG Boca	1800	831
LBP Ojos	59	30
LBP Boca	59	27
Total	3825	1768 + Clase

En la Tabla 4.6 se muestran los resultados de aplicar un algoritmo genético en el conjunto de datos SMIC, donde la reducción total fue del 47,94 % en comparativa con el conjunto SAMM, SMIC se incremento un 1,7. Ambos conjuntos de datos practicamente se redujeron a la mitad utilizando un algoritmo genético. De forma particular el algoritmo genético en el conjunto de datos SMIC redujo en un 83,33 % las características geométricas de los ojos, un 71,42 % las características cromáticas de los ojos y un 66,66 % las características Geométricas boca siendo las dos primeras las características que mayor reducción tuvieron. Las características cromáticas de la boca y las de haralick de los ojos tuvieron una reducción del 57,14 %, las características de HoG tanto en boca como en ojos se redujeron un 51,83 % y las características de LBP para ojos y boca tuvieron una reducción del 50,84 % y finalmente con menor reducción fueron las características texturales de Haralick en la boca con un 46,46 %. Estó nos indica que las características menos redundantes fueron las geométricas en ojos y de color en ojos. En contraste con el conjunto SAMM que las menos discriminativas fueron de color en ojos y haralick en boca, comparten la eliminación de características de color en ojos a una tercera parte, se concluye que las resoluciones de las imágenes intervienen mucho en que características son o no discriminantes debido a cuantos detalles se puedan apreciar sobre los ojos y la boca.

TABLA 4.6: Total de características SMIC

Tipo de Característica	SMIC	
	Características SIN Genético	Características CON Genético
Geométrica Boca	18	6
Geométrica Ojos	18	3
Cromáticas Boca	7	3
Cromáticas Ojos	7	2
Haralick Boca	28	15
Haralick Ojos	28	12
HoG Ojos	1800	867
HoG Boca	1800	867
LBP Ojos	59	29
LBP Boca	59	29
Total	3825	1833 + clase

En las siguientes tablas se muestran todas las métricas con las características reducidas por el genético y sin aplicar el genético.

TABLA 4.7: SAMM - Algoritmo Genético aplicado

Clasificador	SAMM - Con AG					
	Acc	TP	FP	Recall	Fm	ROC
SVM	99.559	0.996	0.001	0.996	0.996	1.0
Bayes	86.075	0.861	0.029	0.861	0.859	0.97
RF	98.87	0.990	0.003	0.990	0.990	1.0
LR	99.312	0.993	0.002	0.993	0.993	1.0

TABLA 4.8: SAMM - Desempeño sin reducción de características

Clasificador	SAMM - Sin AG					
	Acc	TP	FP	Recall	Fm	ROC
SVM	98.721	0.987	0.003	0.987	0.986	0.99
Bayes	85.828	0.858	0.029	0.858	0.856	0.95
RF	98.514	0.985	0.004	0.985	0.985	1.0
LR	98.348	0.983	0.005	0.983	0.981	0.99

En las Tabla 4.7 se muestra en el clasificador SVM una precisión de 99,55 %, esta precisión es utilizando un algoritmo genético para la reducción de las características, en comparación con la precisión mostrada en la tabla 4.8 donde no se utiliza un algoritmo genético es de 98,72 % el cual nos muestra que el reducir las características que no son redundantes para nuestra clasificación hace que nuestra precisión se incremente, en este caso hubo un incremento del 0,838 %. En el clasificador regresión logística se incrementa la precisión un 0,964 %, esta comparación es tomando los clasificadores de mayor precisión de los cuatros, sin embargo, en los cuatro clasificadores se tiene un incremento en su precisión con la reducción de características utilizando un algoritmo genético.

TABLA 4.9: SMIC - Algoritmo Genético aplicado

SMIC - Con AG						
Clasificador	Acc	TP	FP	Recall	Fm	ROC
SVM	98.14	0.981	0.006	0.98	0.981	0.99
Bayes	78.29	0.783	0.063	0.78	0.784	0.93
RF	93.61	0.936	0.021	0.93	0.936	0.99
LR	95.87	0.959	0.013	0.95	0.959	0.99

TABLA 4.10: SMIC - Desempeño sin reducción de características

SMIC - Sin AG						
Clasificador	Acc	TP	FP	Recall	Fm	ROC
SVM	97.68	0.977	0.007	0.97	0.977	0.99
Bayes	72.76	0.728	0.073	0.73	0.729	0.89
RF	93.11	0.931	0.023	0.93	0.931	0.99
LR	95.32	0.953	0.014	0.95	0.953	0.99

La Tabla 4.9 nos muestra las métricas de todos los clasificadores utilizados con una reducción en las características, aplicando un algoritmo genético. En el clasificador SVM podemos observar una precisión del 98,14 % y en la Tabla 4.10 que muestra el desempeño sin la reducción de las características, se muestra una precisión del 97,68 % se aprecia una mejora en la precisión del clasificador con un

incremento del 0,46 % y para el clasificador regresión logística se tiene un incremento del 0,55 %, se observa que en todos los clasificadores tuvimos un incremento por reducir características, por lo tanto el utilizar un algoritmo genético para la reducción de las características da mayor eficiencia en la clasificación.

La clasificación sin utilizar el genético nos da una mejor precisión para el conjunto de datos SAMM que el SMIC, por ejemplo para SVM en SAMM tenemos una precisión de 98,72 %, en regresión logística es de 98,34 y para SVM en SMIC es de 97,68 % en la regresión logística es de 95,32 %. Tomando en cuenta que el tamaño de las imágenes de SAMM es de  $960 \times 650$  y las de SMIC es de  $186 \times 227$  y la resolución del conjunto SAMM es mayor que la del conjunto SMIC.

### 4.0.3. Resultados Metodología III

En las siguientes tablas se muestran los resultados con los diferentes clasificadores para cada una de las zonas del rostro para la identificación de emociones, con el entrenamiento de ambos conjuntos de datos. Se puede observar que en todas las zonas del rostro se presentan las mejores precisiones con el clasificador SVM, del cual se utilizó una normalización previa, discretizando los datos bajo un intervalo de 0 a 1, utilizando un Kernel tipo Polykernel.

TABLA 4.11: SAMM - Desempeño de los clasificadores

Clasificador	Ojos	Boca	Rostro	Ojos-Boca
SVM	99.07	99.02	99.72	99.82
Árbol	98.19	98.06	98.87	99.24
Naive Bayes	76.51	87.04	76.90	82.34
Logistic	98.44	98.56	99.52	99.67

En la Tabla 4.11 podemos determinar que la zona con mayor precisión es la de ojos y boca con el clasificador SVM con 99,82 %; siendo el clasificador que más destaque en su precisión, en comparación de todo el rostro que tiene una precisión de 99,72 % podemos ver que utilizar únicamente los ojos y la boca son suficientes para identificar una emoción e incluso incrementar la precisión.

TABLA 4.12: SMIC - Desempeño de los clasificadores

Clasificador	Ojos	Boca	Rostro	Ojos-Boca
SVM	98.06	99.75	100	99.85
Árbol	94.84	99.45	99.70	99.25
Naive Bayes	64.65	80.71	98.66	79.46
Logistic	95.68	99.45	99.80	99.20

En la Tabla 4.12 podemos ver las precisiones obtenidas por cada clasificador en las zonas del rostro, se puede observar que de igual que en conjunto SAMM el clasificador de mejor desempeño es la SVM. Para el conjunto SMIC podemos ver que la mejor precisión es en la zona del rostro completo con 100 %, seguido por la zona de ojos y boca con una precisión de 99,85 % lo cual nos indica que reduce un 0,15 % en precisión comparando ambas propuestas de las zonas del rostro. Esto es debido a que el conjunto de datos SMIC es de menor resolución que el conjunto SAMM. Pero se concluye que el utilizar la zona de los ojos y la boca es tan favorable o mejor que utilizar el rostro completo. Como todo sistema puede tener variaciones de acuerdo con la calidad de la imagen como hemos visto en la comparativa de ambos conjuntos de datos.

# Capítulo 5

## Conclusiones

En este Capítulo se muestran nuestras conclusiones y discusiones de los resultados obtenidos. Para organizar nuestras conclusiones discutimos los resultados por cada metodología empleada.

### 5.1. Metodología I

El uso de las características de forma individual no es tan favorable, como la unión de los tres descriptores. La combinación de las características de Haralick, HoG y LBP nos ayudan a mejorar significativamente en la precisión de nuestra clasificación. Las precisiones que obtuvimos en nuestra investigación fue alta, teniendo una mejor precisión el clasificador SVM, para el primer conjunto SMIC se tuvo una precisión de 98,01 % y para el segundo conjunto SAMM se tiene una precisión de 99,67 %, con estas precisiones podemos comparar con otros trabajos relacionados a la identificación de las emociones y podemos determinar que nuestra precisión está por encima de las que se han reportado. También concluimos que nuestra investigación se realizó con imágenes de mayor resolución, lo que nos puede tener una desventaja en el tiempo de procesamiento para realizar la identificación.

### 5.2. Metodología II

Los resultados arrojados por aplicar un algoritmo genético para la reducción de las características no indica que para el conjunto SAMM se redujo 2056 características que equivale al 53,75 % y para el conjunto de datos SMIC redujo 1991

características, equivalente al 52,09 %

Para el conjunto de datos SAMM se identificó que las características que redujo significativamente fueron las Cromáticas en la zona de los ojos, Haralick de la boca, HoG de los ojos y boca y LBP de la boca. En contraste con el conjunto de datos SMIC que redujo las características Geométricas de la boca y de los ojos, cromática de la boca, Haralick de los ojos y LBP de los ojos. Cabe mencionar que la resolución de las imágenes juega un papel importante, el conjunto SMIC contiene imágenes tamaño  $186 \times 227$  píxeles y el SAMM imágenes tamaño  $960 \times 650$  píxeles.

### 5.3. Metodología III

Los dos conjuntos de datos con los que trabajamos tienen una resolución diferente, pero cabe destacar algo importante. El conjunto de datos SAMM son 3892 imágenes de las personas a medio torso, y el conjunto de datos SMIC son imágenes de solo el rostro de la persona y difieren en tamaño, SMIC tiene 2016 imágenes con un tamaño de  $186 \times 227$  y el conjunto SAMM maneja imágenes de tamaño  $960 \times 650$ , al momento de obtener solo el encuadre del rostro se tienen imágenes de  $186 \times 227$  mismo tamaño que maneja el conjunto de datos SMIC.

Se destaca el estudio de las zonas de interés identificando como afectan en cada una de las emociones, se concluyó que el solo trabajar con la zona de los ojos realiza un decremento mayor comparado a las demás zonas de interés para la identificación de la emoción de tristeza, ya que el solo tomar la zona de los ojos crea una confusión al reconocer esta emoción de tristeza con la emoción de sorpresa. El utilizar únicamente la zona de la boca nos brinda un menor decremento que solo los ojos respecto a esas dos emociones, cuando se toma solo la zona del encuadre del rostro se tiene una mejora del 50 % respecto a los ojos y boca por separado con la emoción de tristeza.

De forma general estudiando las diferentes zonas del rostro discriminativas, con las técnicas de extracción de características y un clasificador SVM con un Kernel polinomial, muestra a todas las zonas muy favorables para identificar las emociones teniendo precisiones mayores del 98 %. Sin embargo, al analizar la zona de la boca y los ojos realizando una concatenación de dichas características a un solo vector el reconocimiento de las emociones se incrementó un 10 % llegando a una precisión mayor del 99,80 % en ambos conjuntos de datos SMIC y SAMM.

Las zonas más discriminativas son el rostro completo o bien ojos-boca, para el conjunto SAMM el rostro completo arroja una precisión de 99,72% y para SMIC del rostro completo da 100%. Para la zona de ojos-boca en el conjunto SAMM da 99,82% y para el conjunto SMIC 99,85%. El ocupar el rostro completo significa un mayor recorrido de píxeles, que implica mayor tiempo y consumo computacional. Sin embargo, el utilizar únicamente la zona de la boca y ojos nos da una precisión muy favorable, que puede ser equitativa a usar todo el rostro pero con menor cantidad de cálculos porque se reduce el número de píxeles de todo el rostro a usar solo el recorte de los ojos con cejas y la boca.



## Capítulo 6

### Artículos publicados

Derivado de la investigación llevada a cabo durante mi estancia en el programa de Doctorado en Ciencias de la Computación se publicaron dos artículos que describen parte de los resultados obtenidos. A continuación se describen los artículos publicados.

- a) Emotion recognition by eyes region using textural features, lbp and hog [Artículo aceptado].

El reconocimiento de las emociones mediante expresiones faciales, enfocando la investigación en zonas de interés de boca y ojos, como zonas con un aspecto discriminativo para identificar la emoción. La identificación de emociones se llevó a cabo con la extracción de características de Haralick, HoG y LBP. Se trabajó con los conjuntos de datos SMIC y SAMM, se llevaron a cabo pruebas de clasificación con cuatro clasificadores SVM, Regresión Logística, Árboles de decisión y Clasificación Bayesiana. Se obtuvieron resultados muy buenos, destacaron con mejores resultados el clasificador SVM (99,6 %) y Regresión Logística (99,56 %).

## Emotion recognition by eyes region using textural features, lbp and hog

Jalili Laura<sup>a†</sup>, Cervantes Jair<sup>b†</sup>, Espejel Josué<sup>c†</sup>

<sup>†</sup> Universidad Autónoma del Estado de México, Prolongación de Av. Zumpango s/n, Fracc. El Tejocote, Texcoco, México, 52346.

<sup>a</sup> Magister Laura Jalili, Doctorante y Docente de la Universidad Autónoma del Estado de México, Texcoco, México.

Contacto [lydominguezj@uaemex.mx](mailto:lydominguezj@uaemex.mx) <sup>b</sup> Doctor Jair Cervantes profesor de tiempo completo de la Universidad Autónoma del Estado de México, Texcoco, México. Contacto [jcervantesc@uaemex.mx](mailto:jcervantesc@uaemex.mx), ORCID: <https://orcid.org/0000-0003-2012-8151>

<sup>c</sup> Doctor Josué Espejel, Doctor en Ciencias de la Computación. Contacto [jespejelc@uaemex.mx](mailto:jespejelc@uaemex.mx)

### RESUMEN

Objetivo: Reconocimiento de emociones con expresiones faciales de dos zonas de interés (los ojos y la boca).

Metodología: El algoritmo propuesto detecta el rostro y obtiene automáticamente la región de los ojos. Usamos algunas técnicas de extracción de características y hacemos una comparación de desempeño con los clasificadores SVM, Regresión Logística, Regresión bayesiana y Árboles de decisión.

Resultados: los mejores resultados son los obtenidos para SVM y Regresión logística, sin embargo, los resultados para SVM son mejores (0.992) que los obtenidos usando regresión logística (0.960).

Conclusiones: La región de los ojos tiene una precisión del 0,99% utilizando un SVM con el sistema propuesto, no es necesario usar todo el rostro.

Financiamiento: Proyecto 6525/2022CIB Universidad Autónoma del Estado de México

Palabras clave: reconocimiento de emociones, regiones de interés, características, clasificación

### ABSTRACT

Objective: Recognition of emotions with facial expressions of two areas of interest (the eyes and the mouth).

Methodology: The proposed algorithm detects the face and automatically gets the eyes region. We use some feature extraction techniques and make a comparison of the performance of the classifiers: SVM,

FIGURA 6.1: Artículo 1. Emotion recognition by eyes region using textural features, lbp and hog

- b) Emotion Recognition from Facial Expressions Using a Genetic Algorithm to Feature Extraction [[https://doi.org/10.1007/978-3-030-84522-3\\_5](https://doi.org/10.1007/978-3-030-84522-3_5)]. En este artículo, presentamos un reconocimiento de emociones utilizando solo dos zonas del rostro, la boca y los ojos de forma conjunta, utilizando técnicas de extracción de características HoG, LBP y Texturales de Haralick. Se utiliza

un algoritmo genético para realizar una reducción de características, eliminando las características menos significativas y las que introducen ruido a nuestros clasificadores como características redundantes. La precisión que se obtuvo fue de 99,56 %



FIGURA 6.2: Artículo 2. Emotion Recognition from Facial Expressions Using a Genetic Algorithm to Feature Extraction



# Bibliografía

- [1] Ibrahim A. Adeyanju, Elijah O. Omidiora y Omobolaji F. Oyedokun. «Performance evaluation of different support vector machine kernels for face emotion recognition». En: *2015 SAI Intelligent Systems Conference (IntelliSys)*. IEEE, 2015. DOI: [10.1109/intellisys.2015.7361233](https://doi.org/10.1109/intellisys.2015.7361233).
- [2] Timo Ahonen, Abdenour Hadid y Matti Pietik inen. «Face Recognition with Local Binary Patterns». En: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2004, p ags. 469-481. DOI: [10.1007/978-3-540-24670-1\\_36](https://doi.org/10.1007/978-3-540-24670-1_36).
- [3] Alberto Albiol et al. «Face recognition using HOG–EBGM». En: *Pattern Recognition Letters* 29.10 (2008), p ags. 1537-1543. DOI: [10.1016/j.patrec.2008.03.017](https://doi.org/10.1016/j.patrec.2008.03.017).
- [4] Enrique Pajares Alegre Guti rrez. *Conceptos y m todos de vision por computador*. Ed. por Grupo de Vision del Comit  Espa ol. 2016.
- [5] Boulbaba Ben Amor et al. «4D Facial Expression Recognition by Learning Geometric Deformations». En: *IEEE Transactions on Cybernetics* 44.12 (dic. de 2014), p ags. 2443-2457. DOI: [10.1109/tcyb.2014.2308091](https://doi.org/10.1109/tcyb.2014.2308091).
- [6] O. Armagan y M. Kahrman. «Comparison of traditional haar classifiers used in face detection applications with an alternative classifier for four stages filtering». En: *2014 22nd Signal Processing and Communications Applications Conference (SIU)*. IEEE, 2014. DOI: [10.1109/siu.2014.6830676](https://doi.org/10.1109/siu.2014.6830676).
- [7] Juan M. Arriola. «Representaci n matem tica de ondas cerebrales». Tesis doct. Universidad Nacional del sur, 2016.
- [8] Barun Kumar Bairagi et al. «Expressions invariant face recognition using SURF and Gabor features». En: *2012 Third International Conference on Emerging Applications of Information Technology*. IEEE, 2012. DOI: [10.1109/eait.2012.6407888](https://doi.org/10.1109/eait.2012.6407888).

- [9] Herbert Bay, Tinne Tuytelaars y Luc Van Gool. «SURF: Speeded Up Robust Features». En: *Computer Vision ECCV*. Springer Berlin Heidelberg, 2006, págs. 404-417. DOI: [10.1007/11744023\\_32](https://doi.org/10.1007/11744023_32).
- [10] Ashok Bekkanti et al. «Computer Based Detection of Alcohol Consumed Candidates Using Face Expressions with SIFT and Bag of Words». En: *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2021. DOI: [10.1109/icoei51242.2021.9453057](https://doi.org/10.1109/icoei51242.2021.9453057).
- [11] Sergio Benini et al. «Face analysis through semantic face segmentation». En: *Signal Processing: Image Communication* 74 (2019), págs. 21-31. DOI: [10.1016/j.image.2019.01.005](https://doi.org/10.1016/j.image.2019.01.005).
- [12] Fella Berrimi, Khier Benmahammed y Riadh Hedli. «Denoising of degraded face images sequence in PCA domain for recognition». En: *Journal of King Saud University - Computer and Information Sciences* (2019). DOI: [10.1016/j.jksuci.2019.04.014](https://doi.org/10.1016/j.jksuci.2019.04.014).
- [13] Hadjer Boubenna y Dohoon Lee. «Feature selection for facial emotion recognition based on genetic algorithm». En: *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. IEEE, 2016. DOI: [10.1109/fskd.2016.7603226](https://doi.org/10.1109/fskd.2016.7603226).
- [14] Diego Calvo. *Red Neuronal Recurrente â RNN*. Dic. de 2018. URL: <http://www.diegocalvo.es/red-neuronal-recurrente/>.
- [15] Jair Cervantes, Xiaoou Li y Wen Yu. «Imbalanced data classification via support vector machines and genetic algorithms». En: *Connection Science* 26.4 (2014), págs. 335-348. DOI: [10.1080/09540091.2014.924902](https://doi.org/10.1080/09540091.2014.924902).
- [16] A. Chakraborty et al. «Emotion Recognition From Facial Expressions and Its Control Using Fuzzy Logic». En: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 39.4 (2009), págs. 726-743. DOI: [10.1109/tsmca.2009.2014645](https://doi.org/10.1109/tsmca.2009.2014645).
- [17] Abbas Cheddad, Dzulkifli Mohamad y Azizah Abd Manaf. «Exploiting Voronoi diagram properties in face segmentation and feature extraction». En: *Pattern Recognition* 41.12 (2008), págs. 3842-3859. DOI: [10.1016/j.patcog.2008.06.007](https://doi.org/10.1016/j.patcog.2008.06.007).

- [18] Tian Chen et al. «Emotion recognition using empirical mode decomposition and approximation entropy». En: *Computers & Electrical Engineering* 72 (2018), págs. 383 -392. DOI: [10.1016/j.compeleceng.2018.09.022](https://doi.org/10.1016/j.compeleceng.2018.09.022).
- [19] Moi Hoon Yap Chuin Hong Yap Connah Kendrick. «SAMM Long Videos: A Spontaneous Facial Micro- and Macro-Expressions Dataset». En: *15th IEEE International Conference on Automatic Face and Gesture Recognition* (2020).
- [20] Alister Cordiner, Philip Ogunbona y Wanqing Li. «Face detection using generalised integral image features». En: *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2009. DOI: [10.1109/icip.2009.5413646](https://doi.org/10.1109/icip.2009.5413646).
- [21] N. Dalal y B. Triggs. «Histograms of Oriented Gradients for Human Detection». En: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. IEEE. DOI: [10.1109/cvpr.2005.177](https://doi.org/10.1109/cvpr.2005.177).
- [22] Adrian Davison, Walied Merghani y Moi Yap. «Objective Classes for Micro-Facial Expression Recognition». En: *Journal of Imaging* 4.10 (2018), pág. 119. DOI: [10.3390/jimaging4100119](https://doi.org/10.3390/jimaging4100119).
- [23] J. Domke e Y. Aloimonos. «Deformation and Viewpoint Invariant Color Histograms». En: *Proceedings of the British Machine Vision Conference 2006*. British Machine Vision Association, 2006. DOI: [10.5244/c.20.53](https://doi.org/10.5244/c.20.53).
- [24] Dykinson. *Vision por computador*. 2003.
- [25] Nestel Ebel. *Sensores para la técnica de procesos y manipulación*. Festo Didactic, 2000.
- [26] Paul Ekman. «Facial expression and emotion.» En: *American Psychologist* 48.4 (1993), págs. 384-392. DOI: [10.1037/0003-066x.48.4.384](https://doi.org/10.1037/0003-066x.48.4.384).
- [27] Fachruddin et al. «Real Time Detection on Face Side Image with Ear Biometric Imaging Using Integral Image and Haar-Like Feature». En: *2018 International Conference on Electrical Engineering and Computer Science (ICECOS)*. IEEE, 2018. DOI: [10.1109/icecos.2018.8605218](https://doi.org/10.1109/icecos.2018.8605218).
- [28] Laurene Fausett. *Fundamentals of Neural Networks*. 1994.
- [29] J. Geusebroek, A.W.M. Smeulders y J. van de Weijer. «Fast anisotropic gauss filtering». En: *IEEE Transactions on Image Processing* 12.8 (2003), págs. 938-943. DOI: [10.1109/tip.2003.812429](https://doi.org/10.1109/tip.2003.812429).

- [30] Salvador Gutierrez. *TecnologÍa sensorica en agricultura*. Ed. por Esslingen. Parallax, 2019. ISBN: 3-8127-3047-2.
- [31] Xuanyu He y Wei Zhang. «Emotion recognition by assisted learning with convolutional neural networks». En: *Neurocomputing* 291 (2018), págs. 187-194. DOI: [10.1016/j.neucom.2018.02.073](https://doi.org/10.1016/j.neucom.2018.02.073).
- [32] Hao Hu, Ming-Xing Xu y Wei Wu. «GMM Supervector Based SVM with Spectral Features for Speech Emotion Recognition». En: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*. IEEE, 2007. DOI: [10.1109/icassp.2007.366937](https://doi.org/10.1109/icassp.2007.366937).
- [33] Deepak Kumar Jain, Pourya Shamsolmoali y Paramjit Sehdev. «Extended deep neural network for facial emotion recognition». En: *Pattern Recognition Letters* 120 (2019), págs. 69-74. DOI: [10.1016/j.patrec.2019.01.008](https://doi.org/10.1016/j.patrec.2019.01.008).
- [34] Medsker Jain. *Recurrent Neural Networks*. 2001.
- [35] Laura D. Jalili et al. «Improving the Performance of Leaves Identification by Features Selection with Genetic Algorithms». En: *Communications in Computer and Information Science*. Springer International Publishing, 2016, págs. 103-114. DOI: [10.1007/978-3-319-50880-1\\_10](https://doi.org/10.1007/978-3-319-50880-1_10).
- [36] Eiman Kanjo, Eman M.G. Younis y Chee Siang Ang. «Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection». En: *Information Fusion* 49 (2019), págs. 46-56. DOI: [10.1016/j.inffus.2018.09.001](https://doi.org/10.1016/j.inffus.2018.09.001).
- [37] Li Ke y Jingjing Kang. «Eye location method based on Haar features». En: *2010 3rd International Congress on Image and Signal Processing*. IEEE, 2010. DOI: [10.1109/cisp.2010.5646905](https://doi.org/10.1109/cisp.2010.5646905).
- [38] N.M. Khan et al. «A novel SVMNDA model for classification with an application to face recognition». En: *Pattern Recognition* 45.1 (2012), págs. 66-79. DOI: [10.1016/j.patcog.2011.05.004](https://doi.org/10.1016/j.patcog.2011.05.004).
- [39] Ho duck Kim et al. «Genetic Algorithm Based Feature Selection Method Development for Pattern Recognition». En: *2006 SICE-ICASE International Joint Conference*. IEEE, 2006. DOI: [10.1109/sice.2006.315742](https://doi.org/10.1109/sice.2006.315742).
- [40] Hyun-Chul Kim, Daijin Kim y Sung Yang Bang. «Face recognition using the mixture-of-eigenfaces method». En: *Pattern Recognition Letters* 23.13 (2002), págs. 1 549-1558. DOI: [10.1016/s0167-8655\(02\)00119-8](https://doi.org/10.1016/s0167-8655(02)00119-8).

- [41] Bernhard Kratzwald et al. «Deep learning for affective computing: Text-based emotion recognition in decision support». En: *Decision Support Systems* 115 (2018), págs. 24-35. DOI: [10.1016/j.dss.2018.09.002](https://doi.org/10.1016/j.dss.2018.09.002).
- [42] Vladimir Kurbalija et al. «Emotion perception and recognition: An exploration of cultural differences and similarities». En: *Cognitive Systems Research* 52 (2018), págs. 103-116. DOI: [10.1016/j.cogsys.2018.06.009](https://doi.org/10.1016/j.cogsys.2018.06.009).
- [43] Khadija Lekdioui et al. «Facial decomposition for expression recognition using texture/shape descriptors and SVM classifier». En: *Signal Processing: Image Communication* 58 (2017), págs. 300-312. DOI: [10.1016/j.image.2017.08.001](https://doi.org/10.1016/j.image.2017.08.001).
- [44] Bo Li, Chun-Hou Zheng y De-Shuang Huang. «Locally linear discriminant embedding: An efficient method for face recognition». En: *Pattern Recognition* 41.12 (2008), págs. 3813-3821. DOI: [10.1016/j.patcog.2008.05.027](https://doi.org/10.1016/j.patcog.2008.05.027).
- [45] Xiaobai Li et al. «A Spontaneous Micro-expression Database: Inducement, collection and baseline». En: *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE, 2013. DOI: [10.1109/fg.2013.6553717](https://doi.org/10.1109/fg.2013.6553717).
- [46] Xiaobai Li et al. «Towards Reading Hidden Emotions: A Comparative Study of Spontaneous Micro-Expression Spotting and Recognition Methods». En: *IEEE Transactions on Affective Computing* 9.4 (2018), págs. 563-577. DOI: [10.1109/taffc.2017.2667642](https://doi.org/10.1109/taffc.2017.2667642).
- [47] Chung-Wei Liang y Chia-Feng Juang. «Moving object classification using local shape and HOG features in wavelet-transformed space with hierarchical SVM classifiers». En: *Applied Soft Computing* 28 (2015), págs. 483-497. DOI: [10.1016/j.asoc.2014.09.051](https://doi.org/10.1016/j.asoc.2014.09.051).
- [48] Zhen Liang, Shigeyuki Oba y Shin Ishii. «An unsupervised EEG decoding system for human emotion recognition». En: *Neural Networks* 116 (2019), págs. 257-268. DOI: [10.1016/j.neunet.2019.04.003](https://doi.org/10.1016/j.neunet.2019.04.003).
- [49] Hong Liu et al. «Related HOG Features for Human Detection Using Cascaded Adaboost and SVM Classifiers». En: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2013, págs. 345-355. DOI: [10.1007/978-3-642-35728-2\\_33](https://doi.org/10.1007/978-3-642-35728-2_33).

- [50] David G. Lowe. «Distinctive Image Features from Scale-Invariant Keypoints». En: *International Journal of Computer Vision* 60.2 (2004), págs. 91-110. DOI: [10.1023/b:visi.0000029664.99615.94](https://doi.org/10.1023/b:visi.0000029664.99615.94).
- [51] Ji Ma y Yuyu Yuan. «Dimension reduction of image deep feature using PCA». En: *Journal of Visual Communication and Image Representation* 63 (2019), pág. 102578. DOI: [10.1016/j.jvcir.2019.102578](https://doi.org/10.1016/j.jvcir.2019.102578).
- [52] Gonzalo Pajares Martinsanz. *Ejercicios Resueltos de Visión por computador*. Ed. por RA-MA Editorial. 2008.
- [53] Gonzalo Pajares Martinsanz. *Vision Computacional*. Ed. por Alfaomega. 2004.
- [54] Bishwas Mishra et al. «Facial expression recognition using feature based techniques and model based techniques: A survey». En: *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*. IEEE, 2015. DOI: [10.1109/ecs.2015.7124976](https://doi.org/10.1109/ecs.2015.7124976).
- [55] Sandra E. Nope, Humberto Loaiza y Eduardo Caicedo. «Modelo Bio-inspirado para el Reconocimiento de Gestos Usando Primitivas de Movimiento en Visión». En: *Revista Iberoamericana de Automática e Informática Industrial RIAI* 5.4 (2008), págs. 69-76. DOI: [10.1016/s1697-7912\(08\)70179-1](https://doi.org/10.1016/s1697-7912(08)70179-1).
- [56] Yanwei Pang et al. «Efficient HOG human detection». En: *Signal Processing* 91.4 (2011), págs. 773-781. DOI: [10.1016/j.sigpro.2010.08.010](https://doi.org/10.1016/j.sigpro.2010.08.010).
- [57] Tomas Pfister et al. «Recognising spontaneous facial micro-expressions». En: *2011 International Conference on Computer Vision*. IEEE, 2011. DOI: [10.1109/iccv.2011.6126401](https://doi.org/10.1109/iccv.2011.6126401).
- [58] Matti Pietikäinen. «Image Analysis with Local Binary Patterns». En: *Image Analysis*. Springer Berlin Heidelberg, 2005, págs. 115-118. DOI: [10.1007/11499145\\_13](https://doi.org/10.1007/11499145_13).
- [59] Adri Priadana y Muhammad Habibi. «Face Detection using Haar Cascades to Filter Selfie Face Image on Instagram». En: *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIIT)*. IEEE, 2019. DOI: [10.1109/icaiit.2019.8834526](https://doi.org/10.1109/icaiit.2019.8834526).
- [60] Richard E. Woods Rafael C. Gonzalez. *Digital Image Processing*. Ed. por Addison-Wesley Pub. 2001.

- [61] Jisy Raju y C. Anand Deva Durai. «A survey on texture classification techniques». En: *2013 International Conference on Information Communication and Embedded Systems (ICICES)*. 2013, págs. 180-184. DOI: [10.1109/ICICES.2013.6508183](https://doi.org/10.1109/ICICES.2013.6508183).
- [62] K. P. Rao, M. V. P. Chandra Sekhara Rao y N. Hemanth Chowdary. «An integrated approach to emotion recognition and gender classification». En: *Journal of Visual Communication and Image Representation* 60 (2019), págs. 339-345. DOI: [10.1016/j.jvcir.2019.03.002](https://doi.org/10.1016/j.jvcir.2019.03.002).
- [63] Julio Cesar Mello Roman. «Mejora de contraste utilizando morfología matemática multiescala para imágenes en escala de grises e imágenes en color». Tesis doct. Universidad Nacional de Asunción, 2017.
- [64] Vibha. V. Salunke y C.G. Patil. «A New Approach for Automatic Face Emotion Recognition and Classification Based on Deep Networks». En: *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*. IEEE, 2017. DOI: [10.1109/iccubea.2017.8463785](https://doi.org/10.1109/iccubea.2017.8463785).
- [65] R. Santhoshkumar y M. Kalaiselvi Geetha. «Deep Learning Approach for Emotion Recognition from Human Body Movements with Feedforward Deep Convolution Neural Networks». En: *Procedia Computer Science* 152 (2019), págs. 158-165. DOI: [10.1016/j.procs.2019.05.038](https://doi.org/10.1016/j.procs.2019.05.038).
- [66] Shira C. Segal et al. «Children's recognition of emotion expressed by own-race versus other-race faces». En: *Journal of Experimental Child Psychology* 182 (2019), págs. 102-113. DOI: [10.1016/j.jecp.2019.01.009](https://doi.org/10.1016/j.jecp.2019.01.009).
- [67] M. Shamim y Ghulam Muhammad. «Emotion recognition using secure edge and cloud computing». En: *Information Sciences* 504 (2019), págs. 589-601. DOI: [10.1016/j.ins.2019.07.040](https://doi.org/10.1016/j.ins.2019.07.040).
- [68] Ben Krose y Patrick van der Smagt. *An introduction to neural network*. 1996.
- [69] Hamit Soyel y Hasan Demirel. «Improved SIFT matching for pose robust facial expression recognition». En: *Face and Gesture 2011*. IEEE, 2011. DOI: [10.1109/fg.2011.5771463](https://doi.org/10.1109/fg.2011.5771463).
- [70] Fritz Strack, Leonard L. Martin y Sabine Stepper. «Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis.» En: *Journal of Personality and Social Psychology* 54.5 (1988), págs. 768-777. DOI: [10.1037/0022-3514.54.5.768](https://doi.org/10.1037/0022-3514.54.5.768).

- [71] Enrique Sucar. *Visión Computacional*. 2004.
- [72] Jian-Ming Sun, Xue-Sheng Pei y Shi-Sheng Zhou. «Facial emotion recognition in modern distant education system using SVM». En: *2008 International Conference on Machine Learning and Cybernetics*. IEEE, 2008. DOI: [10.1109/icmlc.2008.4621018](https://doi.org/10.1109/icmlc.2008.4621018).
- [73] Felipe Torres. «Adquisición y análisis de señales cerebrales». En: *Adquisición y análisis de señales cerebrales*. 2014.
- [74] Jordi Torres. *DEEP LEARNING Introducción práctica con Keras*. Ed. por Watch this space. 2018.
- [75] P. Viola y M. Jones. «Rapid object detection using a boosted cascade of simple features». En: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. CVPR 2001. IEEE Comput. Soc, 2001. DOI: [10.1109/cvpr.2001.990517](https://doi.org/10.1109/cvpr.2001.990517).
- [76] Jie Wang, K.N. Plataniotis y A.N. Venetsanopoulos. «Selecting discriminant eigenfaces for face recognition». En: *Pattern Recognition Letters* 26.10 (2005), págs. 1470-1482. DOI: [10.1016/j.patrec.2004.11.029](https://doi.org/10.1016/j.patrec.2004.11.029).
- [77] Shangfei Wang et al. «A Natural Visible and Infrared Facial Expression Database for Expression Recognition and Emotion Inference». En: *IEEE Transactions on Multimedia* 12.7 (2010), págs. 682-691. DOI: [10.1109/tmm.2010.2060716](https://doi.org/10.1109/tmm.2010.2060716).
- [78] Baohan Xu et al. «Heterogeneous Knowledge Transfer in Video Emotion Recognition, Attribution and Summarization». En: *IEEE Transactions on Affective Computing* 9.2 (2018), págs. 255-270. DOI: [10.1109/taffc.2016.2622690](https://doi.org/10.1109/taffc.2016.2622690).
- [79] Yu-Feng Yu et al. «Discriminative multi-layer illumination-robust feature extraction for face recognition». En: *Pattern Recognition* 67 (2017), págs. 201-212. DOI: [10.1016/j.patcog.2017.02.004](https://doi.org/10.1016/j.patcog.2017.02.004).
- [80] F. Zago et al. «A survey on facial emotion recognition techniques: A state-of-the-art literature review». En: *Information Sciences* 582 (2022), págs. 593-617. DOI: [10.1016/j.ins.2021.10.005](https://doi.org/10.1016/j.ins.2021.10.005).

- [81] Caili Zhang et al. «Face Detection Algorithm Based on Improved AdaBoost and New Haar Features». En: *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2019. DOI: [10.1109/cisp-bmei48845.2019.8965841](https://doi.org/10.1109/cisp-bmei48845.2019.8965841).
- [82] Kaihao Zhang et al. «Facial Expression Recognition Based on Deep Evolutional Spatial-Temporal Networks». En: *IEEE Transactions on Image Processing* 26.9 (2017), págs. 4193-4203. DOI: [10.1109/tip.2017.2689999](https://doi.org/10.1109/tip.2017.2689999).
- [83] Qiang Zhang et al. «Respiration-based emotion recognition with deep learning». En: *Computers in Industry* 92-93 (2017), págs. 84-90. DOI: [10.1016/j.compind.2017.04.005](https://doi.org/10.1016/j.compind.2017.04.005).
- [84] Yanfeng Zhang y Peikun He. «A revised AdaBoost algorithm: FM-AdaBoost». En: *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*. IEEE, 2010. DOI: [10.1109/iccas.2010.5623209](https://doi.org/10.1109/iccas.2010.5623209).
- [85] Zhenyu Zhang y Xiaoyao Xie. «Research on AdaBoost.M1 with Random Forest». En: *2010 2nd International Conference on Computer Engineering and Technology*. IEEE, 2010. DOI: [10.1109/iccet.2010.5485910](https://doi.org/10.1109/iccet.2010.5485910).
- [86] Zhong-Qiu Zhao, De-Shuang Huang y Bing-Yu Sun. «Human face recognition based on multi-features using neural networks committee». En: *Pattern Recognition Letters* 25.12 (2004), págs. 1351-1358. DOI: [10.1016/j.patrec.2004.05.008](https://doi.org/10.1016/j.patrec.2004.05.008).