UNIVERSIDAD
Panamericana

FACULTAD DE INGENIERÍA

"EVALUACIÓN DE REACCIONES EMOCIONALES A PRODUCTOS ALIMENTICIOS POR MEDIO DE RECONOCIMIENTO AUTOMÁTICO DE EXPRESIONES FACIALES"

TESIS

P R E S E N T A

VÍCTOR MANUEL ÁLVAREZ PATO

PARA OBTENER EL GRADO DE
DOCTOR EN INGENIERÍA

CON RECONOCIMIENTO DE VALIDEZ OFICIAL DE ESTUDIOS DE LA SECRETARÍA DE EDUCACIÓN PÚBLICA, SEGÚN ACUERDO CON EL No. 20171659 DE FECHA 12 DE MAYO 2017

DIRECTOR DE TESIS:
DR. RAMIRO VELAZQUEZ GUERRERO

AGUASCALIENTES, AGS., JUNIO 2022

UNIVERSIDAD
Panamericana

Aguascalientes, Ags., 30 de mayo de 2022

LIBERACIÓN DE TESIS

Por medio de la presente, certifico en calidad de director de tesis, que el trabajo de Víctor Manuel Álvarez Pato, que lleva como título: **EVALUACIÓN DE REACCIONES EMOCIONALES A PRODUCTOS ALIMENTICIOS POR MEDIO DE RECONOCIMIENTO AUTOMÁTICO DE EXPRESIONES FACIALES** cumple con los requisitos establecidos por el reglamento vigente de la Facultad de Ingeniería para presentarse en Examen de Titulación del programa de Doctorado en Ingeniería.

De resultar aprobado, podrá efectuar el trámite para la obtención del grado de Doctor en Ingeniería.

Atentamente,

Dr. Ramiro Velázquez Guerrero
Director de tesis

Agradecimientos

A Dios y a mi familia en primer lugar, porque les debo todo.

Al Dr. Ramiro Velazquez y a la Dra. Teresa Orvañanos por sus orientaciones, a la Dra. Julieta Domínguez, la Dra. Claudia Sánchez por invitarme a colaborar en su investigación. A todos ellos por su apoyo y su compañía.

A la Dra. Guadalupe Ortíz de Landázuri, por resolver los problemas que nadie más podía.

Índice general

Lista de Figuras

Lista de Tablas

Resumen

Las emociones son mecanismos que los seres vivos más complejos han desarrollado a lo largo de millones de años de evolución. Como tales, forman parte integral del ser humano y afectan su comportamiento, muchas veces de manera inconsciente pero determinante. Suelen presentarse acompañadas de manifestaciones fisiológicas entre las que se cuentan las expresiones faciales, cambios en el ritmo cardiaco y la respuesta galvánica de la piel.

En esta tesis desarrollamos un sistema basado en redes neuronales convolucionales para interpretar expresiones faciales. También utilizamos otros algoritmos de inteligencia artificial para estudiar señales biométricas y determinar qué tan útiles pueden resultar para entender las emociones de una persona y eventualmente predecir sus reacciones a ciertos alimentos, lo que puede resultar de capital importancia en el desarrollo de este tipo de productos.

La red neuronal resultante muestra resultados equiparables a los obtenidos en estudios similares por medio de software comercial y los análisis realizados arrojan luz sobre el papel que pueden desempeñar las mediciones de ritmo cardiaco, respuesta galvánica y expresiones faciales en este tipo de investigaciones.

Palabras clave: Predicción de la Aceptación del Consumidor, Análisis Sensorial, Reconocimiento de Expresiones Faciales, Reconocimiento de Emociones, Redes Neuronales, Aprendizaje Automático, Respuesta Galvánica de la Piel.

Capítulo 1

Introducción

Se estima que alrededor del 85 % [1] de los productos de anaquel fracasan al intentar abrirse paso en el mercado, lo que significa pérdidas millonarias para las compañías que los diseñan, fabrican y comercializan. Es por esta razón que se han dedicado grandes esfuerzos a tratar del predecir la aceptación que tendrá un determinado producto de cara a sus consumidores potenciales. Una herramienta útil en este sentido es el análisis sensorial, una disciplina científica utilizada para medir e interpretar reacciones a las características de un producto como son percibidas por los sentidos de la vista, el olfato, el gusto, etc. Un buen programa de calidad sensorial asegura que cualquier problema en el producto pueda ser detectado antes de su exposición al consumidor, pues se estima que por cada consumidor que emite una queja, existen otros diez que simplemente dejarán de comprar el producto [2].

Los estudios de mercado previos al lanzamiento de un nuevo producto se sirven de entrevistas, grupos focales, encuestas y otras técnicas de investigación, principalmente enfocadas a tratar de conocer la opinión de una muestra estadísticamente representativa de personas con respecto al producto en cuestión. Sin embargo, dichas técnicas no son totalmente eficaces, entre otras razones, porque apelan al pensamiento consciente de la persona, mientras que una buena parte de las decisiones de compra se toman de manera más emotiva que racional [3, 4]. Esto ha dado origen a diversas disciplinas, como el *neuromarketing*, que busca obtener datos más precisos sobre la reacción visceral de los consumidores por medio de mediciones psicofísicas. En esta dirección se han orientado muchos estudios recientes, apoyándose en registros de actividad cerebral, cardiaca, muscular, etc. para tratar de profundizar en la respuesta emocional del ser humano a diversos estímulos [5, 6, 7].

A causa de los avances en la tecnología de microprocesadores, la última década ha sido testigo de un crecimiento acelerado en el área de la inteligencia artificial y especialmente en las técnicas de aplicación de redes neuronales, gracias a las cuales contamos ahora con computadoras capaces de aprender de manera pare-

cida —si bien distante— a como lo hacen los humanos [8]. Las redes neuronales han permitido crear programas informáticos capaces de clasificar imágenes y etiquetar con un buen nivel de precisión lo que éstas contienen: automóviles, personas, rostros, animales, etc.

Por otro lado, muchos psicólogos están convencidos de que es posible conocer con cierta exactitud las emociones que experimenta una persona analizando sus expresiones faciales, y que además, esas emociones son comunes a todo el género humano bajo determinadas circunstancias. Esto ha llevado a la comunidad científica a tratar de crear software capaz de identificar las emociones expresadas en el rostro humano de manera automática, con la idea de conocer los sentimientos de una persona por medio una computadora que procese imágenes faciales utilizando —entre otras técnicas— redes neuronales [9, 10, 11].

1.1. Justificación y objetivos

Varios autores han realizado estudios de análisis sensorial y de mercado apoyándose en software comercial para reconocimiento de expresiones faciales, habitualmente *FaceReader*, una herramienta de propósito general [12, 13, 14, 15]. El objetivo de nuestro trabajo es desarrollar un sistema propio de reconocimiento de expresiones faciales que nos permita adaptarlo en un futuro a su uso específico en análisis de aceptación del consumidor y buscar posibles correlaciones entre expresiones faciales y la intención de compra o la aceptación hacia un producto alimenticio.

Proponemos un sistema enfocado a determinar y clasificar las emociones que experimenta una persona al recibir ciertos estímulos, sirviéndonos de inteligencia artificial para interpretar señales biométricas, así como del reconocimiento automatizado de expresiones faciales. Nos interesa establecer si es posible y en qué medida, realizar un programa capaz de clasificar emociones automáticamente y aplicar estos resultados para determinar la aceptación del consumidor.

1.2. Estructura

El presente trabajo está estructurado del siguiente modo:

El Capítulo 2 expone el marco teórico en el que se encuadra la investigación: aborda el concepto de emoción, sus manifestaciones fisiológicas y distintas aproximaciones que se utilizan comúnmente para medirla. Más específicamente, se ocupa de la posibilidad de traducir expresiones faciales y señales biométricas en términos de emociones básicas.

Las principales bases teóricas de este trabajo se describen en el Capítulo 3: la teoría de Ekman sobre la expresión de emociones en el rostro, algunas técnicas de procesamiento de imágenes digitales y las redes neuronales convolucionales son las herramientas básicas utilizadas en este trabajo.

En el Capítulo 4 se describen con detalle los experimentos realizados y el modo en que se procesó la información para obtener los resultados, así como la interpretación de los mismos.

El Capítulo 5 explica una aplicación muy concreta de las técnicas desarrolladas: la predicción de la preferencia del consumidor hacia ciertos tipos de alimentos, midiendo la emoción que éstos pueden producir en quien los prueba.

Finalmente, las conclusiones de la investigación se exponen en el Capítulo 6 junto con algunas posibilidades para continuarla en el futuro.

Capítulo 2

Las Emociones

Las emociones siempre han tenido una gran influencia en toda empresa humana importante. Casi todos los grandes filósofos han especulado sobre su naturaleza, orígenes y efectos. Los teólogos reconocen la relevancia de ciertas emociones en la experiencia religiosa y han dado gran importancia a su educación. Escritores, artistas y músicos siempre han buscado provocar emociones a través de la comunicación simbólica.

«No hay un tema que, a pesar de su enorme influjo en la vida ordinaria, presente un mayor número de opiniones e hipótesis científicas no sólo distintas, sino las más de las veces contrarias. Tal vez esto sea debido a tres motivos: a) la oscuridad que la afectividad presenta a la razón, b) la complejidad que el tema envuelve en sí mismo, c) la pluralidad de enfoques con que se lo puede analizar. En efecto, por una parte, la afectividad parece accesible a cualquier ser humano, en tanto que este es capaz de experimentar una gama muy variada de sentimientos (placer, dolor, odio, amor, ira, esperanza, etc.); por otra, pocas realidades, como la afectividad, son tan complejas y difíciles de explicar. ¿Cuál es su origen? ¿En qué consiste? ¿Qué función desempeña en la vida humana, en particular en el desarrollo de la racionalidad? Son sólo algunas de las preguntas que surgen al examinar el mundo afectivo» [16].

«Por último, por tratarse de una experiencia en que se muestra la complejidad del ser humano (cambios fisiológicos, conciencia de sí, juicios, inclinaciones hacia diferentes acciones, etc.), los métodos usados presentan una gran variedad: se va desde la introspección de la conciencia hasta el análisis del comportamiento, pasando por las neurociencias y la llamada inteligencia artificial» [16].

Tal vez sea a causa de estas dificultades que a pesar del desarrollo reciente de la medicina psicosomática, la psicología clínica y el psicoanálisis —entre otras disciplinas— ha habido cierta reticencia a hablar de la emoción dentro de la psicología hasta la década de 1960 [17]. Durante varios años, la emoción se ha visto como un concepto acientífico caracterizado por el subjetivismo, e incluso

cuando se estudiaba dentro de la psicopatología, su enfoque tradicional no se dirigía a un amplio espectro de emociones, sino sobre todo a la ansiedad [18].

No obstante, en fechas más recientes, el concepto de emoción ha cobrado mayor importancia en el campo de la psicología, como lo demuestra la popularidad del concepto de inteligencia emocional como lo describe Goleman [19], así como la psicología positiva, un enfoque propuesto por Seligman [20] que se aparta del hábito de estudiar la psicología con intención de tratar casos de enfermedad mental para abocarse al fomento de aspectos más positivos como la creatividad, la resiliencia y la felicidad en personas sanas.

2.1. Qué es la emoción

Aunque todos tenemos una idea intuitiva de lo que significa la palabra emoción, aún no existe una definición universalmente aceptada [21, 22, 23].

El diccionario de la Real Academia Española [24] define la voz emoción como «alteración del ánimo intensa y pasajera, agradable o penosa, que va acompañada de cierta conmoción somática», mientras que el diccionario Merriam-Webster [25] la asocia a «una reacción mental consciente (tal como enojo o miedo) experimentada de modo subjetivo como un sentimiento fuerte, usualmente orientado hacia un objeto específico y típicamente acompañado por cambios fisiológicos y conductuales en el cuerpo» [1]. Pueden ser definiciones incompletas para un estudio profundo de la materia, pero funcionan como una primera aproximación.

Desde un punto de vista fisiológico, se dice que las emociones son disposiciones a actuar. Evolucionaron a partir de simples acciones, muchas de las cuales aún forman parte de las posibles respuestas del ser humano. Ciertos estímulos provocan cambios metabólicos en músculos y glándulas en el cuerpo, que lo preparan para acercarse a un estímulo positivo o huir de una amenaza [26].

A causa de las múltiples definiciones de emoción que existen, resulta más sencillo que los estudiosos se pongan de acuerdo en describir sus funciones [22], por lo cual citamos algunas mencionadas por Rolls [27]:

1. Producción de respuestas autónomas y endócrinas que preparan al cuerpo para la acción
2. Flexibilidad de respuestas conductuales hacia estímulos de refuerzo: permite seleccionar entre varios posibles estímulos de recompensa o castigo, junto con sus costos asociados.
3. Motivación
4. Comunicación con otros miembros de la misma especie

[1] Trad. del autor

5. Vinculación social
6. El estado anímico puede afectar la evaluación cognitiva de eventos o recuerdos, lo que da continuidad a la interpretación de los valores de refuerzo encontrados en el ambiente.
7. Facilitar el almacenamiento de recuerdos.
8. Al mantenerse durante cierto tiempo, puede sostener la motivación y producir la redirección del comportamiento necesarios para alcanzar una meta determinada.
9. Recobrar memorias o recuerdos.

En cuanto al origen de las emociones, podemos mencionar varias teorías distintas:

Teoría Evolutiva

Charles Darwin [28] fue quien propuso que las emociones son una ventaja evolutiva que ayuda a la supervivencia de los humanos y otros animales. De acuerdo a esta teoría, las emociones existen porque cumplen una función adaptativa.

Teoría de James-Lange

Sugiere que las emociones aparecen como resultado de reacciones fisiológicas a un estímulo [29]. Según esta teoría, la percepción de un estímulo externo conlleva una reacción fisiológica, pero la reacción emocional depende de la interpretación que el individuo haga de dichas reacciones fisiológicas.

Teoría de Cannon-Bard

En contraste con la teoría de James-Lange, Walter Cannon [30] sugiere que una persona puede experimentar reacciones fisiológicas vinculadas a emociones sin siquiera sentir estas emociones. La teoría de Cannon-Bard afirma que la emoción y los cambios fisiológicos asociados son simultáneos y que ninguna es causa de la otra [31].

Teoría de Schachter-Singer

De acuerdo a este enfoque, lo primero en aparecer es la excitación fisiológica. Posteriormente, el individuo debe identificar la razón por la que la experimenta y etiquetarla como emoción. La emoción es resultado de una interpretación cognitiva [32].

Teoría de la evaluación cognoscitiva

Propone que las emociones son causadas por la valoración que damos a una situación determinada y acepta que existe una amplia variabilidad en las respuestas emocionales de distintos individuos hacia el mismo evento [33].

Teoría de la retroalimentación facial

Los partidarios de esta hipótesis proponen que la activación de ciertos músculos faciales genera emociones [34]. De este modo, una persona que es obligada a sonreír en una reunión social, lo disfrutará más que si hubiera mantenido una cara larga.

2.2. Evolución del cerebro humano

Los seres humanos son capaces de mostrar un repertorio de respuesta a los estímulos mucho más variado de lo que cabe esperar en otros animales menos evolucionados. En los humanos, los estímulos no evocan de manera inmediata y automática un conjunto de comportamientos limitados, sino que sus respuestas pueden ser aplazadas, inhibidas o memorizadas para utilizarse posteriormente en escenarios completamente nuevos.

Estudiar el modo en que el cerebro humano ha evolucionado resulta útil para conocer también su funcionamiento, especialmente en lo relativo al origen de las emociones. Por ello incluimos a continuación una serie de citas que desarrollan este tema.

«La región más primitiva del cerebro, una región que compartimos con todas aquellas especies que sólo disponen de un rudimentario sistema nervioso, es el tallo encefálico, que se halla en la parte superior de la médula espinal. Este cerebro rudimentario regula las funciones vitales básicas, como la respiración, el metabolismo de los otros órganos corporales y las reacciones y movimientos automáticos» [19].

No se puede afirmar que este cerebro primitivo fuera capaz de pensar o aprender, puesto que su función se limitaba a regular el funcionamiento del cuerpo y promover la supervivencia del individuo. Este es el cerebro propio de los reptiles [19].

«De este cerebro primitivo —el tallo encefálico— emergieron los centros emocionales que, millones de años más tarde, dieron lugar al cerebro pensante —o "neocórtex"— ese gran bulbo de tejidos replegados sobre sí que configuran el estrato superior del sistema nervioso. El hecho de que el cerebro emocional sea muy anterior al racional y que éste sea una derivación de aquél, revela con claridad las auténticas relaciones existentes entre el pensamiento y el sentimiento» [19].

«Con la aparición de los primeros mamíferos emergieron también nuevos estratos fundamentales en el cerebro emocional. Estos estratos rodearon al tallo encefálico a modo de una rosquilla en cuyo hueco se aloja el tallo encefálico. A esta parte del cerebro que envuelve y rodea al tallo encefálico se le denominó

sistema "límbico", un término derivado del latín *limbus*, que significa "anillo". Este nuevo territorio neural agregó las emociones propiamente dichas al repertorio de respuestas del cerebro» [19].

«La evolución del sistema límbico puso a punto dos poderosas herramientas: el aprendizaje y la memoria, dos avances realmente revolucionarios que permitieron ir más allá de las reacciones automáticas predeterminadas y afinar las respuestas para adaptarlas a las cambiantes exigencias del medio, favoreciendo así una toma de decisiones mucho más inteligente para la supervivencia» [19].

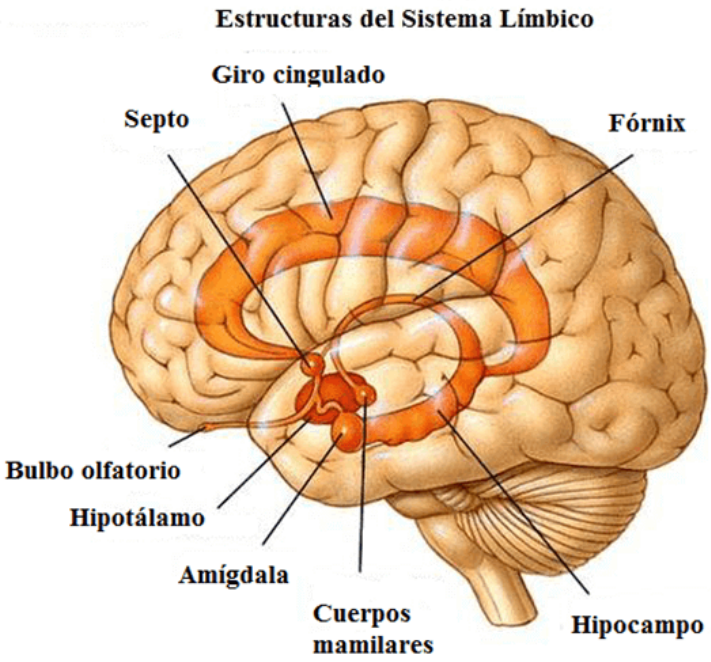


Figura 2.1: Sistema límbico

«En los humanos, el conjunto de estructuras que se conocen como sistema límbico (Figura 2.1), y más precisamente el circuito mesolímbico cortical, tienen importancia en el origen y el control de las emociones. El sistema límbico está localizado inmediatamente debajo de la corteza cerebral. Está formado por varias estructuras entre las que destacan el tálamo, el hipotálamo, el hipocampo y la amígdala» [35].

«Hace unos cien millones de años, el cerebro de los mamíferos experimentó una transformación radical que supuso otro extraordinario paso adelante en el desarrollo del intelecto, y sobre el delgado córtex de dos estratos se asentaron los nuevos estratos de células cerebrales que terminaron configurando el neocórtex» [19].

«El neocórtex del *Homo sapiens*, mucho mayor que el de cualquier otra especie, ha traído consigo todo lo que es característicamente humano. El neocórtex es el asiento del pensamiento y de los centros que integran y procesan los datos registrados por los sentidos. A su vez, agregó al sentimiento nuestra reflexión sobre él y nos permitió tener sentimientos sobre las ideas, el arte, los símbolos y las imágenes» [19].

«Pero el hecho es que estos centros superiores no gobiernan la totalidad de la vida emocional porque, en los asuntos decisivos del corazón —y, más especialmente, en las situaciones emocionalmente críticas—, bien podríamos decir que delegan su cometido en el sistema límbico. Las ramificaciones nerviosas que extendieron el alcance de la zona límbica son tantas, que el cerebro emocional sigue desempeñando un papel fundamental en la arquitectura de nuestro sistema nervioso. La región emocional es el sustrato en el que creció y se desarrolló nuestro nuevo cerebro pensante y sigue estando estrechamente vinculada con él por miles de circuitos neuronales. Esto es precisamente lo que confiere a los centros de la emoción un poder extraordinario para influir en el funcionamiento global del cerebro (incluyendo, por cierto, a los centros del pensamiento)» [19].

2.3. La amígdala y las reacciones automáticas

«La raíz más primitiva de nuestra vida emocional radica en el sentido del olfato o, más precisamente, en el lóbulo olfatorio, ese conglomerado celular que se ocupa de registrar y analizar los olores. En (...) tiempos remotos el olfato fue un órgano sensorial clave para la supervivencia, porque cada entidad viva, ya sea alimento, veneno, pareja sexual, predador o presa, posee una identificación molecular característica que puede ser transportada por el viento» [19].

«El hipocampo y la amígdala fueron dos piezas clave del primitivo "cerebro olfativo" que, a lo largo del proceso evolutivo, terminó dando origen al córtex y posteriormente al neocórtex. La amígdala está especializada en las cuestiones emocionales y en la actualidad se considera como una estructura límbica muy ligada a los procesos del aprendizaje y la memoria» [19].

La amígdala recibe información a través de sus conexiones con prácticamente cada tramo de las rutas de procesamiento sensorial, incluyendo el tálamo, el hipocampo, etc. Esta gran capacidad receptiva hace que la amígdala reciba información sensorial con cualquier grado de procesamiento y permite que el organismo responda de manera «preconsciente» incluso a estímulos que no han sido aún procesados por el neocórtex, el cerebro pensante [36].

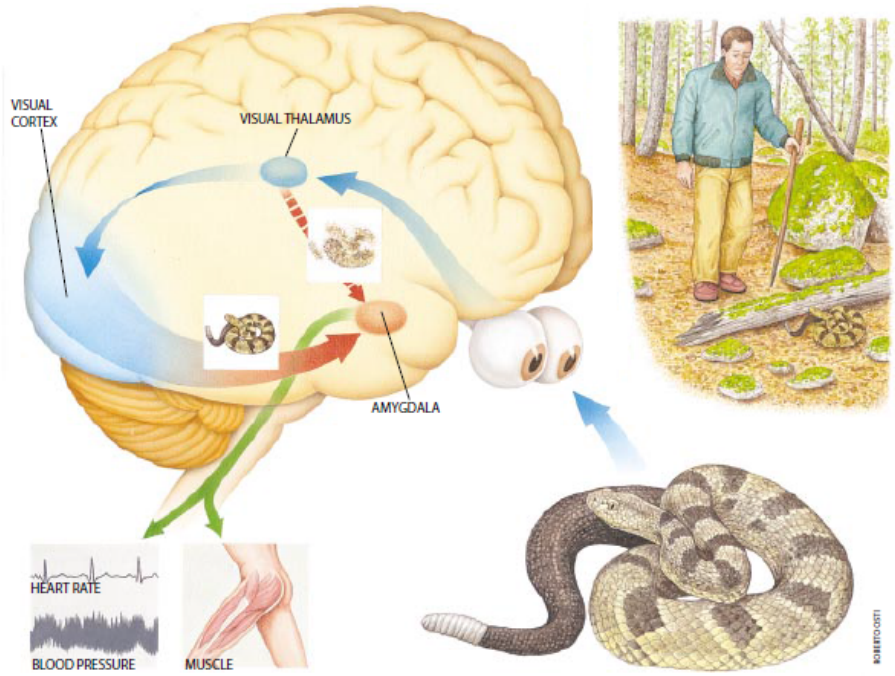


Figura 2.2: Ejemplo de respuesta preconsciente. Imagen obtenida de [37].

Un ejemplo que se suele utilizar para ilustrar este concepto es un hipotético encuentro con una serpiente, tal como lo describe el neurocientífico Joseph LeDoux [38]: supongamos que usted va caminando por el bosque. De pronto, escucha un leve crujido al tiempo que atisba un objeto esbelto y curvo en el camino frente a usted. Antes de que tenga oportunidad de pensar o examinar el objeto, detiene sus pasos al congelarse los músculos de sus piernas, brazos, torso y cuello. Mientras se encuentra inmóvil, la información neuronal correspondiente a la burda imagen visual del objeto alargado viaja a lo largo del nervio óptico, y en unos cuantos milisegundos llega a la corteza visual primaria. Una vez ahí, la información es dividida en componentes más básicos: líneas, ángulos, colores y movimiento, proceso que tarda unos cuantos milisegundos más. A continuación, esta información procesada viaja a través del relativamente lento conducto cortical hacia la corteza visual secundaria, donde se sintetiza como una forma unificada, todavía sin un significado claro. Varias decenas de milisegundos después, la información alcanza las áreas terciarias de asociación, donde la imagen se convierte en una serpiente, asociada con la palabra «serpiente» y con todas las memorias pasadas y el conocimiento de lo que las serpientes representan. En total, pueden pasar alrededor de 100-200 milisegundos desde su reacción inicial hasta el momento en que usted ha procesado completamente la imagen y averiguado si lo que vio era en realidad una serpiente o solamente una rama en medio del camino. Tomará más tiempo aún reaccionar conductualmente a lo

que acaba de pasar, es decir, abandonar su estado de inmovilización y actuar de manera más razonada [38].

Este «corto circuito» en el cerebro puede producir respuestas inadecuadas por parte de una persona en situaciones que requieren un mayor grado de reflexión, pero presenta una ventaja evolutiva que LeDoux resume del siguiente modo: «El costo de confundir una rama con una serpiente es menor, a largo plazo, que el costo de confundir una serpiente con una rama» [36].

«LeDoux descubrió el papel privilegiado que desempeña la amígdala en la dinámica cerebral como una especie de centinela emocional capaz de secuestrar al cerebro. (...) La primera estación cerebral por la que pasan las señales sensoriales procedentes de los ojos o de los oídos es el tálamo y, a partir de ahí y a través de una sola sinapsis, la amígdala. Otra vía procedente del tálamo lleva la señal hasta el neocórtex, el cerebro pensante. Esa ramificación permite que la amígdala comience a responder antes de que el neocórtex haya ponderado la información a través de diferentes niveles de circuitos cerebrales, se aperciba plenamente de lo que ocurre y finalmente emita una respuesta más adaptada a la situación» [19].

«LeDoux destruyó el córtex auditivo de las ratas y luego las expuso a un sonido que iba acompañado de una descarga eléctrica. Las ratas no tardaron en aprender a temer el sonido aun cuando su neocórtex no llegara a registrarlo. En este caso, el sonido seguía la ruta directa del oído al tálamo y, desde allí, a la amígdala, saltándose todos los circuitos principales. Las ratas, en suma, habían aprendido una reacción emocional sin la menor implicación de las estructuras corticales superiores» [19].

«Otra investigación ha demostrado que, durante los primeros milisegundos de cualquier percepción, no sólo sabemos inconscientemente de qué se trata sino que también decidimos si nos gusta o nos desagrada. De este modo, nuestro "inconsciente cognitivo" no sólo presenta a nuestra conciencia la identidad de lo que vemos sino que también le ofrece nuestra propia opinión al respecto. Nuestras emociones tienen una mente propia, una mente cuyas conclusiones pueden ser completamente distintas a las sostenidas por nuestra mente racional» [19].
«Una de las funciones de la amígdala consiste en escudriñar las percepciones en busca de alguna clase de amenaza. De este modo, la amígdala se convierte en un importante vigía de la vida mental, una especie de centinela psicológico que afronta toda situación, toda percepción, considerando una sola cuestión, la más primitiva de todas: "¿Es algo que odio? ¿Que me pueda herir? ¿A lo que temo?" En el caso de que la respuesta a esta pregunta sea afirmativa, la amígdala reaccionará al momento poniendo en funcionamiento todos sus recursos neurales y enviando un mensaje urgente a todas las regiones del cerebro» [19].
«En el caso de que, por ejemplo, suene la alarma de miedo, la amígdala envía mensajes urgentes a cada uno de los centros fundamentales del cerebro, disparando la secreción de las hormonas corporales que predisponen a la lucha o a

la huida, activando los centros del movimiento y estimulando el sistema cardiovascular, los músculos y las vísceras: La amígdala también es la encargada de activar la secreción de dosis masivas de noradrenalina, la hormona que aumenta la reactividad de ciertas regiones cerebrales clave, entre las que destacan aquéllas que estimulan los sentidos y ponen el cerebro en estado de alerta. Otras señales adicionales procedentes de la amígdala también se encargan de que el tallo encefálico inmovilice el rostro en una expresión de miedo, paralizando al mismo tiempo aquellos músculos que no tengan que ver con la situación, aumentando la frecuencia cardiaca y la tensión sanguínea y haciendo más lenta la respiración. Otras señales de la amígdala dirigen la atención hacia la fuente del miedo y predisponen a los músculos para reaccionar en consecuencia. Simultáneamente los sistemas de la memoria cortical se imponen sobre cualquier otra faceta de pensamiento en un intento de recuperar todo conocimiento que resulte relevante para la emergencia presente» [19].

Estudios sobre lesiones cerebrales y análisis de resonancia magnética funcional del cerebro han demostrado el papel preponderante de la amígdala en la detección de estímulos con carga emocional. La amígdala se activa claramente en respuesta a estímulos inherentemente emocionales, tales como las arañas, expresiones faciales, movimientos corporales expresivos, así como estímulos que han recibido su carga emocional a través de condicionamientos relacionados con el miedo. Dichas activaciones ocurren incluso cuando los estímulos se presentan de manera subliminal, es decir, sin que los participantes acusen un registro consciente [36].

Vale la pena hacer notar que el papel de la amígdala en el condicionamiento del miedo no se limita a experiencias personales, también incluye aprendizaje social y a través del ejemplo. En otras palabras, la amígdala también responde a experimentos llevados a cabo en otros individuos, así como a explicaciones verbales de relaciones entre peligros potenciales y estímulos emocionalmente neutrales. Este aprendizaje vicario nos permite tener respuestas emocionales apropiadas, por ejemplo, a que nos apunten con un arma, sin necesidad de tener la experiencia previa de haber recibido un disparo. Sin embargo, no todo el aprendizaje afectivo se basa únicamente en la amígdala. Ésta tiene conexiones con el hipocampo, lo cual sugiere una modulación mutua. Específicamente, mientras que la amígdala es sensible a estímulos discretos, el hipocampo es sensible a contextos con significado emotivo [36].

En resumen, la amígdala es relevante para el estudio de las emociones porque es necesaria para detectar el contenido emocional de un estímulo y suficiente para la detección de estímulos emocionales, puesto que se activa incluso en presencia de estímulos que se presentan de manera subliminal [36].

«El cerebro utiliza un método simple pero muy ingenioso para registrar con especial intensidad los recuerdos emocionales, ya que los mismos sistemas de alerta neuroquímicos que preparan al cuerpo para reaccionar ante cualquier

amenaza —luchando o escapando— también se encargan de grabar vívidamente este momento en la memoria. En caso de estrés o de ansiedad, o incluso en el caso de una intensa alegría, un nervio que conecta el cerebro con las glándulas suprarrenales (situadas encima de los riñones) estimula la secreción de las hormonas adrenalina y noradrenalina, disponiendo así al cuerpo para responder ante una urgencia. Estas hormonas activan determinados receptores del nervio vago, encargado, entre otras muchas cosas, de transmitir los mensajes procedentes del cerebro que regulan la actividad cardiaca y, a su vez, devuelve señales al cerebro, activado también por estas mismas hormonas. El principal receptor de este tipo de señales son las neuronas de la amígdala que, una vez activadas, se ocupan de que otras regiones cerebrales fortalezcan el recuerdo de lo que está ocurriendo. Esta activación de la amígdala parece provocar una intensificación emocional que también profundiza la grabación de esas situaciones. Este es el motivo por el cual, por ejemplo, recordamos a dónde fuimos en nuestra primera cita o qué estábamos haciendo cuando oímos la noticia de la explosión del *Challenger*. Cuanto más intensa es la activación de la amígdala, más profunda es la impronta y más indeleble la huella que dejan en nosotros las experiencias que nos han asustado o nos han emocionado. Esto significa, en efecto, que el cerebro dispone de dos sistemas de registro, uno para los hechos ordinarios y otro para los recuerdos con una intensa carga emocional, algo que tiene un gran interés desde el punto de vista evolutivo porque garantiza que los animales tengan recuerdos particularmente vívidos de lo que les amenaza y de lo que les agrada. Pero, además de todo lo que acabamos de ver, los recuerdos emocionales pueden llegar a convertirse en falsas guías de acción para el momento presente» [19].

«LeDoux ha estudiado el papel desempeñado por la amígdala en la infancia y ha llegado a una conclusión que parece respaldar uno de los principios fundamentales del pensamiento psicoanalítico, es decir, que la interacción —los encuentros y desencuentros— entre el niño y sus cuidadores durante los primeros años de vida constituye un auténtico aprendizaje emocional. En opinión de LeDoux, este aprendizaje emocional es tan poderoso y resulta tan difícil de comprender para el adulto porque está grabado en la amígdala con la impronta tosca y no verbal propia de la vida emocional. Estas primeras lecciones emocionales se impartieron en un tiempo en el que el niño todavía carecía de palabras y, en consecuencia, cuando se reactiva el correspondiente recuerdo emocional en la vida adulta, no existen pensamientos articulados sobre la respuesta que debemos tomar. El motivo que explica el desconcierto ante nuestros propios estallidos emocionales es que suelen datar de un período tan temprano que las cosas nos desconcertaban y ni siquiera disponíamos de palabras para comprender lo que sucedía. Nuestros sentimientos tal vez sean caóticos, pero las palabras con las que nos referimos a esos recuerdos no lo son» [19].

«En circunstancias así, el atajo que va desde el ojo —o el oído— hasta el tálamo y la amígdala resulta crucial porque nos proporciona un tiempo precioso cuando la proximidad del peligro exige de nosotros una respuesta inmediata» [19].

La fase inicial de una respuesta emocional puede reducirse esencialmente a una reacción motora involuntaria que consiste en dos partes: esquelética y endócrina. La respuesta esquelética involuntaria se refiere a un conjunto de activaciones de los músculos esqueléticos —distinta para cada especie— que involucran cambios rápidos e involuntarios de expresiones faciales, vocalización, postura y otros movimientos corporales, que cooperan a la supervivencia del individuo o de un grupo social. Ejemplos de este tipo de respuestas son el gruñido de un perro o el modo en que mueve la cola, el ronroneo de un gato, o la risa y el llanto en los humanos [36].

La respuesta endócrina, por otro lado, se refiere a los cambios hormonales del cuerpo que complementan y facilitan la respuesta de pelea o huida y también los procesos de digestión y reparación [36].

Tabla 2.1: Efectos del sistema nervioso autónomo simpático y parasimpático en distintos órganos [36].

Órgano	**Rama simpática**	**Rama parasimpática**
Glándula adrenal	Adrenalina en el flujo sanguíneo	N/A
Músculos piloerectores	Contracción y erección de los cabellos	N/A
Vasos sanguíneos en la piel	Constricción	N/A
Vasos sanguíneos de los músculos	Dilatación	N/A
Bronquios	Dilatación	Constricción
Ojos	Dilatación	Constricción
Contracción cardiaca	Aumento	Disminución
Ritmo cardiaco	Aumento	Disminución
Ritmo respiratorio	Aumento	Disminución
Glándulas sudoríparas	Aumento de secreción	N/A

El sistema nervioso autónomo (SNA) controla los músculos lisos y se encuentra fuera del influjo consciente o voluntario. Tiene tres ramas: simpática, parasimpática y entérica (que regula los músculos del tracto digestivo). La activación relativa de la rama simpática con respecto a la parasimpática refleja es un indicador universal, aunque no específico, de una respuesta emocional [36]. Algunos efectos de ambos sistemas en los órganos del cuerpo pueden verse en las Tablas 2.1 y 2.2.

Tabla 2.2: Índices de cambio en el sistema autónomo de acuerdo a seis emociones básicas [36].

	Ritmo cardiaco aumenta		**Conductancia de la piel aumenta**	
	Levenson *et al.*	Hamm *et al.*	Levenson *et al.*	Hamm *et al.*
Alegría	Media	Baja	Baja	Media
Tristeza	Alta	Media	Alta	Baja
Miedo	Alta	Media	Alta	Alta
Enojo	Alta	Alta	Media	Baja
Disgusto	Baja	Alta	Alta	Alta
Sorpresa	Media	Alta	Baja	Alta

2.4. Regulación de las emociones humanas

Es muy importante recalcar que aunque el ser humano no puede controlar voluntariamente el sistema motor para generar expresiones emocionales totalmente auténticas, sí puede suprimirlas parcialmente o retrasarlas. Pero también existen estudios que nos muestran que la mayor parte de las personas son incapaces de suprimir completamente expresiones emocionales auténticas, de manera que algunos signos sutiles de la emoción subyacente pueden ser detectados por un observador intuitivo o calificado [36].

«El regulador cerebral que desconecta los impulsos de la amígdala parece encontrarse en el otro extremo de una de las principales vías nerviosas que van al neocórtex, en el lóbulo prefrontal, que se halla inmediatamente detrás de la frente. El córtex prefrontal parece ponerse en funcionamiento cuando alguien tiene miedo o está enojado pero sofoca o controla el sentimiento para afrontar de un modo más eficaz la situación presente o cuando una evaluación posterior exige una respuesta completamente diferente (...). De este modo, el área prefrontal constituye una especie de modulador de las respuestas proporcionadas por la amígdala y otras regiones del sistema límbico, permitiendo la emisión de una respuesta más analítica y proporcionada» [19].

«Nuestra respuesta correspondiente la coordinan los lóbulos prefrontales, la sede de la planificación y de la organización de acciones tendentes a un objetivo determinado, incluyendo las acciones emocionales. En el neocórtex, una serie de circuitos registra y analiza esta información, la comprende y organiza gracias a los lóbulos prefrontales, y si, a lo largo de ese proceso, se requiere una respuesta emocional, es el lóbulo prefrontal quien la dicta, trabajando en equipo con la amígdala y otros circuitos del cerebro emocional. Este suele ser el proceso normal de elaboración de una respuesta, un proceso que —con la sola

excepción de las urgencias emocionales— tiene en cuenta el discernimiento» [19].

«El secuestro emocional parece implicar dos dinámicas distintas: la activación de la amígdala y el fracaso en activar los procesos neocorticales que suelen mantener equilibradas nuestras respuestas emocionales. En esos momentos, la mente racional se ve desbordada por la mente emocional y lo mismo ocurre con la función del córtex prefrontal como un gestor eficaz de las emociones sopesando las reacciones antes de actuar y amortiguando las señales de activación enviadas por la amígdala y otros centros límbicos, como un padre que impide que su hijo se comporte arrebatando todo lo que quiere y le enseña a pedirlo (o a ser paciente). El interruptor que "apaga" la emoción perturbadora parece hallarse en el lóbulo prefrontal izquierdo. Los neurofisiólogos que han estudiado los estados de ánimo de pacientes con lesiones en el lóbulo prefrontal han llegado a la conclusión de que una de las funciones del lóbulo prefrontal izquierdo consiste en actuar como una especie de termostato neural que regula las emociones desagradables. Así pues, el lóbulo prefrontal derecho es la sede de sentimientos negativos como el miedo y la agresividad, mientras que el lóbulo prefrontal izquierdo los tiene a raya. muy probablemente inhibiendo el lóbulo derecho. En un determinado estudio, por ejemplo, los pacientes con lesiones en el córtex prefrontal izquierdo eran proclives a experimentar miedos y preocupaciones catastrofistas mientras que aquéllos otros con lesiones en el córtex prefrontal derecho eran "desproporcionadamente joviales", bromeaban continuamente durante las pruebas neurológicas y estaban tan despreocupados que no ponían el menor cuidado en lo que estaban haciendo» [19].

«Estas averiguaciones condujeron al doctor Damasio a la conclusión contraria a la intuición de que los sentimientos son indispensables para la toma racional de decisiones, porque nos orientan en la dirección adecuada para sacar el mejor provecho a las posibilidades que nos ofrece la fría lógica. Mientras que el mundo suele presentarnos un desbordante despliegue de posibilidades (¿En qué debería invertir los ahorros de mi jubilación? ¿Con quién debería casarme?), el aprendizaje emocional que la vida nos ha proporcionado nos ayuda a eliminar ciertas opciones y a destacar otras. Es así como —arguye el doctor Damasio— el cerebro emocional se halla tan implicado en el razonamiento como lo está el cerebro pensante. Las emociones, pues, son importantes para el ejercicio de la razón. En la danza entre el sentir y el pensar, la emoción guía nuestras decisiones instante tras instante, trabajando mano a mano con la mente racional y capacitando —o incapacitando— al pensamiento mismo» [19].

2.5. Manifestaciones fisiológicas de la emoción

En esta sección mencionamos algunas manifestaciones relacionadas con la emoción que son susceptibles de ser medidas físicamente, también hacemos referencia a dos dimensiones de la emoción humana: la valencia y la activación[2].

[2]Traducción de *valence* y *arousal*.

Se entiende como valencia el atractivo intrínseco (valencia positiva) o la aversión (valencia negativa) que produce un evento, objeto o situación [39], mientras que la activación expresa el grado de entusiasmo o agitación [40]. Aunque se sigue discutiendo la importancia relativa de la valencia y la activación, los investigadores concuerdan en afirmar que estos parámetros son los fundamentos de la emoción [26].

Comúnmente se asume que las emociones activan el sistema nervioso autónomo (SNA). Esto se expresa en varias alteraciones fisiológicas frecuentemente estimuladas a través del SNA: ritmo cardiaco, presión sanguínea, actividad muscular, etc. La principal ventaja de incluir este tipo de variables en análisis de emociones radica en que no pueden ser controladas conscientemente por el individuo [41].

2.5.1. Músculo corrugador

Los músculos corrugadores son los responsables del descenso y contracción de las cejas, una expresión facial asociada con el malestar. Así, es de esperarse un número significativo de activaciones motoras en este músculo al percibir una imagen o sonido considerados como desagradables, incluso si el grado de activación es insuficiente para producir un movimiento visible de las cejas [26]. En la Figura 2.3a se puede apreciar una gráfica que muestra la actividad del corrugador cuando el individuo experimenta sensaciones agradables. Evidentemente, el corrugador no es el único músculo facial que se activa por medio de estímulos emocionales, pero aquí se especifica a manera de ejemplo representativo.

2.5.2. Ritmo cardiaco

Al ver ciertas imágenes diseñadas para producir una sensación agradable, el ritmo cardiaco de una persona suele seguir un patrón de tres fases: desaceleración inicial, aceleración y desaceleración secundaria (ver Figura 2.3b). Los estímulos negativos producen una mayor desaceleración, mientras que los positivos presentan mayores picos de aceleración. También es posible encontrar mayor desaceleración cardiaca frente a imágenes desagradables [26].

2.5.3. Conductancia de la piel

La actividad electrodérmica es un indicador de excitación. Se piensa que se encuentra inervada únicamente por el sistema nervioso simpático, cuyo estado de activación puede detectarse a grandes rasgos por este medio. Algunos estudios muestran que la conductancia de la piel se incrementa linealmente con respecto a la activación [26], sin importar la valencia atribuida al estímulo (ver Figura 2.3c).

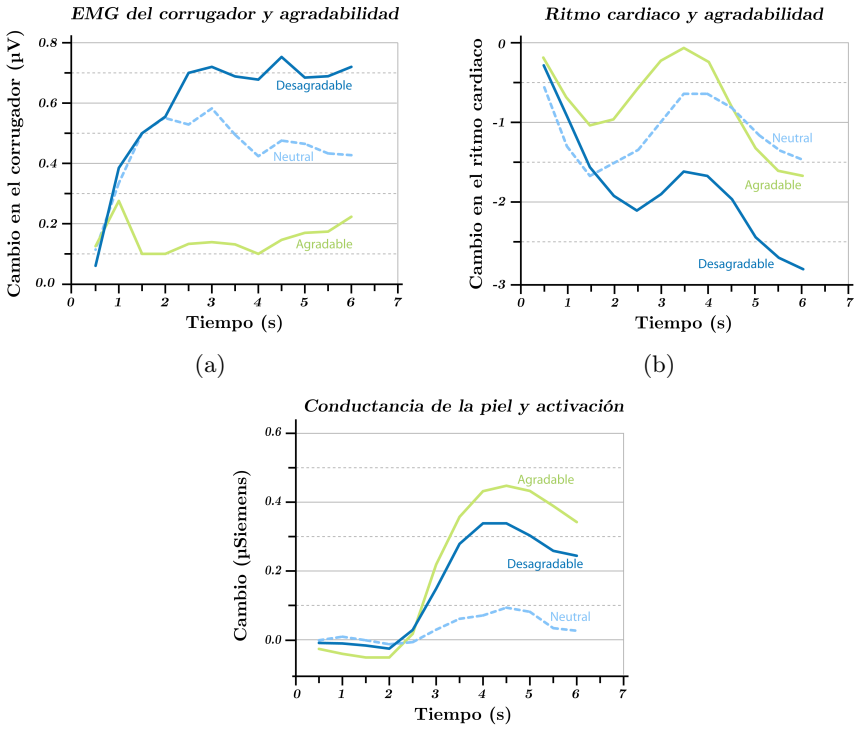


Figura 2.3: Los patrones de respuesta fisiológica asociados a la visualización de imágenes seleccionadas para producir emociones de determinada valencia demuestran que la actividad electromiográfica del músculo corrugador facial (a) y el ritmo cardiaco (b) varían en función de la valencia asociada a cada imagen, mientras que la conductancia de la piel (c) varía de acuerdo a la activación producida (Imágenes adaptadas de [26]).

2.6. Teoría de Paul Ekman sobre la emoción

Paul Ekman es un psicólogo que ha influido de manera decisiva en el estudio de las emociones y sus expresiones faciales asociadas. Postula que dichas expresiones faciales no son producto de un aprendizaje cultural, sino que tienen un origen biológico y evolutivo. A pesar de que sus conclusiones no han convencido a todos los estudiosos del tema, una buena parte de los trabajos realizados sobre reconocimiento de emociones (RE) se apoya en ellas.

Uno de sus experimentos que realizó Ekman consistió en mostrar fotografías de expresiones faciales a personas de distintos países (Chile, Argentina, Brasil, Japón y Estados Unidos) y pedirles que juzgaran la emoción expresada en cada una. Las respuestas obtenidas en los cinco países fueron prácticamente iguales, sugiriendo que las expresiones faciales son universales. En otros estudios llegó a la conclusión de que hay una serie de reglas —distintas en cada cultura— que dictan cuándo los seres humanos exageran, disminuyen o esconden la expresión de sus emociones frente a otras personas, pero las emociones que sus rostros expresan en privado —donde no se aplican estas reglas— sí coinciden de manera universal. Lo mismo encontró entre grupos de ciegos congénitos: su repertorio de expresiones faciales no pudo ser aprendido por imitación y sin embargo concuerda con el del resto del mundo. Los resultados de sus experimentos apuntan a que las emociones de alegría, enojo, disgusto y tristeza son claramente separables, mientras que el miedo y la sorpresa son más difíciles de distinguir entre sí, aunque no con respecto a las anteriores [42].

En 1978 publicó el sistema de codificación de acciones faciales (FACS por sus siglas en inglés): un estándar común para clasificar sistemáticamente las expresiones faciales a través del análisis de la activación de ciertos músculos del rostro (llamadas unidades de acción), que se sigue utilizando ampliamente al día de hoy [43], sin embargo, la utilización de este sistema requiere una cantidad considerable de tiempo de análisis, además de una capacitación extensa.

2.7. Herramientas

En esta sección se explican algunos instrumentos comúnmente utilizados para evaluar distintas condiciones físicas asociadas con la emoción.

2.7.1. Termografía infrarroja

Las cámaras de video suelen registrar ondas electromagnéticas dentro del espectro visible, es decir, longitudes de onda entre 0.4 y 0.7 μm aproximadamente, pero una cámara infrarroja puede percibir radiación entre 0.7 y 14.0 μm, lo que le permite medir con precisión la temperatura de un rostro sin utilizar cables o electrodos, ya que la temperatura de un objeto está directamente relacionada con la cantidad de radiación infrarroja que emite (ver Figura 2.4). Por otro lado,

la evidencia existente [44] sugiere que las respuestas emocionales se manifiestan como distintos patrones de calor en la cara.

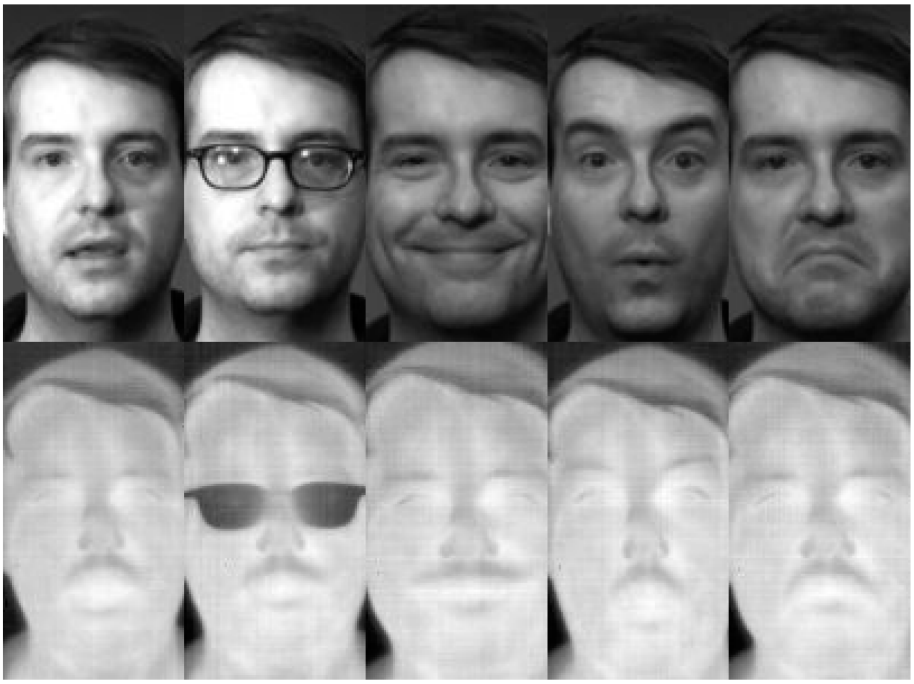


Figura 2.4: Ejemplo de imágenes en espectro visible y su correspondiente emisión infrarroja.

2.7.2. Electroencefalogramas

Se conoce como electroencefalograma (EEG) al registro de las señales eléctricas producidas por la acción conjunta de las células del cerebro. Dicha actividad puede medirse por medio de electrodos colocados sobre el cuero cabelludo [45] y esto permite detectar patrones asociados a distintas funciones del cerebro como la atención y el sueño. También existen muchos estudios sobre reconocimiento de emociones por este medio, una revisión de la literatura relativa a este tema puede encontrarse en [5].

2.7.3. Actividad cardiaca

La actividad cardiaca es un fenómeno fácil de medir, y ha sido estudiado a profundidad por parte de la comunidad médica. El ritmo cardiaco y su variabilidad se han utilizado como medida para algunas emociones como miedo y enojo [46]. La presión sanguínea y el ritmo cardiaco pueden registrarse a través de un dispositivo llamado pletismógrafo. Un pletismógrafo tradicional utiliza

Figura 2.5: Paciente con electrodos para lectura de EEG.

un extensómetro o galga extensométrica dentro de una banda que la presiona contra la vena, misma que cambia de volumen con la presión sanguínea. El extensómetro de la banda traduce este cambio de volumen en un cambio de resistencia eléctrica que puede medirse con facilidad por medio de un circuito sencillo. En el caso del pletismógrafo fotoeléctrico (Figura 2.6), se hace incidir luz infrarroja en el flujo sanguíneo de arterias y venas subcutáneas. Los tejidos absorben la mayor parte de esta radiación, pero entre un 5 y un 10 % alcanza los vasos subcutáneos. La magnitud de luz reflejada depende de la densidad de glóbulos rojos contenidos en su interior y se registra por medio de un fotosensor para luego ser amplificada y convertida en un diferencial de voltaje [47]. No es posible calibrar un pletismógrafo fotoeléctrico y por ello sus medidas no tienen una unidad asignada [48].

2.7.4. Respuesta galvánica de la piel

Es una medida de la conductividad de la piel. La estimulación del SNA a través de las emociones influye en las glándulas sudoríparas para que produzcan más sudor. Esto se traduce en un incremento de la conductividad en la piel, misma que puede ser cuantificada por medio de un circuito eléctrico relativamente sencillo [41].

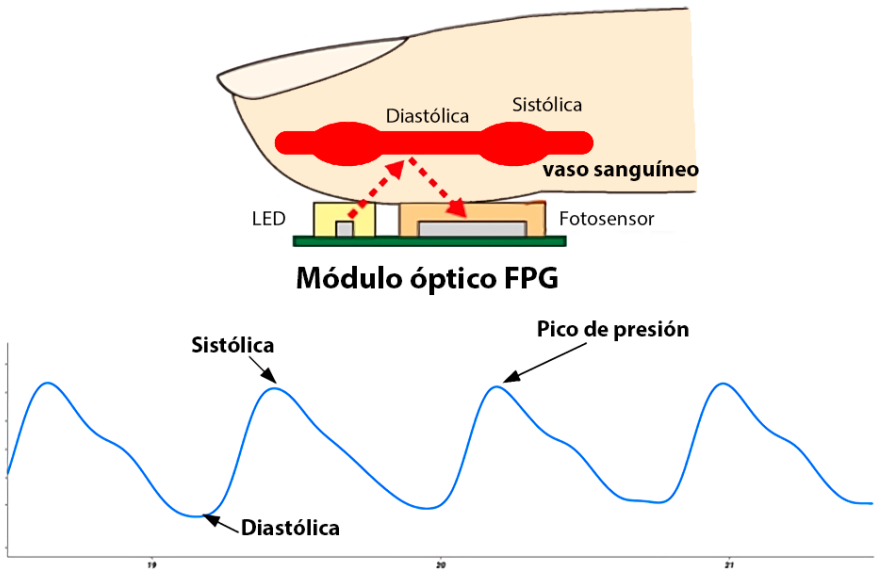


Figura 2.6: Principio de funcionamiento del pletismógrafo fotoeléctrico.

2.7.5. Expresión verbal

También a través del habla se pueden expresar emociones de manera consciente o inconsciente y existen sistemas informáticos capaces de clasificar estas emociones. El primer paso consiste en extraer características acústicas o lingüísticas del audio digitalizado. Las características acústicas pueden ser: entonación, intensidad, perturbaciones, etc. Entre los rasgos lingüísticos detectables se cuentan los fonemas y palabras; también se pueden tomar en cuenta signos paralingüísticos como suspiros o risas. Una vez extraídas las características apropiadas, suelen alimentarse a una inteligencia artificial para su clasificación [49].

2.7.6. Seguimiento ocular

Existen muchos tipos distintos de dispositivos capaces de monitorear movimientos oculares. Uno de los métodos más comúnmente utilizados para determinar el punto en el que una persona fija la mirada consiste en registrar mediante cámaras de video la posición de la pupila y el reflejo de luz infrarroja en la córnea (ver Figura 2.7) [50]. El seguimiento ocular también se ha utilizado en el reconocimiento de emociones [51], muchas veces combinado con algún tipo de inteligencia artificial [52]. Por ejemplo, la dilatación de la pupila puede estar vinculada a una activación del sistema nervioso simpático y un parpadeo puede indicar emociones defensivas.

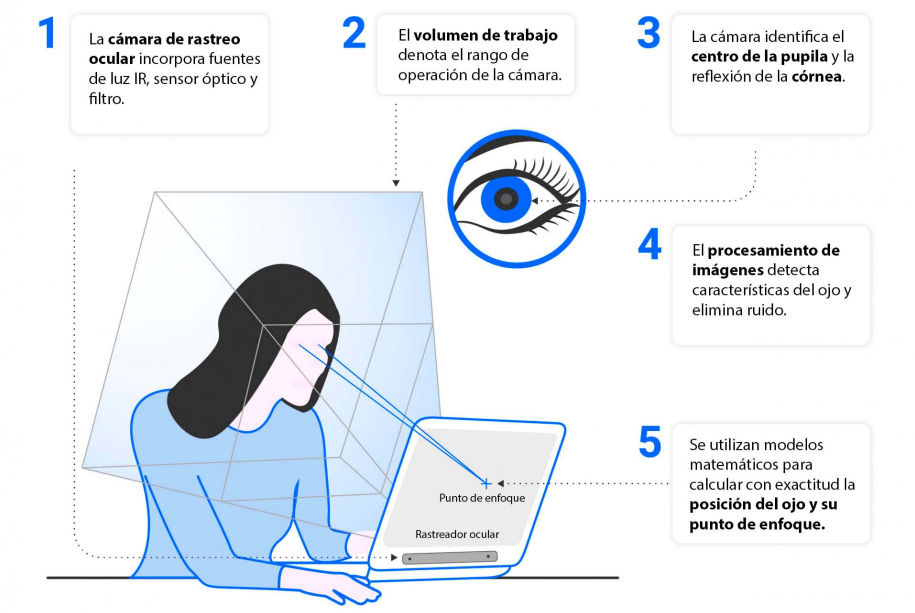


Figura 2.7: Esquema de funcionamiento de un seguidor ocular.

2.7.7. *Kinect*

El dispositivo *Kinect* de *Microsoft* fue concebido como un instrumento para interactuar con videojuegos por medio de movimientos corporales —dejando de lado los controladores tradicionales operados por botones y palancas— con intención de ofrecer una experiencia más cercana a la realidad, pero debido a su amplia disponibilidad en el mercado y su costo relativamente bajo, muchos investigadores en campos como ingeniería electrónica y ciencias computacionales lo han empleado para realizar tareas tan diversas como ayudar a niños con autismo [53] asistir a cirujanos en salas de operaciones [54]. El aparato contiene varios sensores avanzados: cuatro micrófonos, cámara a color y sensor de profundidad, que lo hacen muy adecuado para captura de movimiento corporal en 3D, reconocimiento facial, registro de gestos, entre otras posibilidades [55].

2.7.8. Aprendizaje profundo

Kaplan y Haenlein [56] definen la inteligencia artificial como la habilidad de un sistema para interpretar correctamente datos externos, aprender de ellos y utilizar ese aprendizaje para lograr ciertas metas y realizar tareas a través de una adaptación flexible. Algunos sostienen que en un futuro existirá software capaz de imitar o incluso superar la capacidad intelectual del ser humano. En cualquier caso, las últimas décadas han producido avances considerables en el campo de la

Figura 2.8: Dispositivo *Kinect*.

rama de la inteligencia artificial conocida como aprendizaje automático: existen hoy aplicaciones concretas fundamentadas en la capacidad de las computadoras para «aprender» ciertas características comunes a los elementos de un conjunto de entrenamiento e identificarlas después en elementos similares fuera del conjunto. Probablemente la técnica de aprendizaje automático que ha cobrado mayor popularidad en los últimos años sea el llamado aprendizaje profundo, que consiste en el empleo de varias capas de redes neuronales en tareas de clasificación o generación de imágenes. Por ejemplo, el aprendizaje profundo hace que sea posible alimentar un programa computacional con una imagen digital y que éste determine si la imagen contiene autos, personas, árboles, animales, o todos al mismo tiempo. Es por esta razón que el aprendizaje profundo se ha utilizado también para identificación de rostros y reconocimiento de expresiones faciales. Un ejemplo de software aplicado al reconocimiento de expresiones faciales es *FaceReader* [57], que utiliza también redes neuronales, además de modelos activos de apariencia. La empresa que lo comercializa reporta un 89 % de efectividad en la clasificación de emociones. El Capítulo 3 contiene una explicación más detallada de las redes neuronales y el aprendizaje profundo.

2.7.9. Texto y análisis de sentimiento

El análisis de sentimiento o minería de opinión es una disciplina que estudia, entre otras cosas, la manera de clasificar automáticamente grandes cantidades de textos como positivos o negativos en función de las opiniones que expresan sobre un determinado tema [58] o detectar en sus autores sentimientos de enojo, tristeza, etc. Este tipo de análisis se realiza a través de procesamiento de lenguaje natural y algoritmos de inteligencia artificial. Se utiliza frecuentemente para evaluar la percepción de los usuarios de redes sociales, calibrar la reputación de una empresa, predecir tendencias o arrojar luz sobre el modo en que piensan los consumidores potenciales de un producto.

2.8. Estado del arte en reconocimiento de emociones

Chia-Yin Yua *et al.* [59] hicieron experimentos para conocer las reacciones emotivas de algunos voluntarios producidas por imágenes con distintos estilos gráficos y niveles de estilización y concluyeron que estos pueden producir diferencias significativas en mediciones de alegría, enojo, sorpresa y disgusto utilizando *FaceReader*. Rodas *et al.* [6] utilizaron herramientas de *neuromarketing* como seguidores de visión y *FaceReader* para evaluar el impacto emocional y patrones de atención de los consumidores producidos por un video comercial.

Ruiz-García *et al.* [60] investigaron sobre la factibilidad de dotar a robots de asistencia social con capacidades de reconocimiento de emociones en tiempo real, con la idea de facilitar su aceptación por parte de los usuarios a los que buscan ayudar. Alhussein [61] propone un sistema de reconocimiento de emociones para una aplicación de diagnóstico médico a distancia (*eHealthcare*), con el objetivo de incluir el estado emocional del paciente como parte de la información relevante para el diagnóstico.

Candra *et al.* [62] subrayan la importancia que tiene para un psicoterapeuta el identificar las emociones de su paciente a través de expresiones faciales durante las sesiones de acompañamiento, como una herramienta útil para ofrecer un tratamiento óptimo. Por ello, proponen un algoritmo de REF basado en máquinas de vectores de soporte. Maison y Pawłowska [63] se dieron a la tarea de determinar la eficacia de anuncios publicitarios que buscan llamar la atención por medio de imágenes escandalosas o controvertidas, para lo cual se sirvieron —entre otros recursos— del software *FaceReader*.

En un artículo de Gilda *et al.* [64] se describe un reproductor de música afectivo, que ofrece al usuario recomendaciones musicales a partir de sus preferencias determinadas con anterioridad, así como de su estado de ánimo actual, establecido por medio de algoritmos de aprendizaje profundo que toman como entrada imágenes faciales del usuario.

Otra aplicación potencial de los algoritmos de REF es la detección de dolor en pacientes hospitalizados. Aunque normalmente se evalúa por medio de reportes escritos por los mismos pacientes, algunos de ellos pierden la capacidad de comunicarse a causa de sus padecimientos y la supervisión continua por parte del personal médico es inviable, además de estar sujeta a sesgos subjetivos. Entre otros investigadores que han abordado el problema, Hassan *et al.* [65] publicaron una revisión de varios artículos sobre posibles soluciones de REF basadas en visión por computadora.

En el ramo de los juegos de video también existen trabajos que buscan aplicar REF, por ejemplo, como apoyo en el diseño de interfaz para un juego educativo

[66], estudio que también se valió de entradas de audio, texto y patrones de comportamiento del usuario para conocer su estado emocional. Bahreini *et al.* [67] investigaron qué tanto podía ser de ayuda el reconocimiento de emociones por medio de imágenes faciales y registro de voz para mejorar un juego serio abocado al entrenamiento de habilidades comunicativas, mientras que Psaltis *et al.* [68] se sirvieron de pautas de movimiento facial y corporal obtenidas por medio de un sensor *Kinect* para evaluar las distintas emociones que experimentan los jugadores a lo largo de varias escenas de un juego.

Existen también sistemas patentados que utilizan RE uno de ellos busca inferir el estado emocional de distintos clientes que llaman a un centro de soporte, para poder comunicarlos automáticamente con el agente de ventas más adecuado para lograr un mayor grado de satisfacción del cliente [69], mientras que otro [70] se aplica a videoconferencias para detectar si un cliente se encuentra molesto o enojado y poder avisar a su interlocutor para que pueda actuar en consecuencia.

Soroush *et al.* [5] presentan una reseña de varias investigaciones relativas al RE por medio de electroencefalogramas (EEG), método que se utiliza con frecuencia por ser económico y ofrecer una buena resolución. Otras técnicas buscan medir emociones registrando el audio correspondiente a expresiones habladas, Lugović *et al.* [49] produjeron una revisión de varios trabajos en esta línea, con vistas a su aplicación en registro de interacciones con clientes en centros de llamadas o *call centers* o incluso, detección de mentiras.

Las ramas del conocimiento llamadas computación afectiva e interacción hombre-máquina buscan reducir la frustración del usuario, crear aplicaciones capaces de procesar información afectiva y crear herramientas capaces de asistir en el desarrollo humano de capacidades socioemocionales, entre otros objetivos [71], todos los cuales se benefician en buena medida del desarrollo del RE.

También son relevantes los esfuerzos dedicados a desarrollar algoritmos capaces de clasificar las emociones expresadas en un texto: sistemas que buscan determinar los sentimientos de grupos sociales [72] o su opinión, así como algoritmos de recomendación que puedan omitir productos que hayan recibido retroalimentación negativa; aunque suelen clasificarse en rubros un tanto diferentes al RE: minería de opinión y análisis de sentimiento.

Por otra parte, el efecto emocional que puede experimentar un consumidor al probar ciertos productos alimenticios también ha sido objeto de estudio desde el punto de vista del RE. En [14], He *et al.* registraron expresiones faciales de participantes expuestos a olores de naranja y pescado. Leitch *et al.* midieron la respuesta a endulzantes de té a través de una escala hedónica, un cuestionario con términos emocionales y expresiones faciales [73].

Viejo *et al.* evaluaron EEG, ritmo cardiaco, temperatura y expresiones faciales en consumidores de cerveza [7]. Danner *et al.* reportaron la medición de

cambios en el nivel de conductancia y temperatura de la piel, ritmo cardiaco, pulso y expresiones faciales de voluntarios mientras probaban diferentes tipos de jugo de naranja [12]. De manera similar, otros autores han realizado estudios con jamón ahumado [13] y sabores amargos [74].

2.9. Síntesis

La Figura 2.9 presenta una comparación entre los métodos anteriormente citados para medir las posibles manifestaciones fisiológicas de la emoción humana, de acuerdo a nuestra propia valoración.

	Costo captura	Facilidad de aplicación	Facilidad de interpretación	Procesamiento	Cantidad / calidad de información	**TOTAL**
Cuestionarios	5	5	4	4	3	**21**
Electroencefalograma	3	3	1	1	2	**10**
Electromiografía	4	3	4	4	3	**18**
Expresión verbal	4	4	2	3	3	**16**
Expresiones faciales	4	4	3	4	4	**19**
Respuesta galvánica	4	4	4	4	3	**19**
Ritmo cardiaco	4	4	3	5	3	**19**
Seguimiento ocular	3	5	2	4	2	**16**
Temperatura	2	5	3	4	2	**16**
Texto	5	5	2	2	4	**18**

menos conveniente — más conveniente

1 2 3 4 5

Figura 2.9: Cuadro comparativo para métodos de medición.

Los cuestionarios escritos son una herramienta de uso común en análisis sensorial: permiten registrar la información relevante para estudios de este tipo de acuerdo con objetivos claros, terminología común y un marco de trabajo estructurado [75]. La utilización de estos cuestionarios tiene muchas ventajas: son económicos y fáciles de aplicar, pero no son capaces de medir la emoción en el momento en que se produce. Es necesario que quien contesta el cuestionario realice cierta introspección para tratar de identificar conscientemente sus propias emociones, lo cual resta espontaneidad: un elemento clave cuando se busca determinar las emociones inconscientes.

Contamos también con un dispositivo electroencefalográfico relativamente económico y fácil de colocar, aunque algo incómodo para quien lo utiliza. El problema con un electroencefalograma es que la señal obtenida resulta sumamente difícil de decodificar y de relacionar directamente con las emociones básicas que estamos buscando, por lo que no es nada práctico en una primera instancia. Es por eso que, a pesar de haberlo utilizado en los experimentos, fue necesario posponer el análisis de los resultados que obtuvo.

Un análisis electromiográfico exige la colocación de varios electrodos en el rostro del participante. Esto resulta incómodo, pero sobre todo, interfiere con la visión del rostro, lo cual es clave para el reconocimiento de expresiones faciales por medio de una cámara, técnica a la que decidimos dar preferencia.

Encontrar las emociones expresadas a través del habla es posible, pero además de precisar investigación adicional, esta técnica es más apta para detectar disposiciones emocionales estables a lo largo del tiempo que los cambios inmediatos que produce un estímulo puntual.

Para el diagnóstico del estado emocional a través de imágenes faciales contamos con mejores posibilidades: una cámara con la resolución adecuada es fácil de conseguir e instalar, además de ser un método no invasivo de adquisición de datos. La dificultad consiste en clasificar automáticamente las imágenes obtenidas, pero disponemos de los recursos adecuados para crear una red neuronal convolucional que realice la tarea.

Evaluar la conductancia de la piel y el ritmo cardiaco no es complicado y puede dar pistas importantes, aunque un tanto genéricas, sobre la emoción. Los sensores son económicos y solo parcialmente invasivos, pero las señales obtenidas requieren cierto procesamiento para interpretarse.

Al igual que una cámara, un sistema de seguimiento ocular es sencillo en su funcionamiento y no resulta incómodo para los participantes, sin embargo, es poco lo que puede aportar en términos descriptivos: puede decirnos con precisión hacia dónde está mirando una persona, pero no lo que está sintiendo, emocionalmente hablando. Una cámara infrarroja ofrece ventajas similares, pero además de su elevado costo, se requiere una mayor cantidad de trabajo para interpretar

las señales que obtiene. Lo mismo podemos decir del análisis de sentimientos en textos: los algoritmos necesarios no son fáciles de desarrollar y tampoco permiten observar cambios emocionales repentinos.

Por lo anterior, decidimos enfocarnos principalmente en el desarrollo de un sistema de REF por medio de cámaras de espectro visible, apoyándonos también en cuestionarios y sensores de conductancia y ritmo cardiaco a manera de validación y profundización de los resultados.

Capítulo 3

Fundamentos Teóricos

Una gran cantidad de productos nuevos se lanzan al mercado cada año. Sin embargo, mientras unos pocos consiguen permanecer en el gusto del consumidor y generar ingresos proporcionados a la inversión requerida, una alta proporción de los mismos no consigue este objetivo. Si fuera posible conocer la aceptación del consumidor potencial hacia un producto determinado antes de decidir producirlo a gran escala, podrían ahorrarse importantes cantidades de dinero.

Es por esto que muchas veces se busca predecir esa aceptación del consumidor por medio de estudios de mercado. Sin embargo, estos normalmente se apoyan más en respuestas conscientes de *focus groups* que en las emociones básicas que —según estudios de psicólogos y mercadólogos— impulsan las decisiones de compra del consumidor.

Identificar esas emociones y su efecto en los procesos de intención de compra ha sido el objetivo de muchos trabajos de investigación entre los cuales puede contarse el nuestro. Hemos buscado profundizar en la determinación de emociones por medio de técnicas de procesamiento de imágenes, inteligencia artificial y reconocimiento de expresiones faciales.

Además de ahondar en lo dicho anteriormente, este capítulo aclara algunos conceptos clave aplicados en nuestra investigación: explica cómo se compone y procesa una imagen digital; describe la configuración y funcionamiento de las redes neuronales y más concretamente las redes neuronales convolucionales, que son especialmente aptas para la clasificación de imágenes.

3.1. Redes neuronales

Una red neuronal es un conjunto interconectado de elementos de procesamiento cuya funcionalidad se asemeja a la de una neurona [76]. Estos elementos de procesamiento se conocen comúnmente con el nombre de perceptrones.

La Figura 3.1 muestra su estructura básica: se multiplican entradas numéricas (x_1, x_2 y x_3 en este caso) por sendas cantidades o pesos y se suman los resultados para obtener a y después aplican una función de activación $\sigma(a)$ a esta suma, lo que produce el resultado final y.

$$\begin{aligned} a &= \mathbf{xw} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \\ y &= \sigma(a) \end{aligned} \tag{3.1}$$

Para cualquier vector de entrada $\mathbf{x}$ de longitud n, este proceso también puede expresarse por medio de las siguientes ecuaciones.

$$\begin{aligned} a &= \sum_{i=1}^{n} w_i x_i \\ y &= \sigma(a) \end{aligned} \tag{3.2}$$

Existen muchos tipos distintos de funciones de activación, pero para conseguir que la red neuronal clasifique información no lineal, dicha función debe ser también no lineal [77]. En la siguiente sección, hablaremos brevemente sobre algunas funciones de activación comunes y sus aplicaciones.

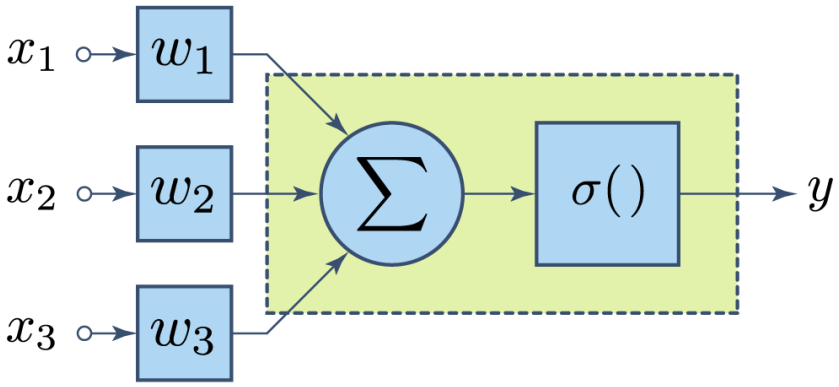


Figura 3.1: Esquema de un perceptrón o neurona

Llamamos capa a un arreglo de varios perceptrones o neuronas. La Figura 3.2 presenta una estructura de este tipo. En el caso mostrado, cada elemento del vector $\mathbf{x}$ se alimenta a cada neurona después de haber sido multiplicado por su peso correspondiente w_{ij} donde i representa el número de entrada y j la neurona a la que está conectada. Esto permite a la capa modelar las interacciones

entre las distintas entradas y mapear dichas entradas a un mayor número de salidas. De este modo, entre mayor sea el número de neuronas en la capa, más compleja será la información que la capa pueda modelar. La Ecuación 3.3 indica cómo obtener el vector $\mathbf{a}$ de entradas para las funciones de activación, previa multiplicación del vector $\mathbf{x}$ por la matriz de pesos W. Cada elemento de $\mathbf{a}$ se alimenta a una función de activación para obtener los elementos del vector de salida $\mathbf{y}$.

$$\mathbf{a} = \mathbf{x}W = \begin{bmatrix} x_1 & x_2 & \dots & x_m \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & & \vdots \\ w_{m1} & w_{m2} & \dots & w_{mn} \end{bmatrix} \tag{3.3}$$

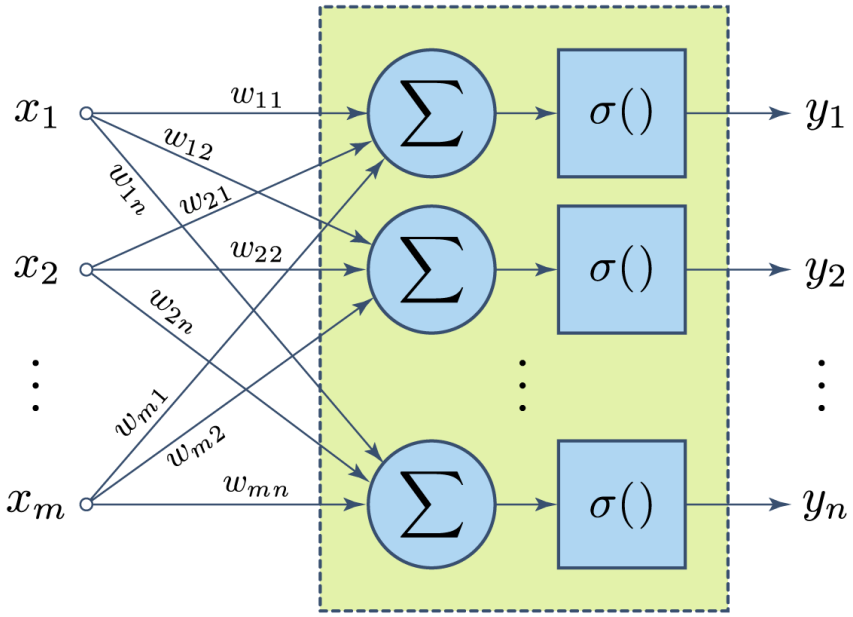


Figura 3.2: Representación de una capa de neuronas

Pueden colocarse varias capas una detrás de otra para construir una red más compleja, capaz de realizar tareas de clasificación más difíciles. La Figura 3.3 representa una red de dos capas (para facilitar su lectura, no se muestran los pesos individuales), $\mathbf{x}$ está compuesto por n_0 entradas, el vector $\mathbf{h}^{[1]}$ contiene todas las n_1 salidas correspondientes a la primera capa oculta y $\mathbf{h}^{[2]}$ las de la segunda, que aquí reescribimos como $\mathbf{y}$, los n_2 elementos de la salida. Por tanto, el proceso completo puede representarse como

$$\begin{aligned} \mathbf{a}^{[1]} &= \mathbf{x}W^{[1]} \\ \mathbf{h}^{[1]} &= \sigma(\mathbf{a}^{[1]}) \\ \mathbf{a}^{[2]} &= \mathbf{x}W^{[2]} \\ \mathbf{h}^{[2]} &= \sigma(\mathbf{a}^{[2]}) \\ \mathbf{y} &= \mathbf{h}^{[2]} \end{aligned} \tag{3.4}$$

Una red neuronal como esta, puede clasificar un conjunto de n_0 variables de entrada en una de n_2 categorías. Por ejemplo, podría utilizarse como un sistema de recomendación en un servicio de música vía *streaming*. Se toman ciertos datos del usuario en forma de entradas numéricas (edad, localización, etc.) y encontramos el género musical que con mayor probabilidad resulte ser de su agrado con el objeto de recomendarle canciones pertenecientes a dicho género.

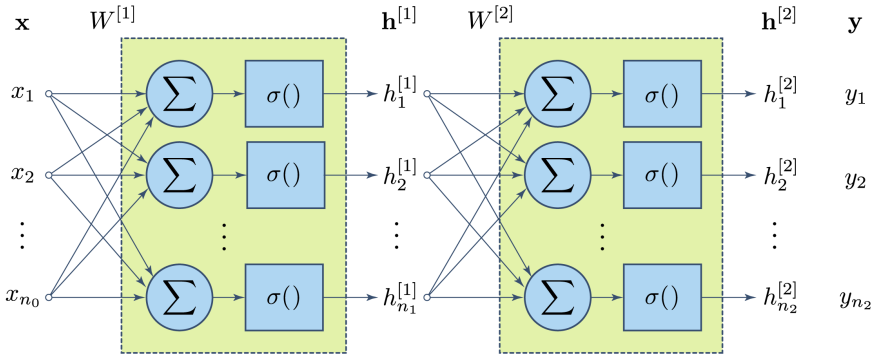


Figura 3.3: Red de dos capas

La exactitud de esta clasificación está en función de los pesos. Idealmente, una correcta arquitectura de red, configurada con el conjunto adecuado de pesos conseguirá buenas clasificaciones en la mayoría de los casos, una vez que la red haya sido entrenada. Cuando hablamos de entrenamiento de una red neuronal, nos referimos al proceso de calibración de los pesos con el cual se obtiene la salida deseada en función de la entrada correspondiente, que se repite con un número suficiente de pares entrada-salida llamado conjunto de entrenamiento. Al término de este proceso, la red debería ser capaz de clasificar correctamente entradas que no están incluidas en el conjunto de entrenamiento.

3.1.1. Función de activación

Una neurona artificial puede actuar como un interruptor o compuerta, dejando pasar información hacia otras neuronas o capas si es relevante para la tarea de

clasificación, o bloqueándola en caso contrario. Este comportamiento es denominado activación neuronal: si la suma de las entradas ponderadas sobrepasa cierto umbral, se dice que la neurona está activada. La función de activación determina si la neurona se activará o no de acuerdo a ciertas entradas y normaliza la salida para mantenerla en un rango determinado, habitualmente entre 0 y 1. A diferencia de otros parámetros, la función de activación suele ser la misma para gran parte de las neuronas de una red, cuando no de todas ellas [78].

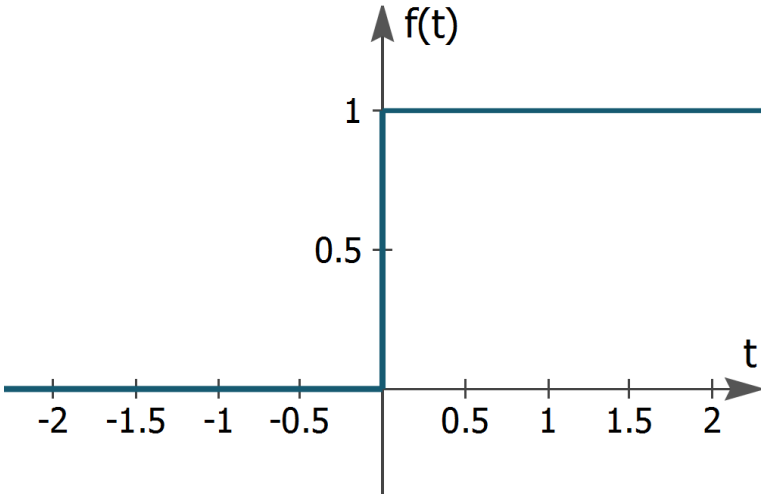


Figura 3.4: Función escalón

Dado que solamente puede tomar dos valores, una función escalón binaria —como la que se muestra en la Figura 3.4– no resulta útil para la mayoría de las tareas de clasificación, mientras que una función de activación lineal hace que el comportamiento de cualquier número de capas sea idéntico al de una sola, reduciendo la capacidad clasificativa de la red e impidiendo que pueda ser entrenada a través de retropropagación, un algoritmo de entrenamiento frecuentemente utilizado.

Las redes neuronales modernas se sirven de funciones de activación no lineales como la sigmoide y la ReLU, entre otras. La sigmoide (Figura 3.5) proporciona un gradiente suave (lo que evita cambios bruscos en los valores de salida) y es derivable, una condición importante para la retropropagación.

$$sigmoide(a) = \frac{1}{1 + e^{-a}} \tag{3.5}$$

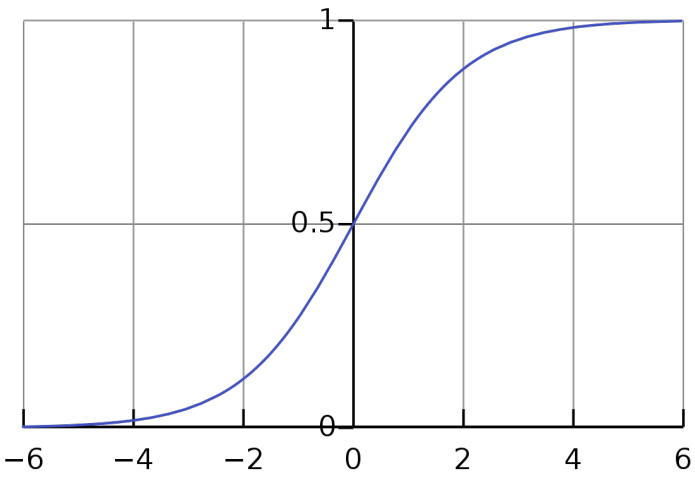


Figura 3.5: Función sigmoide

Sin embargo, para redes con alto número de neuronas, esta función resulta muy demandante desde el punto de vista computacional, lo cual puede resolverse utilizando en su lugar unidades lineales rectificadas (ReLU por sus siglas en inglés). Una ReLU también posibilita el uso de retropropagación, pero su derivada es mucho más fácil de calcular.

$$R(a) = max(0, a) \tag{3.6}$$

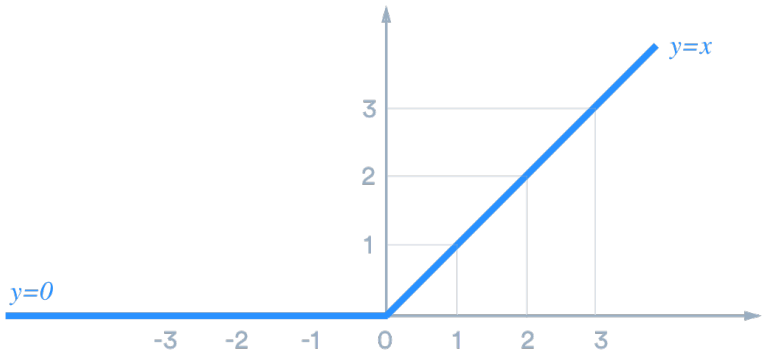


Figura 3.6: Función ReLU

Para clasificar una entrada dentro de una etiqueta de varias disponibles, frecuentemente se emplea la función exponencial normalizada o *softmax*.

$$softmax(a) = \frac{e^{a_i}}{\sum_{j=1}^{K} e^{a_j}} \tag{3.7}$$

3.1.2. Función de pérdida

Una vez que se ha conseguido implementar la red neuronal, es necesario determinar los pesos que minimicen su error de clasificación haciéndola pasar por un proceso de entrenamiento, mismo que requiere una función de pérdida para cuantificar qué tan cerca está la red de su estado de aprendizaje óptimo.
Para cualquier entrada x_i, la red producirá un resultado de clasificación $\hat{y}_i$, mismo que se busca hacer lo más cercano posible a la etiqueta de clasificación real y_i, previamente conocida. Dicho de otro modo, tratamos de minimizar el error de clasificación $y_i - \hat{y}_i$ para todas las entradas. Este error conjunto se cuantifica por medio de una función conocida como función de pérdida. El error cuadrático medio (ECM), el error absoluto medio (EAM) y la entropía cruzada (EC) son funciones de pérdida de uso común.

$$ECM = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.8}$$

$$EAM = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.9}$$

$$EC = \frac{1}{n}\sum_{i=1}^{n} y_i \cdot log(\hat{y}_i) \tag{3.10}$$

Estas funciones permiten conocer qué tanto mejora la capacidad clasificativa de una red durante el proceso de entrenamiento y con qué exactitud podrá clasificar información no incluida en el conjunto de entrenamiento.

3.1.3. Descenso de gradiente

Es un algoritmo iterativo de optimización para encontrar los valores mínimos de una función. Es útil para encontrar el mínimo error posible en la función de pérdida, y por ello, la configuración de la red que ofrece los mejores resultados. Comienza calculando la pendiente (o gradiente) de la función para un punto arbitrario de su superficie y después mueve este punto en la dirección que presente el descenso más pronunciado. Después de repetir este procedimiento varias veces, el punto de prueba finalmente alcanzará el punto más bajo de la curva, es decir, el mínimo de la función (Figura 3.7) El proceso de entrenamiento de una red comienza por seleccionar un valor arbitrario para cada uno de los pesos, alimentar la red con algunas de las entradas disponibles y evaluar la función de pérdida correspondiente a las salidas esperadas y las obtenidas por la red.

Después es necesario encontrar el gradiente de la función de pérdida, esto es, todas las derivadas parciales de la función con respecto a cada uno de los pesos, de manera que podamos tener una idea clara de cómo hay que modificar los pesos en la siguiente iteración para encontrar un menor valor de la función de pérdida.

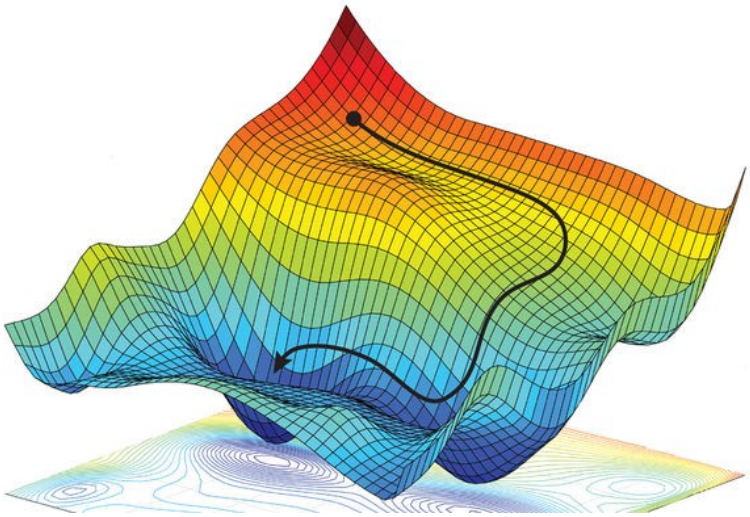


Figura 3.7: Descenso de Gradiente

3.1.4. Retropropagación

Es un método numérico que busca el mínimo de una función de error en el espacio de los pesos, utilizando el descenso de gradiente [8]. Es el método más popular de entrenamiento para redes neuronales, a causa de su simplicidad y su aptitud para entrenar redes neuronales de conectividad arbitraria [79]. La retropropagación es también un caso especial de diferenciación automática, un modo de calcular derivadas de funciones en un punto determinado, aplicando la regla de la cadena a los valores numéricos correspondientes, lo cual hace a este método mucho más sencillo que la diferenciación simbólica y más preciso que la diferenciación numérica, dos de las alternativas más comunes.
Volviendo a la red de la Figura 3.3, entrenarla mediante retropropagación precisa el uso de la derivada $\frac{dL}{dW_2}$ para optimizar los valores de W_2, de modo que podamos aplicar la regla de la cadena y obtener:

$$\frac{dL}{dW_2} = \frac{dL}{d\mathbf{h}^{[2]}} \frac{d\mathbf{h}^{[2]}}{d\mathbf{a}^{[2]}} \frac{d\mathbf{a}^{[2]}}{dW_2} \tag{3.11}$$

Lo cual nos permite calcular el nuevo conjunto de pesos W_2^* para cada iteración

$$W_2^* = W_2 - \alpha \frac{dL}{dW_2} \tag{3.12}$$

Donde α se conoce como tasa de aprendizaje y representa qué tan grandes deben ser los pasos hacia el mínimo de la función. Valores bajos de α harán que el proceso de aprendizaje sea más lento, pero pasos más grandes pueden pasar por alto el mínimo, impidiendo que el algoritmo converja hacia una solución, de modo que la elección de un valor apropiado resulta importante.

Todas las ecuaciones involucradas en el conjunto (3.4) deben ser evaluadas numéricamente primero. A esto se le llama paso hacia adelante (*forward pass*) o propagación hacia adelante. Una vez que se cuenta con dichos valores numéricos, es posible calcular las derivadas (Ecuación 3.11) desde la última hacia la primera, de aquí el nombre retropropagación.

3.2. Imágenes digitales

Para que una computadora sea capaz de guardar y procesar imágenes, éstas deben codificarse en términos numéricos. Una manera de conseguir esto es dividiendo la imagen en pequeños cuadrados que contengan un solo color y acomodarlos en una rejilla rectangular, de modo similar a un mosaico o una pintura puntillista. Estos cuadrados reciben el nombre de pixeles (apócope del inglés *picture element*), y el color que contienen puede expresarse por medio de uno o tres números, para imágenes en escala de grises o a color, respectivamente. A causa del modo en que se retienen dentro de la computadora, dichos números suelen ser enteros dentro del rango entre 0 y 255. Para un pixel en escala de grises, el 0 normalmente representa un color negro, mientras que el 255 representa el blanco. Los números intermedios corresponden a distintas tonalidades de gris, ordenadas por luminosidad. Comúnmente, un pixel a color tiene tres componentes: rojo, verde y azul, cuyos valores también se encuentran entre 0 y 255. Esto se debe a que la mayor parte de los colores que el ojo humano es capaz de percibir pueden obtenerse combinando estos tres colores básicos. Expresado en términos más formales, una imagen digital es una imagen que ha sido discretizada tanto en términos de coordenadas espaciales como en términos de luminosidad o brillo [80]. Se representa por medio de un arreglo bidimensional de enteros como se muestra a en la Ecuación 3.13 y en la Figura 3.8, o por una serie de arreglos, uno por cada componente de color, de manera similar a como se representa en la Figura 3.9.

$$f(x,y) = \begin{bmatrix} f(1,1) & f(1,2) & \dots & f(1,N) \\ f(2,1) & f(2,2) & \dots & f(2,N) \\ \vdots & \vdots & & \vdots \\ f(M,1) & f(M,2) & \dots & f(M,N) \end{bmatrix} \tag{3.13}$$

3.2.1. Ajuste de contraste en imágenes

De acuerdo con González [81], el histograma de una imagen digital con niveles de gris en el rango $[0, L-1]$ es una función discreta $h(r_k) = n_k$, donde r_k es el k-ésimo nivel de gris y n_k es el número de pixeles en la imagen con nivel de gris r_k. Una práctica común consiste en normalizar un histograma, dividiendo cada uno de sus valores entre el número total de pixeles en la imagen.

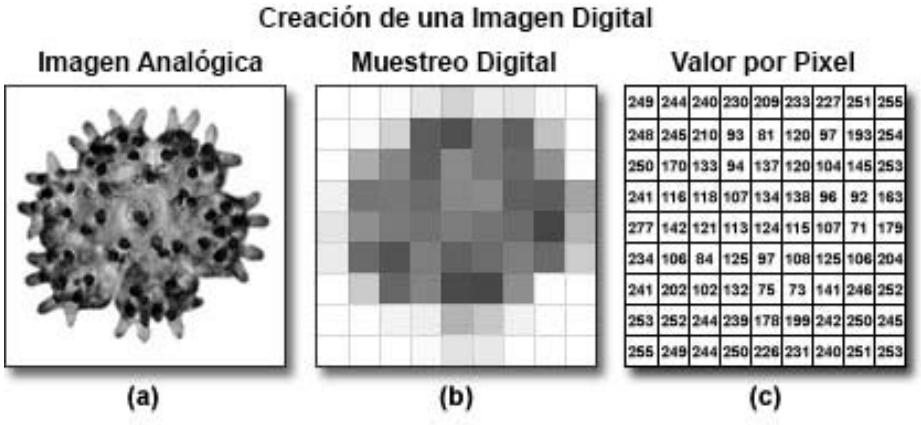


Figura 3.8: Composición de una imagen digital

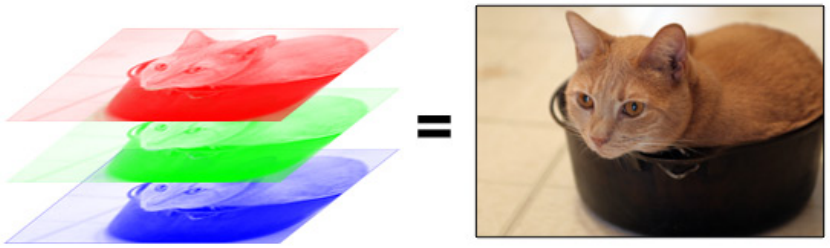


Figura 3.9: Composición de una imagen digital a color

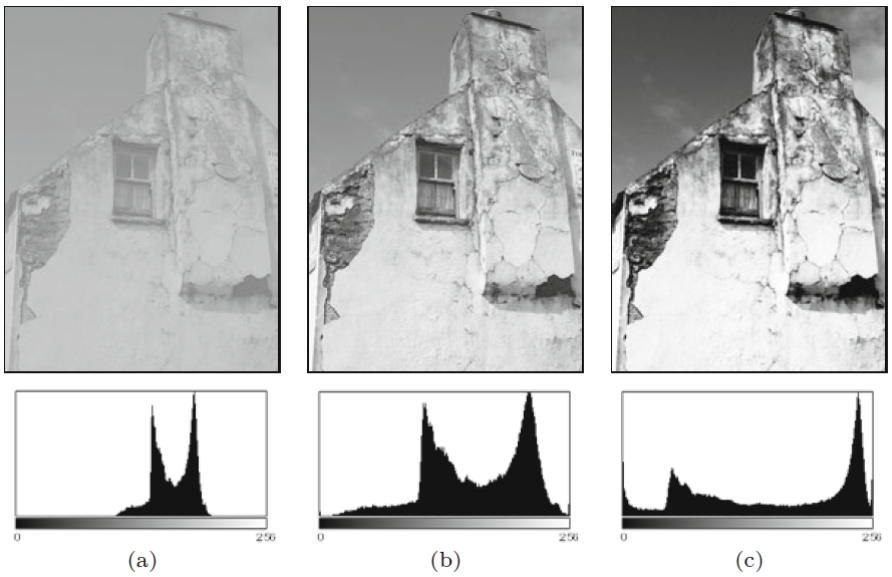


Figura 3.10: Ejemplos de histogramas correspondientes a modificaciones de una misma imagen.

Este método normalmente incrementa el contraste global de la imagen, especialmente cuando la información útil que contiene está representada por valores de contraste muy cercanos entre sí. Por medio de este ajuste, las intensidades se distribuyen de mejor manera a lo largo del histograma y se consigue aumentar el contraste entre áreas de bajo contraste local.

Por otro lado, la ecualización adaptativa de histogramas (EAH) [82] consiste en mapear cada pixel de la imagen a un nivel de intensidad que sea proporcional al rango que ocupa entre los pixeles que lo rodean. Aunque mejora el contraste, tiene el inconveniente de amplificar también el ruido de la imagen. Para corregir esto, suele utilizarse lo que se conoce como limitación de contraste. Se puede definir el mejoramiento del contraste como la pendiente de la función que relaciona la intensidad de entrada con la intensidad de salida. Supongamos que ambas intensidades son iguales, en ese caso, una pendiente unitaria significa que no hay ninguna modificación, mientras que pendientes mayores implican mayor diferencia de contrastes. Así, la limitación en el cambio de contraste restringe la pendiente en la función de mapeo.

3.3. Redes neuronales convolucionales

Se utilizan habitualmente para detectar características clave en imágenes. Otros algoritmos requieren que un ser humano decida previamente qué características

deberían detectar, mientras que una red neuronal convolucional (RNC) es capaz de aprender por sí misma a detectar las características clave de cada imagen durante el proceso de entrenamiento.

Una práctica común consiste en aplicar varias operaciones de convolución con núcleos distintos (también llamados filtros) para transformar una entrada de n canales a una salida de m canales, donde m es el número de núcleos.

Por convención, la salida de una capa convolucional con n filtros se conoce como mapa de características ($w^{(t)} \times h^{(t)} \times n$), porque su estructura ya no está vinculada a una imagen específica, sino que representa la superposición de varios detectores de características.

En RNCs suelen utilizarse arreglos bidimensionales con cualquier número de canales (como imágenes en escala de grises o RGB). Por simplicidad, analizaremos la convolución aplicada a un solo canal. Si $X \in \mathbb{R}^{w \times h}$ y $x \in \mathbb{R}^{n \times m}$, la convolución $X * k$ se define como:

$$(X * k)(x, y) = \sum_{i \in [0, n-1], j \in [0, j-1]} k(i, j) X(x + i, y + j) \tag{3.14}$$

Cuando los índices comienzan en cero. En la Figura 3.11 se indica esquemáticamente cómo se realiza la operación de convolución con un núcleo de 3×3 sobre una imagen.

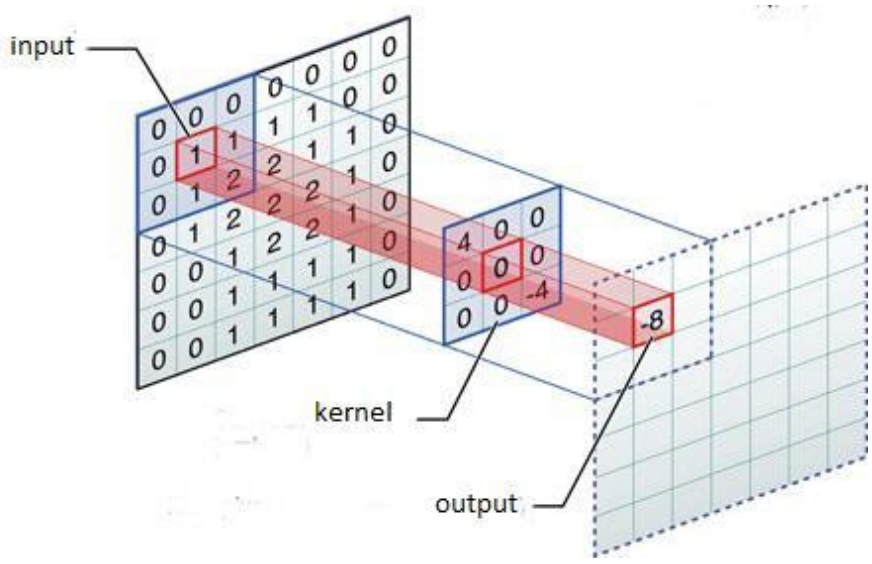


Figura 3.11: Convolución de una imagen con un núcleo de 3×3

El núcleo se desplaza horizontal y verticalmente, obteniendo la sumatoria de todas las multiplicaciones casilla por casilla de los elementos correspondientes. Cada operación arroja como resultado el valor de un solo pixel. Se puede apreciar qué se obtiene al aplicar un núcleo específico a una imagen en la Figura 3.12.

Como puede verse, el resultado de aplicar la convolución es una imagen que enfatiza los bordes de las distintas formas que contiene la imagen de entrada. Los núcleos pueden calibrarse para cumplir ciertos requerimientos, sin embargo, ya no es necesario que una persona lo haga manualmente, la red convolucional delega esta tarea al proceso de entrenamiento, orientado a un objetivo preciso que se expresa por medio de la función de pérdida. La convolución aplicada a una imagen da como resultado otra de menor tamaño, por esto es habitual rellenar los bordes de la primera imagen con ceros, como puede apreciarse en la Figura 3.14. De esta manera, al aplicar la convolución se obtiene una imagen con las mismas dimensiones de la original.

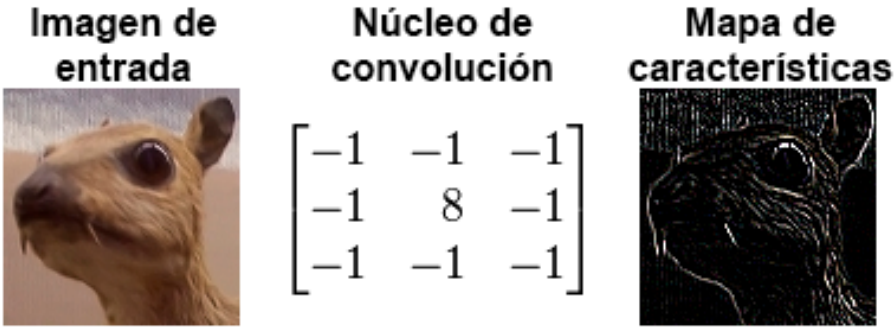


Figura 3.12: Resultado de la convolución con el núcleo mostrado

Si se aplican varias operaciones de convolución (también llamadas filtros) en paralelo, se obtienen superposiciones complejas que pueden simplificar la extracción de las características más relevantes para la clasificación. La principal diferencia entre una capa completamente conectada y una capa convolucional es la habilidad que tiene esta última para trabajar con la geometría de la imagen y determinar todas las particularidades que resultan útiles para distinguir un objeto de otro y descartar aquellas que no lo son [83].

Las capas convolucionales son capaces de extraer características relevantes una vez que las han aprendido a partir de un conjunto de entrenamiento [84]. Las neuronas de una capa convolucional están agrupadas en mapas de características: cada neurona en un mapa de características tiene un campo receptivo que se conecta a las neuronas vecinas de la capa anterior por medio de un conjunto de pesos, comúnmente conocidos como banco de filtros [85].

3.3.1. Capas de agrupación y abandono

Para reducir el tamaño y necesidad de procesamiento de estos mapas de características, además de conseguir invariancia espacial a distorsiones y traslaciones en las imágenes de entrada, se utilizan capas de agrupamiento. [86]. Las capas de agrupamiento dividen una imagen en varias subsecciones y las resumen en un solo valor numérico, habitualmente el promedio o el máximo (Figura 3.13).

Entrenar una red neuronal para ajustarla a un conjunto particular de entrenamiento no garantiza que obtenga buenas predicciones para el conjunto de prueba, incluso si sus predicciones son perfectas durante el entrenamiento. En otras palabras, siempre hay una diferencia entre el desempeño de una red durante su entrenamiento y durante su aplicación a los datos de prueba. Esta diferencia suele ser especialmente significativa cuando la red es compleja y el conjunto de datos es pequeño [87]. Cuando una red arroja buenas clasificaciones para el conjunto de entrenamiento pero no para el de prueba, se dice que sufre de sobreajuste (*overfitting*).
El abandono o *dropout* es una técnica sencilla para reducir el sobreajuste. Consiste en sustituir por ceros algunos de los pesos de la red de manera relativamente aleatoria. Se establece un parámetro π entre 0 y 1 que representa la probabilidad de que un peso en particular sea sustituido por cero durante alguna de las épocas de entrenamiento, de modo que no se actualice durante esa iteración. Esto obliga a que la red memorice ciertas redundancias, mejorando su capacidad de aislar las propiedades esenciales del conjunto de datos. Un valor típico de π es 0.2 [88], pero al igual que otros parámetros, debe ajustarse de acuerdo al conjunto de validación.

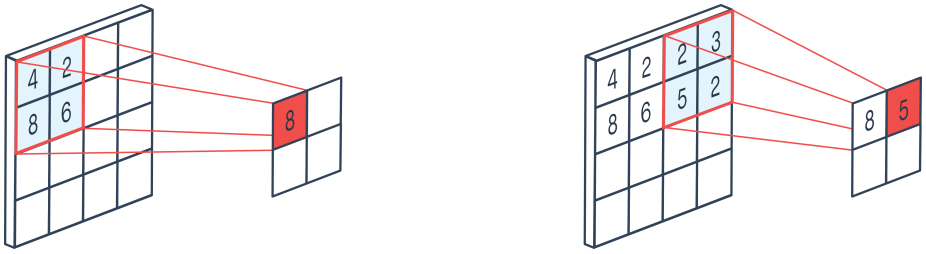


Figura 3.13: Capa de agrupación máxima

3.4. Aportes

Este trabajo describe una propuesta original para predecir la aceptación del consumidor por medio de mediciones de ritmo cardiaco y respuesta galvánica de la piel y reconocimiento de expresiones faciales, es decir, enfocada a registrar directamente las emociones generadas por experiencias sensoriales en un consumidor de manera inadvertida y mínimamente invasiva. A diferencia de estudios análogos que se apoyan en software comercial, nuestro trabajo parte de un siste-

ma con algoritmos propios que permite una configuración flexible y adaptable, basado en una red neuronal convolucional sencilla, de alto desempeño, y con resultados de exactitud cercanos a los del resto de la literatura científica.

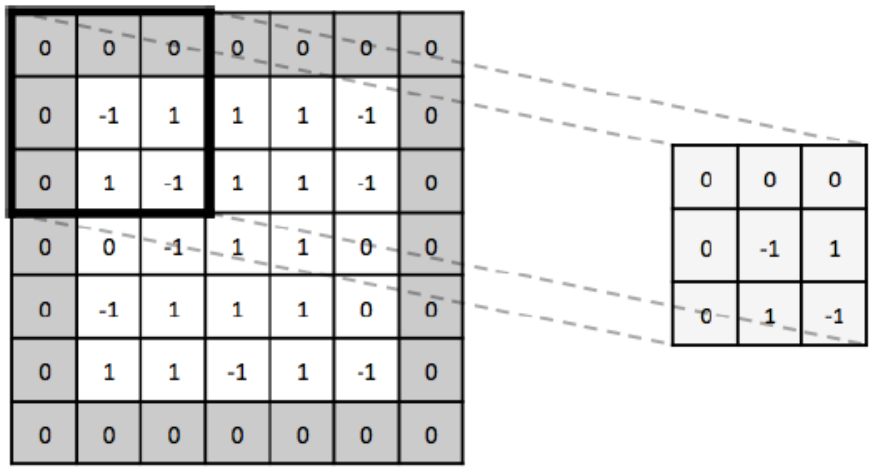


Figura 3.14: Relleno con ceros en los bordes de la imagen

Capítulo 4

Metodología

4.1. Detalles de la implementación

4.1.1. Captura de los datos

Se desarrolló un sencillo programa de captura para gestionar y archivar la información proveniente de los sensores conectados a una PC de trabajo. La interfaz de dicho programa permite registrar el nombre del participante, fecha y hora de cada experimento realizado, así como las lecturas del electroencefalograma, respuesta galvánica y ritmo cardiaco, video del rostro visto de frente y lateralmente: todo en carpetas bien organizadas.

Se utilizaron dos sensores de la compañía NeuLog: el NUL-217 para medir la respuesta galvánica y el NUL-208 para medir el ritmo cardiaco. El dispositivo NUL-208 es capaz de medir la cantidad de latidos del corazón por minuto en un rango de 0 a 240 pulsos por minuto, o utilizar unidades arbitrarias análogas para mostrar funciones de onda con una resolución de 0 a 1023 a una frecuencia máxima de 100 muestras por segundo. El sensor puede utilizarse para monitorear y comparar ritmos cardiacos en distintas condiciones de ejercicio o reposo y puede mostrar cambios de volumen o flujo sanguíneo en un dedo, a lo largo del tiempo. El sensor está basado en el principio de un pletismógrafo: consiste en un transmisor LED infrarrojo y un fotoresistor infrarrojo acoplado, que funciona como receptor [89].

Por otro lado, el NUL-217 mide la conductividad de la piel. Tiene dos unidades de medida: *microsiemens* (μS) o unidades arbitrarias, también utilizadas para mostrar ondas, frecuencias o períodos. Tiene una velocidad máxima de 100 muestreos por segundo y puede registrar información en intervalos de entre 1 segundo y 20 días.

4.1.2. Preprocesamiento de imágenes

OpenCV [90] es una biblioteca de software de código abierto con rutinas útiles para visión por computadora que ofrece interfaces para utilizarse con programas escritos en C++, Python, Java y MATLAB. Dlib es otra biblioteca de código abierto con algoritmos de aprendizaje máquina programados de manera nativa en C++. Se utiliza en robótica, dispositivos empotrados y teléfonos móviles, entre otras aplicaciones [91].

- Se convirtieron las imágenes RGB a escala de grises para reducir a una tercera parte la cantidad de información que tendrá que procesar la red neuronal, sobre todo durante su entrenamiento. Para esto se utilizó la función `cv2.cvtColor()` de OpenCV, que realiza la siguiente operación para cada pixel: $V = 0{,}299 \cdot R + 0{,}587 \cdot G + 0{,}114 \cdot B$, donde V es el valor en escala de grises y R, G y B son los valores asociados a los canales rojo, verde y azul, respectivamente.

- Para determinar si hay algún rostro en la imagen y su posición dentro de la misma, se empleó un algoritmo [1] implementado en Dlib a partir de la técnica llamada histograma de gradientes orientados (HGO). Este algoritmo es útil para detectar diversos tipos de objetos semirrígidos en imágenes, pero esta implementación en particular se entrenó para detectar rostros humanos, principalmente en posición frontal.

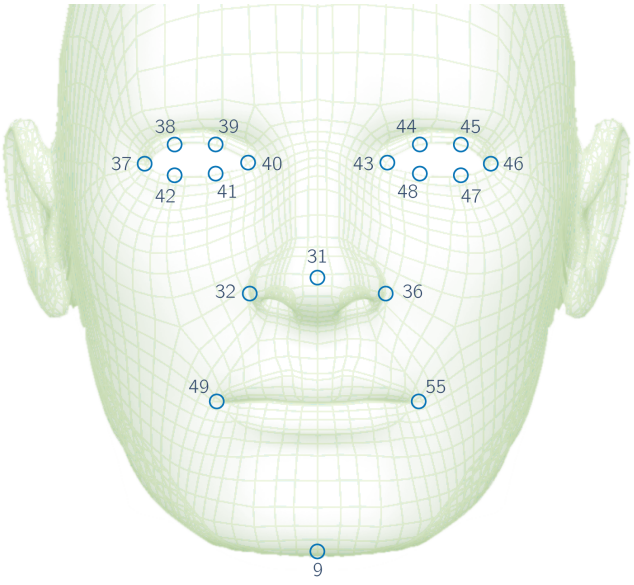
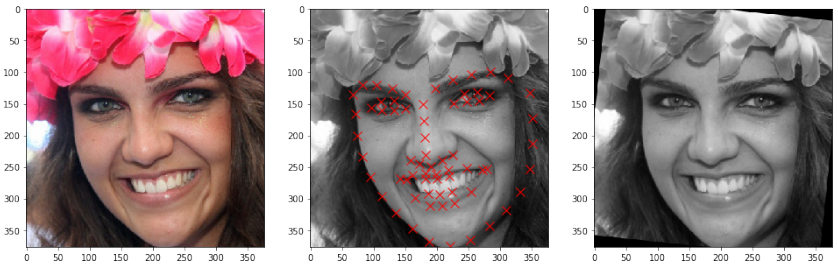


Figura 4.1: Principales puntos clave.

[1] *face detection* `http://dlib.net/face_detection_ex.cpp.html`

- Posteriormente, la parte de la imagen original que contiene el rostro se somete a otro clasificador [2] similar al anterior, también construido sobre HGO, combinado con un clasificador lineal, ventana deslizante y pirámide de imagen. Este algoritmo determina la posición de 68 puntos clave en el rostro analizado, que corresponden a características útiles para la determinación de emociones: cejas, ojos, boca, etc. La Figura 4.1 muestra los puntos utilizados en este proceso. Para ver la localización de todos los puntos disponibles, referirse a la Figura 5.5.



(a) (b) (c)

Figura 4.2: (a) Imagen original a color (b) Convertida a escala de grises y mostrando sobreimpuestos los puntos clave (c) Imagen rotada

- Para evitar que la red neuronal pierda efectividad tratando de clasificar rostros con distintos grados de rotación, se alinearon todos utilizando el método `cv2.warpAffine()` de OpenCV. Para ello se tomó en cuenta la inclinación de la línea que va del punto 40 al 43 (de un lagrimal a otro) con respecto a la horizontal (Figura 4.3).

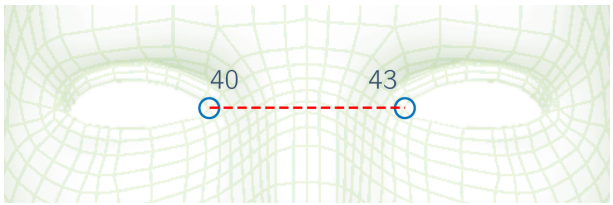
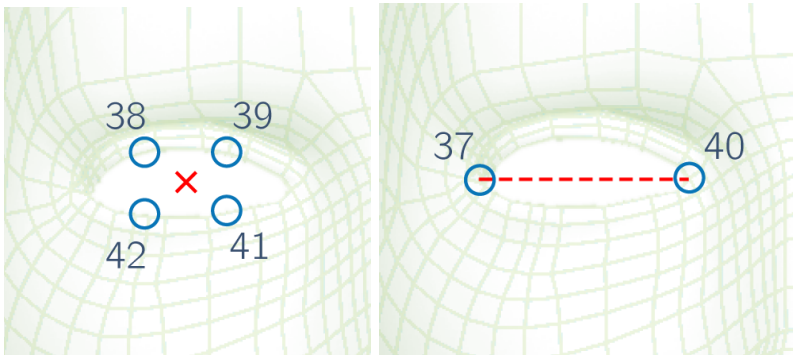


Figura 4.3: Referencia horizontal a partir de la línea entre los lagrimales: puntos 40 y 43.

- Con la intención de aprovechar la simetría del rostro para reducir la complejidad de la red neuronal y aprovechar mejor el material de entrenamiento, cada imagen registrada de un rostro se dividió en cuatro cuadrantes. Si asumimos que el lado izquierdo puede ser procesado de manera similar al derecho, podemos reflejar uno de ellos y así utilizar ambos para alimentar

[2] *face landmark detection* `http://dlib.net/face_landmark_detection_ex.cpp.html`

una misma red neuronal, en lugar de construir y entrenar dos redes distintas que cumplirían la misma función. La determinación de los cuadrantes se realizó de acuerdo al siguiente procedimiento:



(a) (b)

Figura 4.4: (a) Centro del rectángulo superior (b) Distancia *dp*

Cuadrantes superiores

- Se localiza el centro aproximado del ojo izquierdo promediando las coordenadas de los puntos 38, 39, 42 y 41. (Figura 4.4a)
- Se define *dp* como 0.4 veces la distancia entre los puntos 37 y 40 (Fig. 4.4b)
- Se construye un rectángulo cuyos lados se encuentran a una distancia $3dp$ del centro, excepto por el inferior, que se encuentra a $2dp$. (Fig. 4.5)

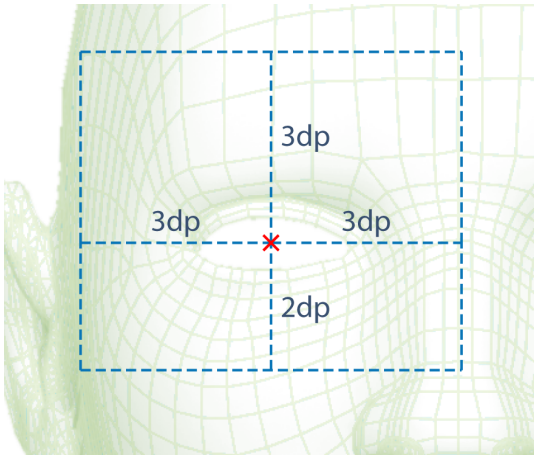
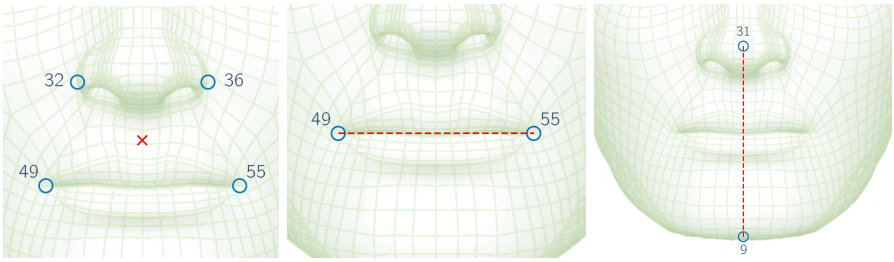


Figura 4.5: Rectángulo para ojos y cejas

- Este rectángulo constituye el área del cuadrante superior izquierdo. El mismo método se aplica a los puntos correspondientes del ojo derecho para determinar el otro cuadrante superior.

Cuadrantes inferiores

- Se determina un punto central promediando las coordenadas de los puntos 32, 36, 49 y 55 (Fig. 4.6a).
- Se define *dpx* como la distancia entre los puntos 49 y 55, que representan las comisuras de la boca (Fig. 4.6b), y *dpy* como la mitad de la distancia entre el punto 31 en la nariz, hasta el punto 9 en el extremo inferior de la barbilla (Fig. 4.6c).
- A partir del punto central, se colocan los lados de un rectángulo una distancia de *dpy* hacia arriba del centro, 2*dpy* hacia abajo y *dpx* hacia la izquierda (o en dirección contraria para el cuadrante derecho) como se muestra en la Figura 4.7.



(a) (b) (c)

Figura 4.6: (a) Centro del rectángulo inferior (b) Distancia *dpx* (b) Distancia *dpy*

- Después, se cambiaron las proporciones de cada cuadrante para que las dimensiones de cada imagen sean de 64 x 64 pixeles, ya que este es el tamaño de entrada que aceptará la red neuronal.
- Se utilizó el método `cv2.createCLAHE()` para optimizar el contraste de los cuatro cuadrantes por separado y así propiciar que la red neuronal encuentre las características relevantes de cada imagen con mayor facilidad. El resultado del preprocesamiento puede apreciarse en la Figura 4.8.
- La red neuronal es capaz de procesar conjuntos de valores entre 0 y 1, pero los valores de escalas de gris en las imágenes se encuentran en el rango [0,255], por lo cual se normalizan dividiéndolos entre su máximo valor posible, que es de 255.

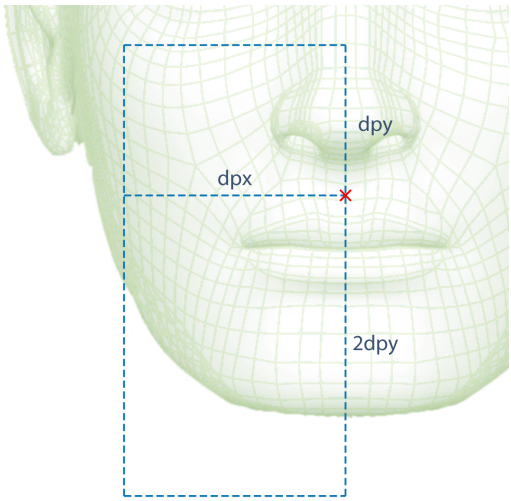


Figura 4.7: Rectángulo de boca y nariz

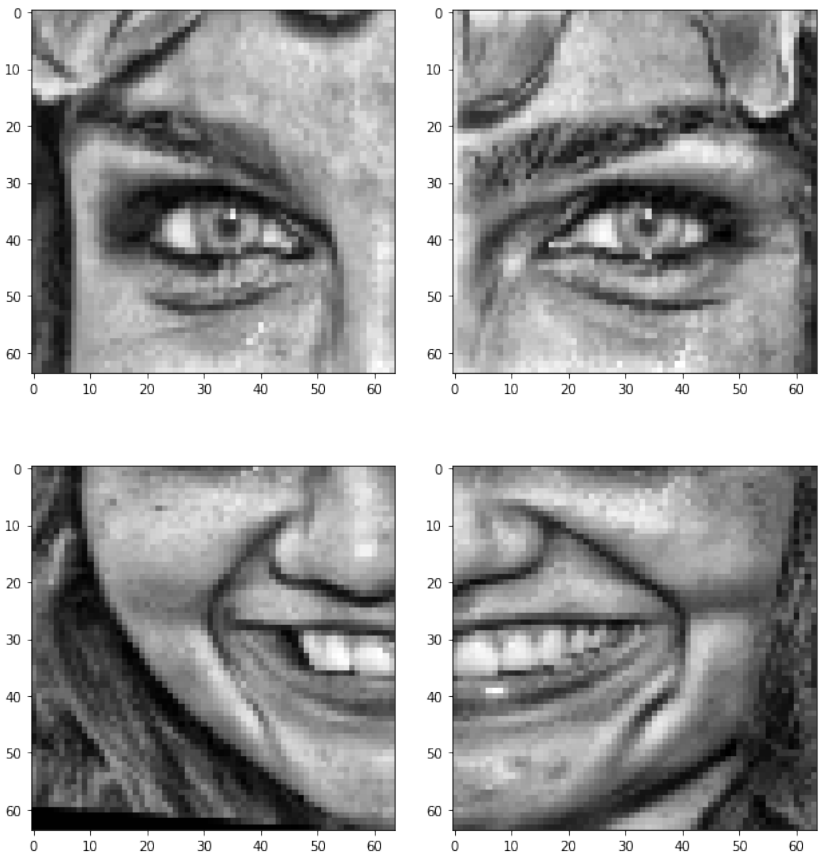


Figura 4.8: Resultado de dividir la imagen original en cuadrantes, escalarlos a un tamaño de 64 x 64 pixeles y ajustar el contraste.

4.1.3. Construcción de la red neuronal

Se realizaron varios experimentos con distintas arquitecturas de red neuronal, pero solamente describiremos la que consiguió mejores resultados de clasificación.

Dado que cabe esperar mejores clasificaciones de varias redes procesando la misma información en paralelo que de una sola, se dividió el problema de clasificar las expresiones faciales en dos partes: una red para clasificar los cuadrantes superiores (que llamaremos red A) y otra para los inferiores (red B). Como ya se mencionó, las imágenes del lado derecho se invierten para ser clasificadas por las mismas redes que procesarán el lado izquierdo, como se muestra en la Figura 4.9.

Las redes A y B comparten la misma arquitectura que se puede apreciar en la Figura 4.10: reciben una matriz de 64×64 valores correspondientes a la imagen de uno de los cuadrantes y clasificarán la entrada dentro de una de diez posibles categorías, es decir, su salida será un vector de 10 números reales en el rango [0,1] que representan la probabilidad de que la imagen de entrada se relacione con una de las siguientes diez etiquetas: neutral, alegría, tristeza, sorpresa, miedo, disgusto, enojo, desdén, ninguna e indeterminada. Las capas que la componen están listadas en la Tabla 4.1, cuyas columnas contienen el tipo de cada capa, de acuerdo a lo que ofrece la librería Keras; sus dimensiones y el número de parámetros o pesos susceptibles de entrenamiento que contiene. La red tiene 4,217,658 parámetros en total. Todas las funciones de activación son ReLU, excepto para la última capa, donde se utiliza *softmax*.

Tabla 4.1: Capas que conforman las redes A y B.

Tipo de capa	**dimensiones**	**parámetros**
InputLayer	(64, 64, 1)	0
ZeroPadding2D	(66, 66, 1)	0
Conv2D	(64, 64, 32)	320
MaxPooling2D	(32, 32, 32)	0
Dropout	(32, 32, 32)	0
ZeroPadding2D	(34, 34, 32)	0
Conv2D	(32, 32, 64)	18,496
MaxPooling2D	(16, 16, 64)	0
Dropout	(16, 16, 64)	0
Flatten	(16,384)	0
Dense	(256)	4,194,560
Dropout	(256)	0
Dense	(16)	4,112
Dense	(10)	170

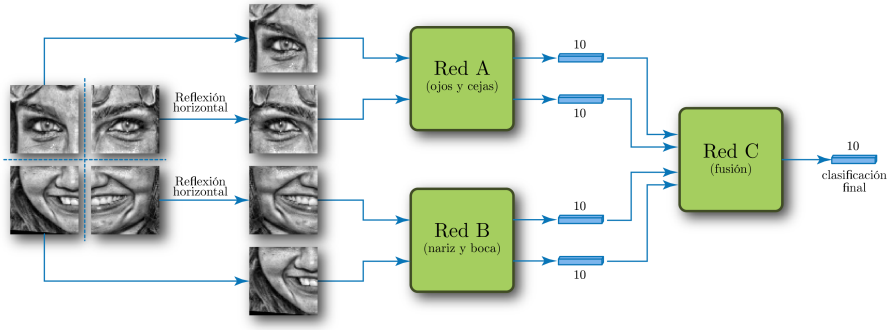


Figura 4.9: Distribución de la información a través de las distintas redes.

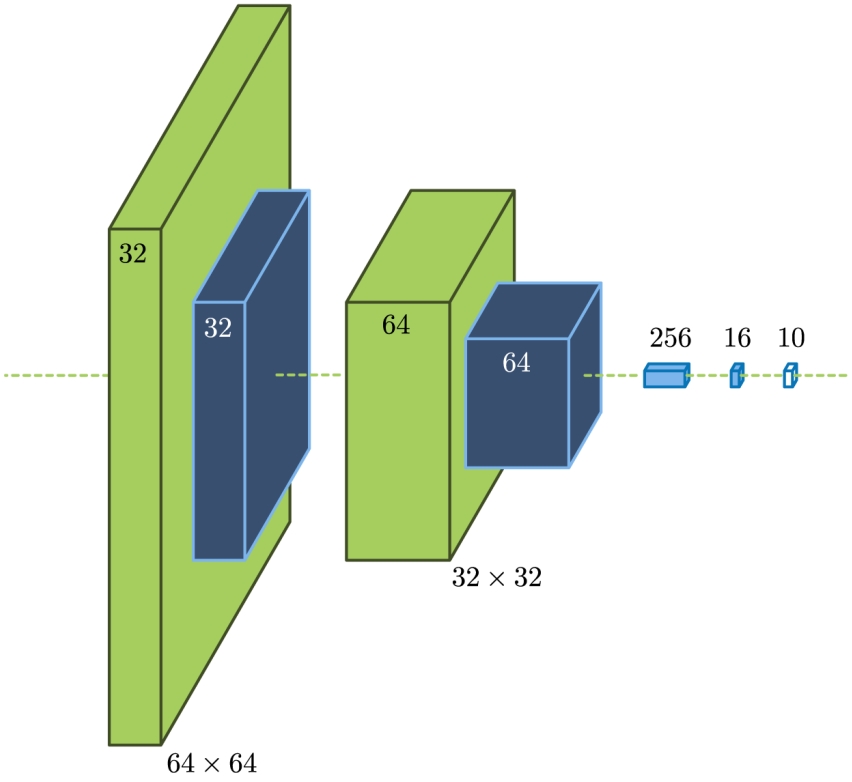


Figura 4.10: Arquitectura de las redes A y B. Las capas se describen a detalle en la Tabla 4.1.

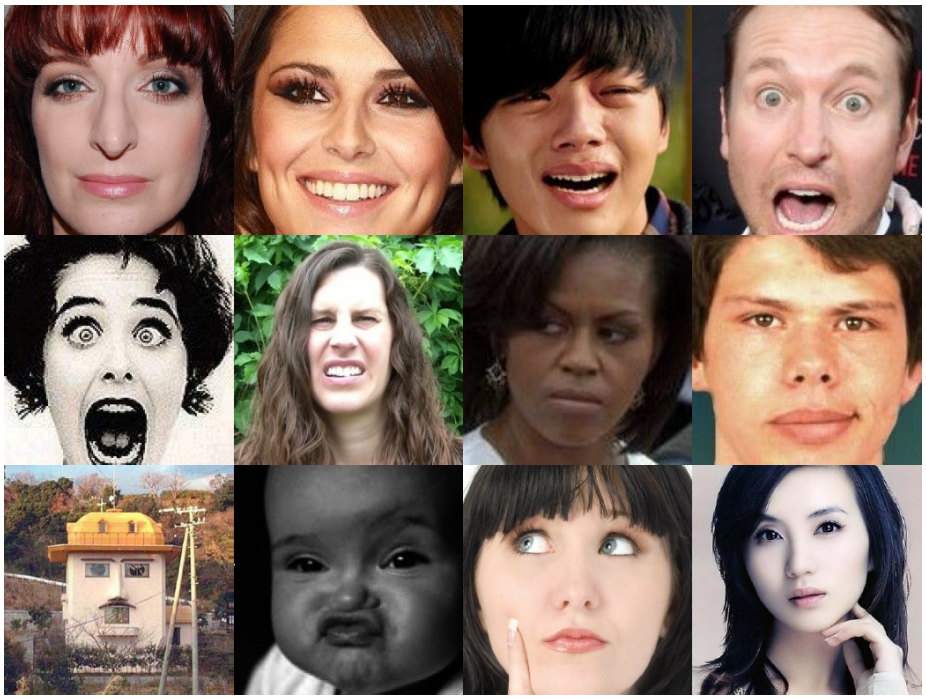


Figura 4.11: Ejemplos de imágenes de *AffectNet*, obtenidas de [92]. Las etiquetas asociadas por renglón son: 1) neutral, felicidad, tristeza, sorpresa; 2) miedo, disgusto, enojo, desdén; 3) no es un rostro, indeterminada, ninguna, ninguna

Una vez que han procesado los cuadrantes, las redes A y B devuelven cuatro vectores de diez elementos, mismos que se agrupan para obtener un solo vector de 40 elementos. Este vector se alimenta a la red C, que realiza una fusión de los resultados anteriores para dar un veredicto final en un solo vector de 10 elementos. Las capas de las que consta se muestran en la Tabla 4.2 y entre todas suman un total de 3,700 parámetros entrenables. También aquí las funciones de activación son ReLU en general y *softmax* para la capa de salida.

Tabla 4.2: Capas que conforman la red C.

Tipo de capa	**dimensiones**	**parámetros**
Dense	(40)	1,640
Dense	(30)	1,230
Dense	(20)	620
Dense	(10)	210

Figura 4.12: Ejemplos de imágenes de CK+, obtenidas de [93]. Las etiquetas asociadas por renglón son: 1) disgusto, felicidad, sorpresa, miedo; 2) enojo, desdén, tristeza, neutral

4.2. Entrenamiento de la red

Las imágenes que utilizamos para entrenar las redes pertenecen a la base de datos *AffectNet* (Figura 4.11), compuesta por imágenes extraídas de internet y clasificadas a mano de acuerdo a dos modelos distintos de emoción: uno discreto (10 etiquetas) y otro continuo. Las imágenes que incluye varían ampliamente en sus características: hay imágenes en color y en escala de grises que muestran personas de todo el mundo en posturas y con expresiones muy diferentes, y aunque el tamaño promedio es de 425 x 425 pixeles, la variación estándar es de 349 x 349 [92]. Su variabilidad es una de las características por las cuales seleccionamos este conjunto para el entrenamiento, pues esto amplía la capacidad de reconocimiento de la red. También nos servimos de la base de datos extendida de Cohn-Kanade [93] o CK+ (Figura 4.12) para hacer pruebas posteriores, ya que el número de imágenes que contiene es más reducido y éstas están clasificadas en solo 7 etiquetas. Además, los rostros que muestra se fotografiaron en condiciones más controladas y por ello más restrictivas en términos de aprendizaje potencial de la red neuronal.

Seleccionamos 40,336 imágenes de la base de datos de *AffectNet* y 327 de CK+. Después, un programa recibió la dirección en disco duro de cada una de estas imágenes y las sometió a todo el procesamiento previo: detección de rostros, determinación de puntos clave, recorte de cada imagen en cuatro cuadrantes, ajuste de contraste, etc. (el proceso completo se detalla en la sección 4.1.2 de este capítulo) y el resultado se guardó en un archivo, junto con la etiqueta emocional asignada a cada una de las imágenes.

Más tarde, se extrajeron de este archivo los cuadrantes y las etiquetas necesarias para entrenar las redes A y B por separado.

Después se utilizaron estas redes para procesar la misma información empleada

para entrenarlas y los resultados se guardaron en otro archivo, que posteriormente sirvió como entrenamiento para la red C.

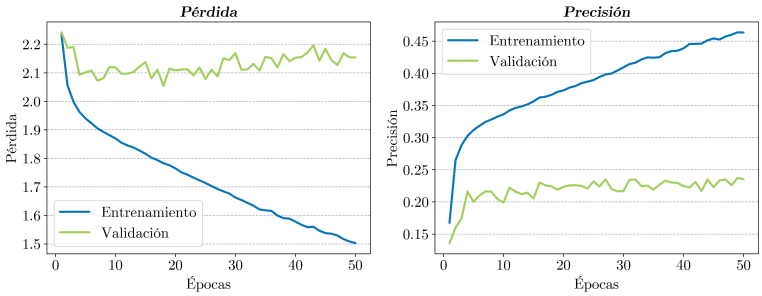


Figura 4.13: Entrenamiento de la red A.

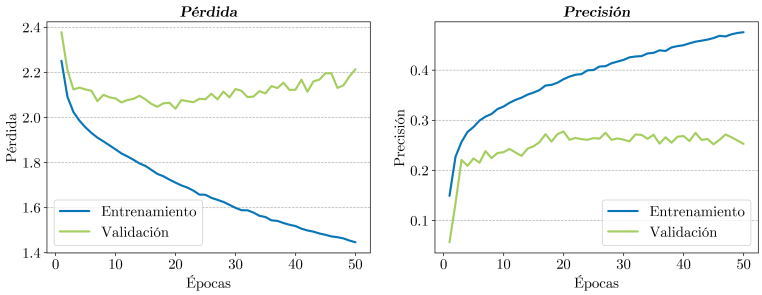


Figura 4.14: Entrenamiento de la red B.

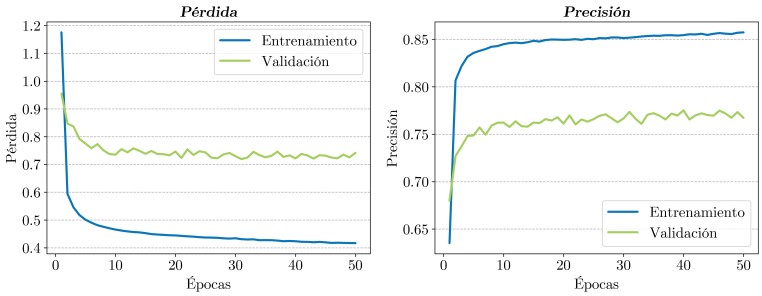


Figura 4.15: Entrenamiento de la red C.

El proceso de entrenamiento de las redes se llevó a cabo en una PC con procesador Intel(R) Core(TM) i5-3340 CPU a 3.10GHz, 8 GB en RAM y una tarjeta gráfica NVIDIA GeForce GTX 1050. Dicho proceso abarcó 50 ciclos con un tamaño de lote de 128. El porcentaje de datos reservados para validación fue de

20 % para las redes A y B, pero 15 % para la red C. Se empleó un 0.4 como índice de abandono. Las Figuras 4.13, 4.14 y 4.15 muestran la evolución de las mediciones de pérdida y precisión para cada una de las redes a lo largo de su proceso de entrenamiento. A lo largo del tiempo se realizaron varios entrenamientos, la Tabla 4.3 contiene los resultados del más reciente. Para una estimación más certera de los resultados que se pueden obtener en general, puede referirse a las matrices de confusión que se incluyen en la sección 4.3.

Cada una de las redes se probó con los datos de CK+ para evaluar su comportamiento en un conjunto distinto de datos, con objeto de detectar un posible sobreajuste.

Tabla 4.3: Resultados y tiempos de entrenamiento para las tres redes.

	red A	**red B**	**red C**
Pérdida en entrenamiento	1.313	1.391	0.469
Precisión en entrenamiento	0.587	0.555	0.842
Pérdida en pruebas	1.174	1.253	0.897
Precisión en pruebas	0.657	0.583	0.752
Tiempo de entrenamiento	0.422 hrs.	0.438 hrs.	0.0257 hrs.

4.3. Resultados en reconocimiento de emociones

Como puede apreciarse, la estructura de la red mostrada en la Figura 4.10 difiere de la mostrada en el Capítulo 5, ya que corresponde al resultado de una optimización posterior.

Al entrenar y probar la red con elementos extraídos únicamente de *AffectNet* se obtenían muy buenos resultados, pero al aplicarla a imágenes de otros conjuntos (CK+), se obtenían clasificaciones de menor precisión, lo cual indica que la red estaba sobreajustada.

Para corregir este problema, se hicieron numerosas pruebas con distintos metaparámetros y arquitecturas de red, además de añadir capas de abandono. Así se consiguió mejorar la precisión de los resultados de la red para CK+, aunque ésta se haya visto reducida en el caso de *AffectNet*. También es notable el hecho de que la precisión en las redes A y B es relativamente baja, puesto que cada una de ellas recibe únicamente una parte de cada imagen facial y además de esto, se permitió que disminuyera su efectividad parcial a fin de reducir el sobreajuste en el sistema completo.

A continuación mostramos las matrices de confusión correspondientes a las tres partes de la red. Estas matrices se obtuvieron después de un proceso de validación cruzada de 10 iteraciones, es decir: cada una de las partes de la red se entrenó 10 veces con ordenamientos distintos del mismo conjunto de entrenamiento. Esto, aunado al hecho de que el entrenamiento es un proceso heurístico,

hace que cada entrenamiento obtenga una red ligeramente distinta. Finalmente, se obtiene una matriz de confusión para cada iteración y al final se promedian los 10 resultados.

En la Figura 4.16 se muestra la matriz de confusión para la red A, alimentada con el conjunto de datos extraído de *AffectNet*. Las etiquetas del eje vertical indican el resultado obtenido por la red, mientras que las del eje horizontal presentan las etiquetas asociadas a cada una de las imágenes de la base de datos, o dicho de otro modo, la clasificación que la red debería haber reportado en un caso ideal de entrenamiento perfecto. En las casillas negras está escrito el número de imágenes totales asociadas a cada etiqueta, junto con el porcentaje de las mismas que fue clasificado correctamente en verde; o incorrectamente, en rojo. Las casillas de la diagonal nos dicen cuántas imágenes fueron correctamente clasificadas, así como el porcentaje que representan del total de 80,625 cuadrantes alimentados a la red. En el resto de las casillas se reportan las clasificaciones erróneas.

Así, en las Figuras 4.16, 4.18 y 4.20, podemos ver que la precisión de la red A fue de 55.87 %, 58.73 % para la red B y 84.45 % para la red C, cuando se aplicaron al conjunto de datos *AffectNet*. Como referencia, las Figuras 4.17, 4.19 y 4.21 muestran los resultados correspondientes para el conjunto de datos CK+: 66.08 % en la red A, 62.92 % en la B y 78.46 % en la C, que es la que reporta el resultado final de todo el sistema. En el caso de CK+, las matrices de confusión no incorporan datos para todas las etiquetas, ya que este conjunto de datos no las incluye algunas de las que sí se encuentran en *AffectNet*.

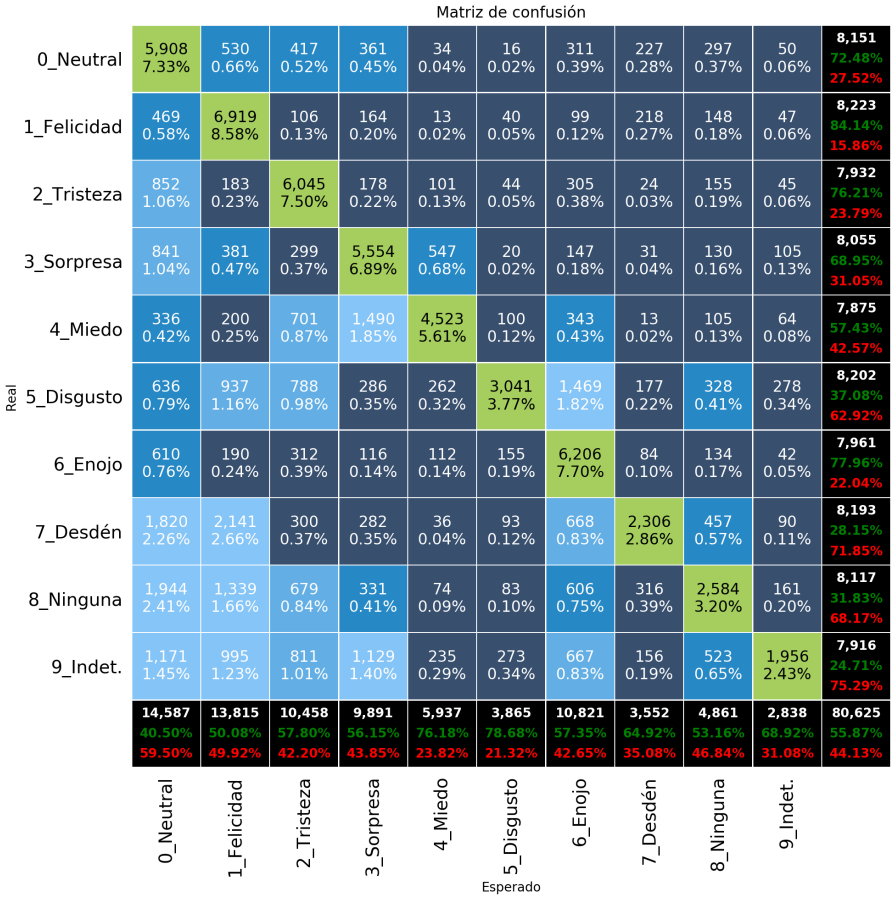


Figura 4.16: Matriz de confusión de la red A aplicada a los datos de *AffectNet*

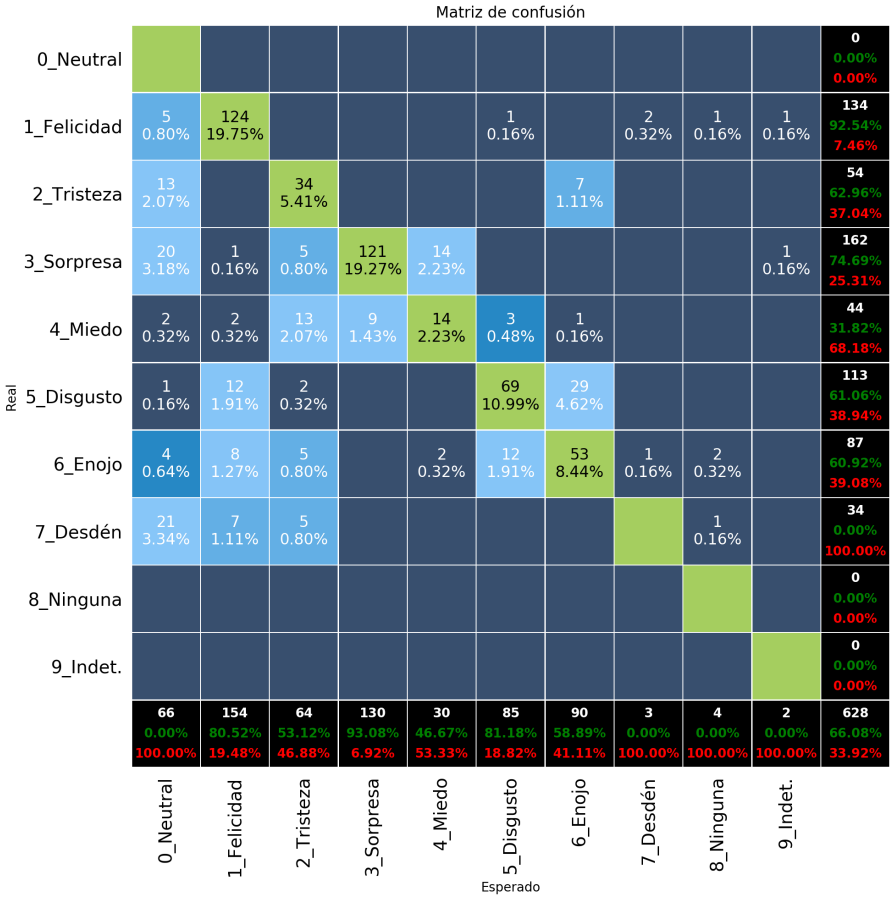


Figura 4.17: Matriz de confusión de la red A aplicada a los datos de CK+

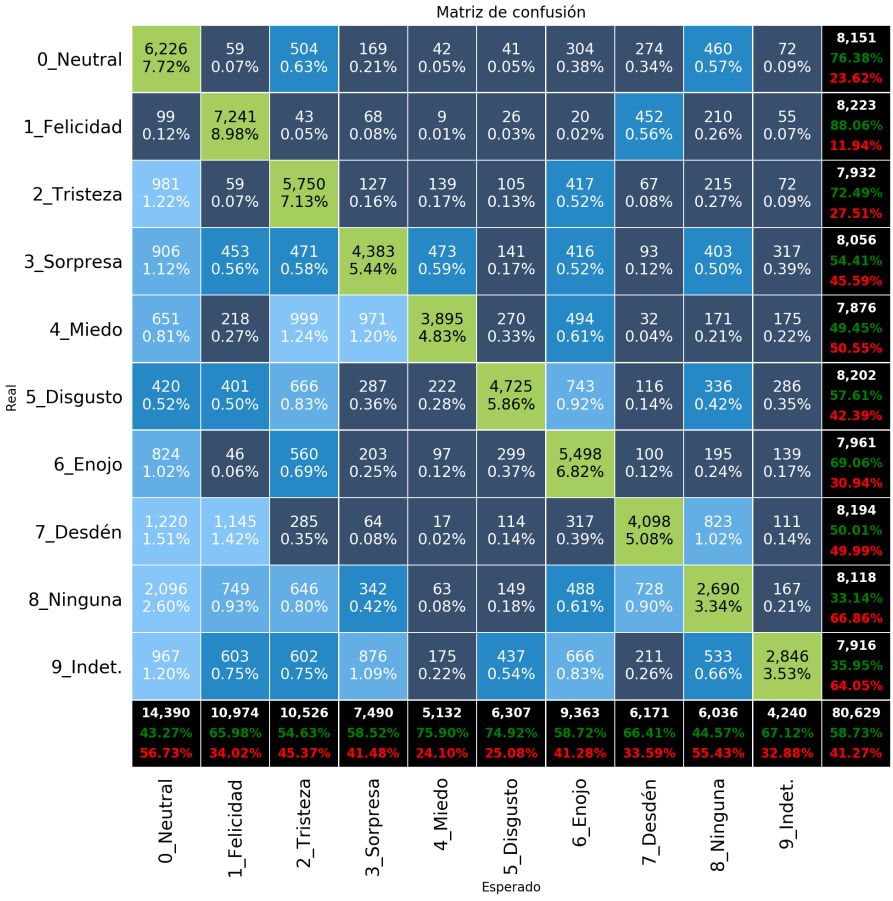


Figura 4.18: Matriz de confusión de la red B aplicada a los datos de *AffectNet*

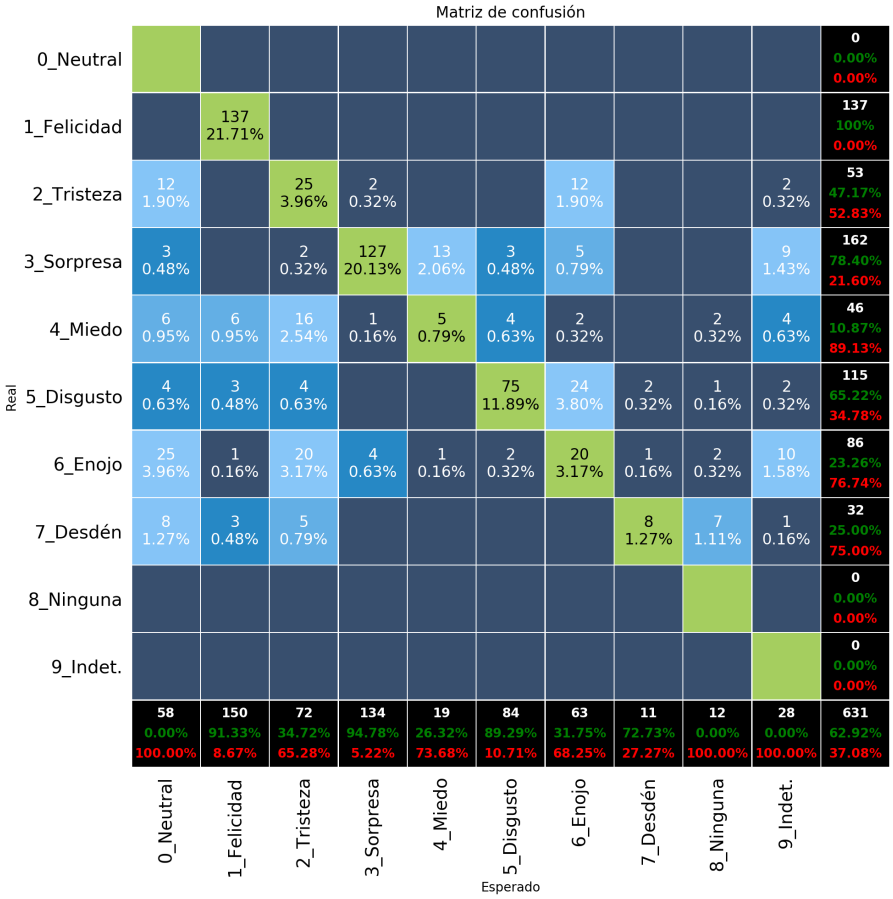


Figura 4.19: Matriz de confusión de la red B aplicada a los datos de CK+

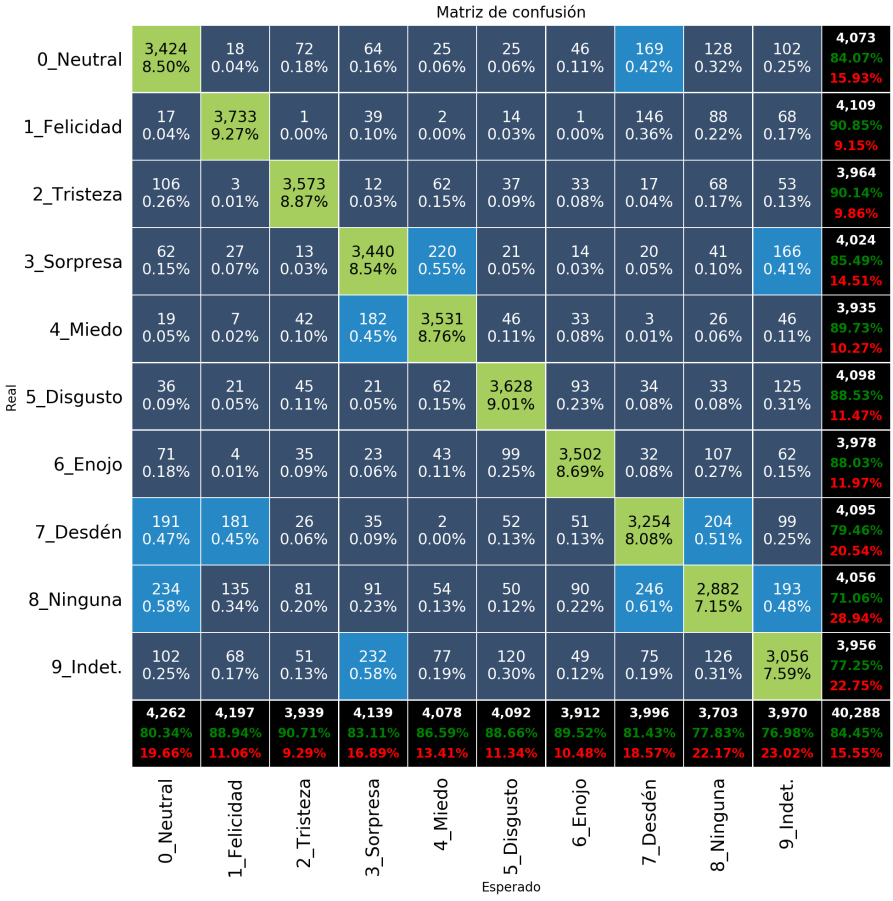


Figura 4.20: Matriz de confusión de la red C aplicada a los datos de *AffectNet*

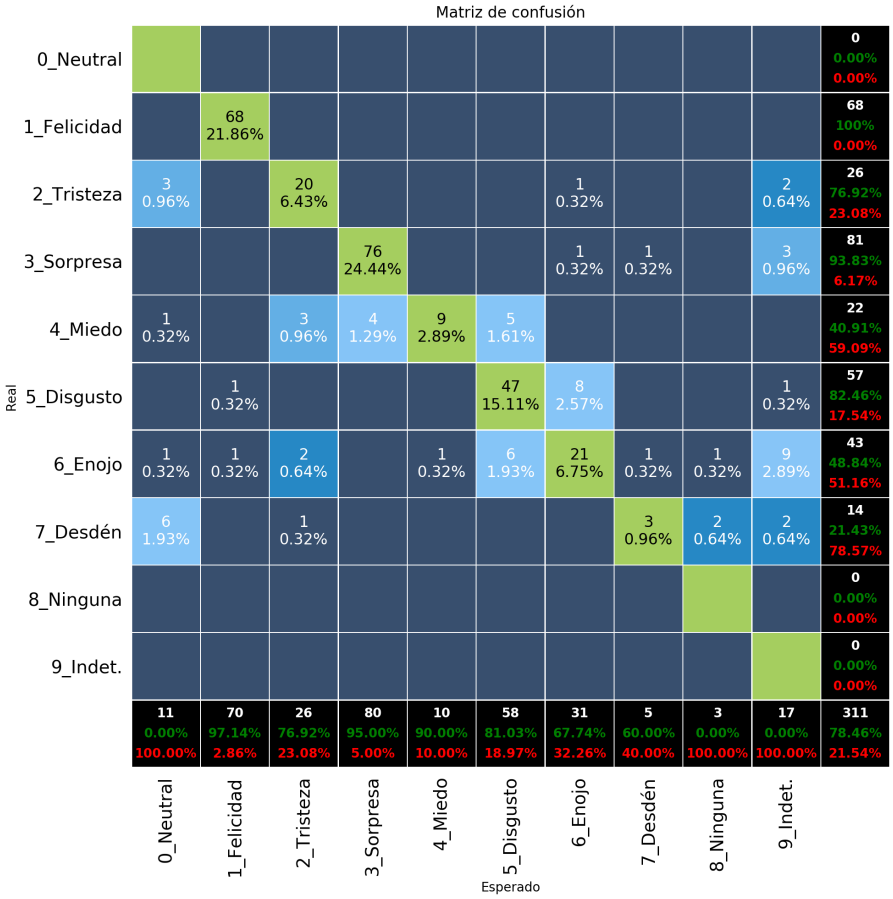


Figura 4.21: Matriz de confusión de la red C aplicada a los datos de CK+

Capítulo 5

Aplicaciones en Aceptación del Consumidor

5.1. Introducción

La respuesta del consumidor y sus decisiones de compra tienen una fuerte componente emocional. Un producto puede causar en quien lo compra distintos sentimientos: exclusividad, orgullo de pertenecer a un grupo social, confianza, etc.

Las reacciones emocionales de un consumidor hacia un producto tienen una mayor influencia en sus elecciones que los atributos sensoriales [94, 95]. Se han realizado muchos intentos de medir dichas reacciones para predecir el desempeño de un producto en el mercado: ritmo cardiaco, temperatura corporal, respuesta galvánica de la piel (RGP), electroencefalografía (EEG), expresiones faciales y otros indicios potenciales para determinar las emociones de un consumidor. Aún así, el reconocimiento de emociones causadas por un producto alimenticio es una disciplina nueva y los algoritmos requeridos todavía no han sido desarrollados [13]. En este contexto, Viejo *et al.* evaluaron EEG, ritmo cardiaco, temperatura y expresiones faciales en consumidores de cerveza. En [14], He *et al.* registraron expresiones faciales de participantes expuestos a olores de naranja y pescado. Leitch *et al.* midieron la respuesta a endulzantes de té a través de una escala hedónica, un cuestionario con términos emocionales y expresiones faciales [73].

Danner *et al.* reportaron la medición de cambios en el nivel de conductancia y temperatura de la piel, ritmo cardiaco, pulso y expresiones faciales de voluntarios mientras probaban diferentes tipos de jugo de naranja [12]. De manera similar, otros autores han realizado estudios con jamón ahumado [13] y sabores amargos [74]. Una característica común en muchos de estos proyectos es que utilizan *FaceReader* [57], un software comercial y de propósito general para reconocimiento de expresiones faciales.

El reconocimiento de expresiones faciales (REF) ha cobrado gran interés dentro del área de análisis sensorial, que comprende dos enfoques principales [96]: el modelo continuo y el modelo categórico. El primero propone un amplio espectro continuo de emociones distintas, mientras que el segundo se limita a un conjunto discreto de emociones básicas.

Concretamente, el modelo categórico propuesto por Paul Ekman en [43] sigue siendo el más generalizado. A través de un entrenamiento especializado, el método de Ekman permite identificar expresiones faciales por medio del análisis de ciertas activaciones de músculos faciales. Sin embargo, se requiere cerca de una hora de análisis por cada minuto de video [97] y el entrenamiento necesario también requiere muchas horas. Es por esto que se ha dedicado una gran cantidad de trabajos de investigación a encontrar algoritmos computacionales que sean capaces de superar el nivel actual de evaluación en expresiones faciales.

Las redes neuronales convolucionales (RNC) han obtenido buenos resultados en aplicaciones prácticas [98] además de ser robustas y reducir el impacto de variaciones entre rostros [96]. Entre los distintos trabajos basados en RNCs para REF se encuentra el de Cai *et al.* [98] que propone una nueva función de pérdida para maximizar las diferencias entre clases en la clasificación de emociones realizada por RNCs. Zhao *et al.* mostraron en [99] una RNC con arquitectura tridimensional para determinar y aprender características relevantes en imágenes faciales y secuencias de flujo óptico. Li *et al.* [100] utilizaron un mecanismo de atención en RNCs para clasificar expresiones faciales en rostros parcialmente cubiertos enfocándose en diferentes regiones de una imagen facial y ponderándolas de acuerdo al nivel de oclusión que presentan y su nivel de importancia en la clasificación. Wang *et al.* [101] buscaron mejorar la precisión de reconocimiento combinando múltiples regiones ponderadas de imágenes faciales. Liong *et al.* diseñaron una red neuronal tridimensional de tres flujos y sólo dos capas que es capaz de extraer características de alto nivel así como microexpresiones a través de la determinación de propiedades de flujo óptico.[102]

Adicionalmente, otros tipos de análisis —entre los que se cuentan el registro de emociones a través de cambios fisiológicos producidos en consumidores al probar nuevos productos alimenticios— han permitido entender con mayor profundidad las respuestas del consumidor [15].

En general, existen dos tipos de evaluaciones sensoriales: explícitas e implícitas. El análisis explícito se sirve cuestionarios que utilizan términos descriptivos de tipo verbal o no verbal [103, 104, 105]. Este tipo de análisis resulta fácil de entender para los consumidores y recopilar la información es relativamente sencillo. Sin embargo, los resultados pueden ser afectados por sesgos cognitivos [105] y no permiten medir la experiencia del consumidor en el momento exacto en que está probando el producto.

Por otro lado, los métodos implícitos se enfocan a las expresiones faciales y otros cambios fisiológicos. Una revisión exhaustiva de cómo estos últimos se relacionan con ciertas emociones puede encontrarse en [106]. Otros métodos implícitos miden el ritmo cardiaco, conductancia y temperatura de la piel así como la dilatación de la pupila, entre otros cambios fisiológicos y respuestas del sistema nervioso autónomo [103, 104, 106].

Aunque algunos estudios proponen que la percepción de sabores básicos está vinculada a expresiones faciales específicas (por ejemplo, el sabor ácido produce contracción de los labios) [107], hay muchas variables distintas que pueden afectar tanto al REF como a los cambios fisiológicos: qué tanta hambre tiene un consumidor, el tipo de comida que está probando, el tiempo transcurrido desde el comienzo de la prueba, etc. Incluso en un intervalo de 10 s, un consumidor puede mostrar varias expresiones faciales distintas [108]. Más aún, los cambios en las expresiones faciales son más difíciles de determinar en pruebas de productos alimenticios que aquellas en las que se aspira un perfume o se mira un video [103], dado que el movimiento de la mandíbula al masticar y las ocasionales oclusiones del rostro (cuando la mano lleva el alimento a la boca) causan con frecuencia ruido en los algoritmos de REF. Éstas pueden ser algunas de las razones por las que estudios similares no parecen obtener conclusiones más sólidas [7, 109, 110].

El presente trabajo busca avanzar un paso hacia una predicción más confiable de la aceptación del consumidor por medio de RNCs y otros algoritmos de aprendizaje automático capaces de interpretar expresiones faciales y encontrar posibles correlaciones entre medidas obtenidas con sensores biométricos, análisis facial y opiniones descritas explícitamente por el consumidor.

Presentamos un sistema de REF desarrollado por nosotros mismos, dado que no todos los estudios basados en soluciones de tipo comercial han tenido éxito. El programar nuestra propia aplicación nos permite explorar distintos métodos y modelos emocionales con mayor flexibilidad, siendo que las emociones se expresan de modo multimodal [111], y que la fusión de varios canales de información permite mejorar las predicciones [112], decidimos incluir también sensores biométricos en el análisis. La RNC mostrada en este trabajo comprende cuatro canales, uno para cada cuadrante de las imágenes faciales, pues estudios anteriores sugieren que varias redes tienen un mejor desempeño que una sola [96].

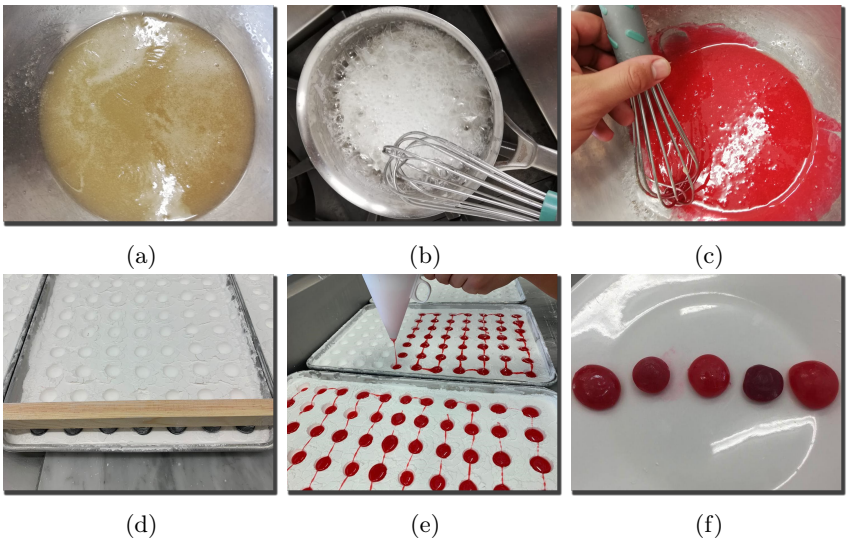
El resto de este documento está organizado de la siguiente manera: La sección 2 describe los materiales y métodos utilizados para la implementación del sistema de evaluación sensorial propuesto. La sección 3 presenta los resultados obtenidos, que se discuten en la sección 4. Finalmente, la sección 5 cierra el documento resumiendo las principales contribuciones del trabajo y planteando perspectivas para futuros trabajos.

5.2. Materiales y métodos

5.2.1. Análisis sensorial

Descripción de las muestras de sabor y olor

Para los experimentos descritos, se utilizaron los siguientes ingredientes y porcentajes (w/w) para preparar dulces blandos de cinco sabores distintos: glucosa (36.5 %, Deiman, USA), azúcar (33.21 %, Gelita, México), agua (23.38 %), gelatina sin sabor (5.3 %, Gelita, México), ácido cítrico (1.28 %, ENSIGN, China), saborizante (0.3 %) y colorante rojo (0.03 %, Deiman, USA). En dos de los cinco sabores (almeja y queso) se utilizó maltodextrina en lugar de azúcar.



(a) (b) (c)

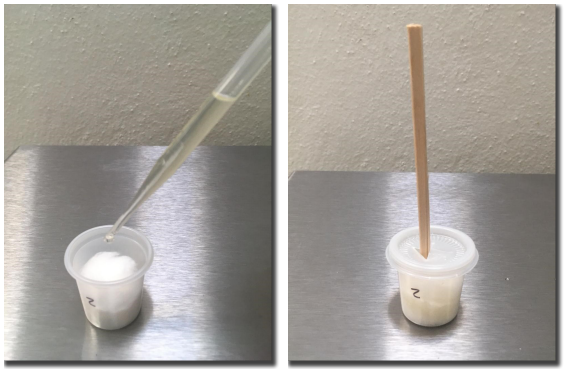
(d) (e) (f)

Figura 5.1: Proceso de elaboración de los dulces: (a) gelatina en agua, (b) mezcla de agua, glucosa y azúcar, (c) solución de gelatina, (d) cama de almidón, (e) mezcla vertida en la cama de almidón, (f) dulces terminados.

El proceso de elaboración para los dulces puede verse en la Figura 5.1 y se describe a continuación:

1. Disolver la gelatina sin sabor en agua (10.6 g/L) durante 30 minutos.
2. Mezclar agua y azúcar (11.5 g/L) y calentar a 70°C, añadir glucosa y luego aumentar la temperatura hasta 108°C.
3. A 100 °C, añadir a la mezcla la solución de gelatina sin sabor, el colorante, el saborizante y el ácido cítrico diluido (1.28 g/L).
4. Finalmente, moldear la mezcla vertiéndola en una cama de almidón y dejarla reposar durante 18 horas.

Se buscó que los distintos dulces tuvieran una apariencia muy similar (Figura 5.1f) para evitar que los voluntarios pudieran predecir su sabor. Estos dulces permiten que se libere el sabor en el momento preciso en que el voluntario prueba el producto, para que la expresión facial pueda ser registrada al mismo tiempo en que el voluntario percibe el estímulo. Seleccionamos los sabores de los dulces de manera que contáramos con 5 estímulos sensoriales distintos entre sí; tres considerados como agradables: menta (Deiman, USA), piña (Deiman, USA) y fresa (Deiman, USA) y dos que podríamos calificar como desagradables: almeja (Bell, USA) y queso Gouda (Bell, USA).



(a) (b)

Figura 5.2: Muestra de olor: (a) Solución vertida en algodón. (b) Contenedor sellado y palillo de madera.

También preparamos un conjunto de muestras de olor empapando algodón en distintas sustancias y guardándolo en un contenedor sellado. Cada participante podía utilizar un palillo de madera para acercar la sustancia a su nariz (Figura 5.2). Los aromas utilizados para el experimento fueron: piña (Ungerre, USA), menta (Deiman, USA), vinagre (Ungerer, USA) queso Gouda (Bell, USA) y humo (Castells, USA)

Participantes y puesta a punto

Un grupo de 120 estudiantes, profesores y administrativos de la Universidad Panamericana se ofrecieron a realizar la prueba. El experimento tuvo lugar en la cabina de un laboratorio sensorial con iluminación controlada. La cabina cuenta con un dispositivo *Kinect*, que integra varios sensores distintos: cámara a color, cámara infrarroja y sensor de profundidad, que capturan la vista frontal y la geometría del rostro de cada participante utilizando un sólo dispositivo, lo cual elimina la necesidad de preparar y sincronizar muchas fuentes de información distintas.

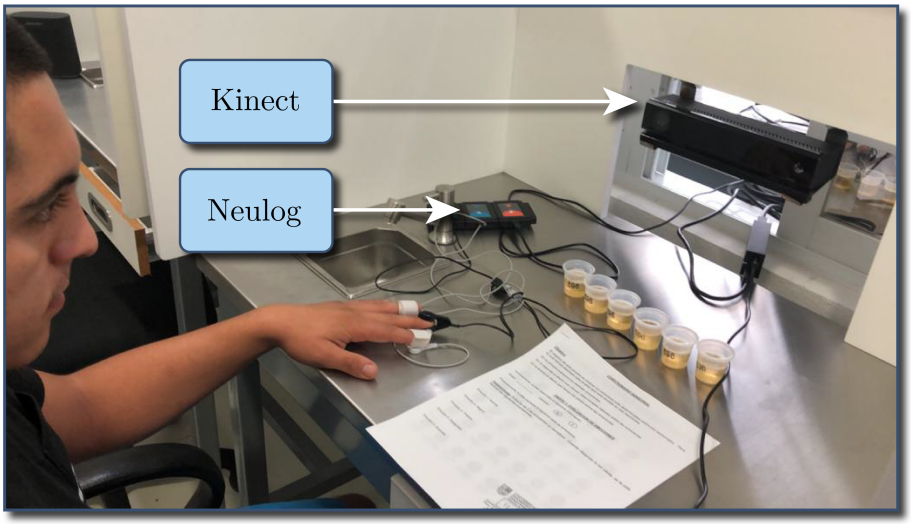


Figura 5.3: Instalación de la cabina en el laboratorio sensorial.

Para este estudio, nos enfocamos únicamente en la imagen frontal de cada participante; el resto de la información será analizado en trabajos futuros. Durante la prueba, cada voluntario llevaba un sensor NeuLog NUL-217 en sus dedos medio y anular, además de un NeuLog NUL-208 en el índice para medir respuesta galvánica de la piel y pulso cardiaco respectivamente. Se construyó un pequeño semáforo para indicarle a cada participante el momento en que debía probar cada muestra, con la intención de sincronizar el comienzo de la captura de video con las reacciones del voluntario. Después de consumir cada muestra, los participantes tomaban agua y galletas saladas para neutralizar los sabores anteriores. Finalmente, se le pidió a cada voluntario que respondiera un cuestionario sensorial.

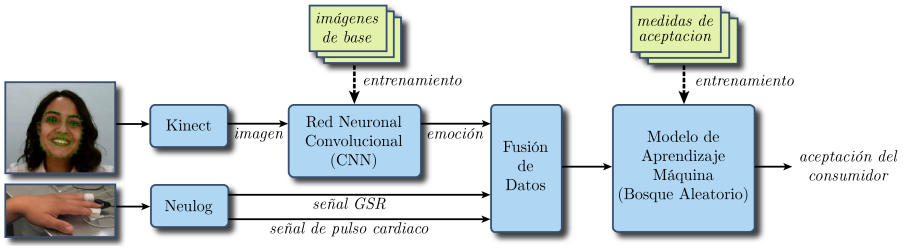


Figura 5.4: Esquema del sistema de análisis

Cuestionario

Se utilizó un cuestionario compuesto por escalas hedónicas de siete puntos para cada una de las muestras de sabor y olor. Este tipo de cuestionarios se emplean

frecuentemente en ciencia sensorial para monitorear la aceptación de distintos tipos de productos alimenticios. Los resultados obtenidos se compararon con los registros de expresiones faciales utilizando métodos de inteligencia artificial.

5.2.2. Sistema de análisis

La Figura 5.4 presenta los módulos principales del sistema utilizado para recabar y analizar la información. El sistema tiene tres entradas: imágenes faciales, señales de RGP y de pulso cardiaco.

Las imágenes faciales fueron procesadas por una RNC obtener una clasificación de las emociones que expresan, misma que se alimentó a un modelo de aprendizaje automático junto con las señales de pulso y RGP para predecir la aceptación del consumidor. El modelo de aprendizaje automático está basado en el método de clasificación por bosques aleatorios. Fue entrenado con la información de los cuestionarios y los resultados de la fase de fusión de datos. Las siguientes secciones expondrán los módulos del sistema a mayor detalle.

Conjuntos de datos de expresiones faciales

Se utilizaron dos distintos conjuntos de expresiones faciales: *AffectNet* [92] y CK+ [93] para entrenar y probar la red neuronal, respectivamente. *AffectNet* contiene más de 420,000 imágenes faciales clasificadas de acuerdo a 11 categorías discretas. Sin embargo, para lograr un entrenamiento más balanceado, solamente se emplearon 3,800 imágenes para cada una de las siguientes categorías: neutral, felicidad, tristeza, sorpresa, miedo, disgusto, enojo, desdén, ninguna e indeterminada. Se excluyeron imágenes que no contenían rostros según la clasificación original. Decidimos utilizar CK+ para evaluar la red dado que es un conjunto muy utilizado en el ramo y contiene una menor cantidad de imágenes categorizadas.

Preprocesamiento de imágenes

Algunas imágenes del conjunto de entrenamiento presentan características irregulares que la red neuronal no puede procesar adecuadamente. Por tanto, se requieren ciertos pasos de preprocesamiento para que la red reciba únicamente información consistente, concretamente:

1. Descartar la información de color, convirtiendo imágenes de tipo RGB a escala de grises, con la intención de reducir su tamaño y tiempo de procesamiento.
2. Detectar todos los rostros mostrados en la imagen, junto con sus rectángulos delimitadores, aplicando un algoritmo basado en histograma de gradientes orientados(HGO) [113].
3. Localizar 68 puntos de interés en el primer rostro detectado, utilizando el algoritmo de Kazemi [114].

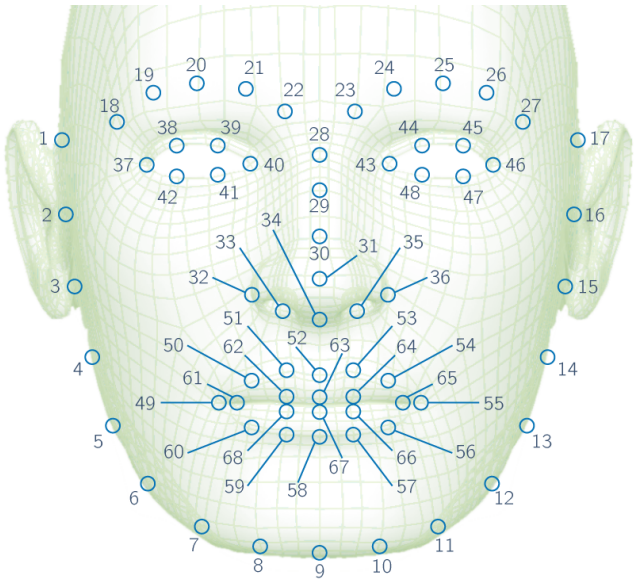


Figura 5.5: Puntos de interés numerados en el rostro.

4. Girar la imagen y sus respectivos puntos de interés para conseguir que la línea que une los puntos 40 y 43 sea horizontal, esto consigue que todos los rostros queden alineados (Figura 5.5).

5. Dividir la imagen facial en cuatro secciones: izquierda y derecha para los ojos y para nariz y boca.

6. Reflexión horizontal de las secciones derechas para poder alimentarlas a las mismas redes que procesan las secciones izquierdas.

7. Aplicar ecualización adaptativa de histogramas limitada por contraste a cada una de las secciones [115].

8. Normalizar el valor de cada pixel convirtiendo del rango (0,255) a (0,1).

Todas estas operaciones se realizan utilizando las librerías Dlib [116] y OpenCV [117] en el lenguaje de programación Python, mientras que la red neuronal fue construida y entrenada utilizando la librería de aprendizaje profundo Keras [118] ejecutada sobre TensorFlow [119].

Arquitectura de la red

La primera etapa está compuesta por dos redes entrenadas de distinta manera ambas construidas según la misma arquitectura mostrada en la Figura 5.6: cada sección (64×64 pixeles) se alimenta a través de tres capas convolucionales y una de agrupación máxima. Posteriormente, otros dos bloques de filtros reducen aún

más la información bidimensional para introducirla en otras cuatro capas densas lineales, la última de las cuales emite una clasificación preliminar de la entrada en una de las 10 categorías posibles por medio de una función de transferencia de tipo *softmax*. Las demás funciones de transferencia son unidades de rectificación lineal (ReLUs).

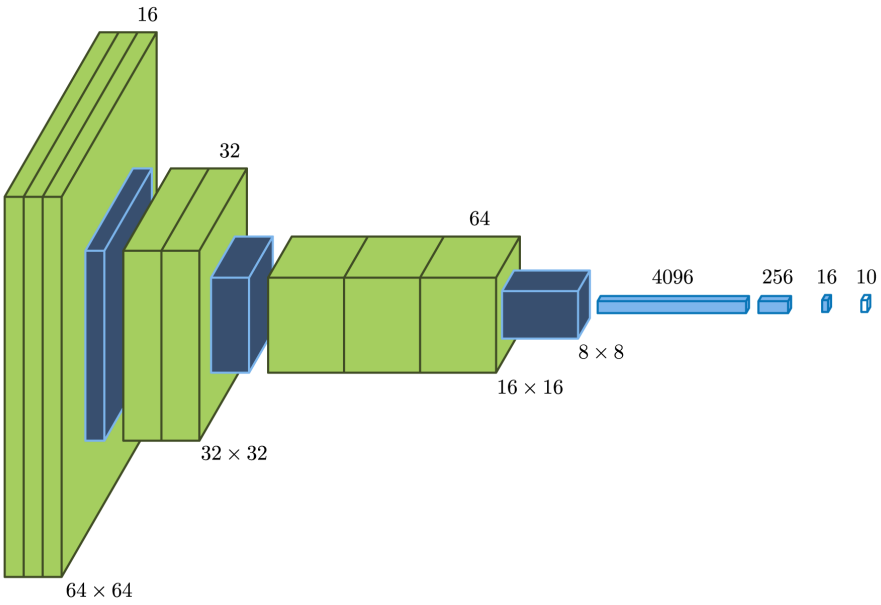


Figura 5.6: Arquitectura para la primera fase de la red neuronal.

La red A produce un vector de 10 números para las secciones correspondientes a los ojos, mientras que la red B hace lo propio para las secciones asociadas a nariz y boca. los cuatro vectores resultantes configuran una entrada de 40 números para la segunda etapa de la red, misma que se compone de dos capas densas ReLU y una capa de salida *softmax* que genera la clasificación final.

Entrenamiento de la red

Solamente 40,336 rostros del conjunto seleccionado fueron utilizados para el entrenamiento, dado que los algoritmos de detección no funcionaron adecuadamente en todos los casos. Reflejando las secciones derechas, se obtuvo un total de 80,672 imágenes faciales para entrenar las redes A y B. Ambas fueron entrenadas a lo largo de 50 épocas con un tamaño de lote de 128 y 20 % del conjunto de entrenamiento para validación. También se fijó un índice de abandono de 0.4 para reducir las probabilidades de sobreajuste.

Posteriormente, las redes procesaron todas las imágenes disponibles para obte-

ner 80,672 vectores de 40 elementos que se utilizaron para entrenar la segunda fase. En el proceso de entrenamiento correspondiente se aplicaron los mismos parámetros que en el primero, salvo por el porcentaje de validación, que esta vez fue del 15 %.

Reconocimiento de emociones

La red ya entrenada se alimentó con todas las imágenes preprocesadas correspondientes a 111 participantes. Con esto se consiguió igual número de archivos CSV con las siguientes columnas: índice de la imagen, número de rostros detectados (o -1 si el algoritmo no pudo encontrar ninguno), nombre del archivo de imagen y probabilidad de clasificación para todas las categorías mencionadas en la subsección 5.2.2.

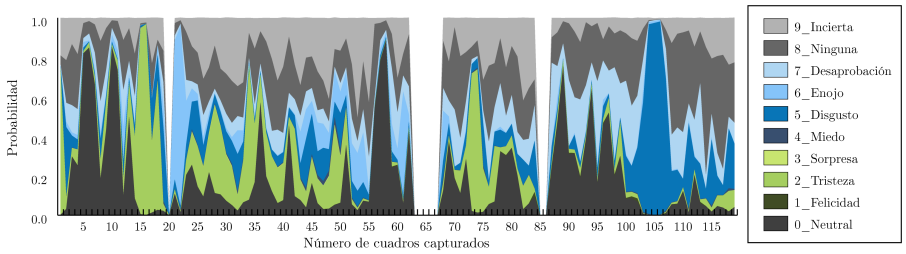


Figura 5.7: Ejemplo de probabilidades para cada emoción detectada.

Fusión de datos

Cada experimento obtiene información de tres fuentes distintas: (1) imágenes que muestran las probabilidades de cada expresión facial en el rango (0,1), (2) RGP y (3) pulso. Como puede apreciarse en la Figura 5.7, las mediciones de los sensores están dispersas a lo largo del tiempo y para cada experimento se registraron varias mediciones. Utilizamos cuatro métricas estadísticas para representar la información obtenida por los sensores: el promedio (avg), la desviación estándar (std), el valor mínimo (min) y el valor máximo (max). En resumen, cada experimento cuenta con 44 características obtenidas a partir de las cuatro métricas estadísticas aplicadas a nueve expresiones faciales, RGP y pulso.

Predicciones de aceptación

Utilizamos técnicas de regresión comunes en aprendizaje automático para predecir la aceptación que los consumidores asignaron a cada prueba. Para cada experimento, extrajimos 44 características de entrada, como se explicó anteriormente, y una salida: el nivel de aceptación asignado por el consumidor a la prueba en cuestión. Cada consumidor evaluó 10 pruebas diferentes.

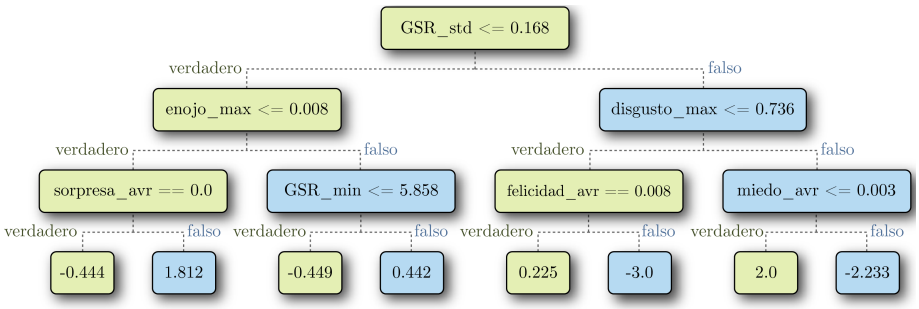


Figura 5.8: Ejemplo de árbol de decisión. Para predecir la evaluación del consumidor, las preguntas deben ser respondidas de arriba hacia abajo, siguiendo la ruta de las respuestas. Al final de la ruta, el último nodo contiene la predicción de la evaluación.

El modelo de aprendizaje automático que seleccionamos fue el de bosque aleatorio, propuesto por Breiman [120]. Consta de un conjunto de árboles de decisión (30 para este trabajo), cada uno creado con un subconjunto aleatorio de pruebas y características extraídas del conjunto de entrenamiento. Un árbol de decisión es un modelo de predicción basado en una serie de preguntas relacionadas con valores específicos de las características (Figura 5.8). La información multidimensional se separa por medio de hiperplanos, que se determinan en función de las preguntas. La idea principal es que pruebas con valores similares tienden a concentrarse en la misma región. Elegimos utilizar bosques aleatorios porque son capaces de calibrar cuánto contribuye cada característica al modelo final (Figura 5.13). Los árboles de decisión establecen criterios de decisión al buscar minimizar la impureza de la información asociada a cada nodo. En este caso, la impureza se calcula como el error cuadrático medio (ECM), formalmente definido en la ecuación (5.1)

$$ECM(\vec{y}, \vec{\hat{y}}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{5.1}$$

Donde $\vec{y}$ y $\vec{\hat{y}}$ son las salidas reales y predichas (los valores de aceptación reportados en el experimento) respectivamente, y n es el número de muestras. Cuando se define una regla de clasificación, la información del nodo se divide en dos regiones. Se probaron varios valores y características, y el par de característica-valor que minimiza la impureza fue el que se seleccionó como regla de clasificación.

La importancia de cada característica es proporcional a la reducción de impureza de todos los nodos relacionados con dicha característica. La reducción de impureza RI en cada nodo j que representa una regla específica, puede calcularse con la ecuación (5.2):

$$RI_j = w_j I_j - (w_{izq} I_{izq} + w_{der} I_{der}), \tag{5.2}$$

Donde *izq* y *der* representan los nodos hijos del nodo j, I representa la impureza de cada nodo, los pesos w son la proporción de cada muestra en los nodos: son calculados dividiendo el número de muestras en el nodo entre el total de muestras. Una vez que se conoce la reducción de impureza en todos los nodos, la importancia de la característica k, IC_k, se calcula mediante la ecuación (5.3):

$$IC_k = \frac{\sum_{j \in N_k} RI_j}{\sum_{j \in N} RI_j}. \tag{5.3}$$

Donde N_k representa el conjunto de todos los nodos que fueron divididos utilizando la variable j, y N representa todos los nodos del árbol de decisión.

Se validaron los resultados por medio de una validación cruzada de 10 iteraciones. Esto quiere decir que el conjunto de datos fue dividido aleatoriamente en diez bloques. Después se entrenó el modelo diez veces, utilizando diez bloques para entrenamiento y uno para pruebas. Se utilizó el error absoluto medio (EAM) para calcular el error del modelo (ecuación 5.4).

$$EAM(\vec{y}, \vec{\hat{y}}) = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{5.4}$$

donde $\vec{y}$ y $\vec{\hat{y}}$ son las salidas reales y predichas, respectivamente. El EAM se calculó cada vez que el modelo fue entrenado y probado. los resultados finales son el promedio de todas las corridas. Decidimos presentar estos resultados con EAM en lugar del ECM utilizado para entrenar el modelo por ser más sencillo de interpretar.

5.3. Resultados

Las Figuras 5.9 y 5.10 muestran los resultados acumulativos para las escalas hedónicas asociadas a las pruebas de sabor y olor, respectivamente. Las barras, centradas en cero, representan cuántos participantes calificaron cada sabor u olor.

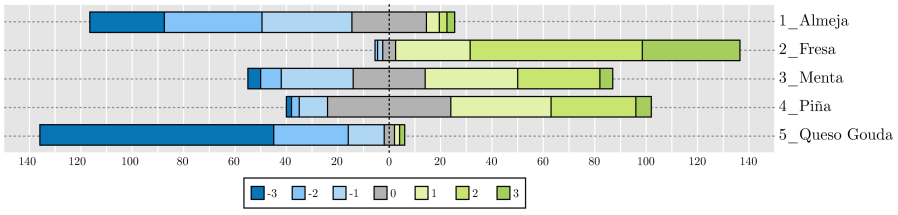


Figura 5.9: Resultados de aceptación para evaluaciones de sabor.

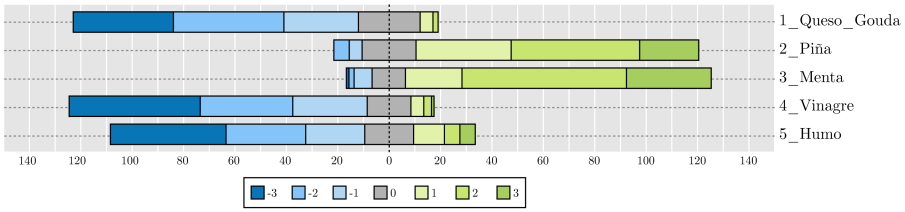
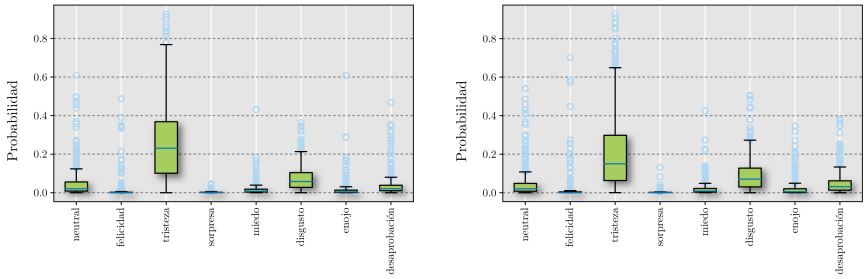


Figura 5.10: Resultados de aceptación para evaluaciones de olor.

El sabor más frecuentemente reportado como agradable es el de fresa, mientras que el de queso parece provocar la mayor reacción de desagrado, dado que su calificación más común es -3 y casi toda la barra queda en el lado izquierdo de la gráfica. El sabor a almeja también obtuvo una calificación general negativa. En cuanto a las pruebas de olor, piña y menta tuvieron buena aceptación, a diferencia de las de queso, vinagre y humo. Parece haber un buen contraste entre los resultados de evaluación asociados a muestras agradables y desagradables.



(a) (b)

Figura 5.11: Emociones detectadas en experimentos de (a) sabor y (b) olor.

La Figura 5.11 muestra las emociones que fueron reconocidas durante los experimentos de sabor (Figura 5.11a) y olor (Figura 5.11b). La gráfica de caja representa el valor promedio de cada emoción para todos los consumidores a lo largo de los cinco experimentos. Puede verse que tristeza es la emoción que aparece con mayor frecuencia durante los experimentos, seguida de disgusto. Estos resultados concuerdan con los obtenidos por He *et al.* [108], quienes reportan haber medido los cambios de expresiones faciales para pruebas de sabor idénticas, similares y distintas. Concluyeron que el agrado producido al consumir un producto alimenticio disminuye rápidamente y encontraron predominancia en las emociones de tristeza y enojo. Adicionalmente, suponemos que las expresiones de tristeza y disgusto pueden deberse a cierto nerviosismo o expectativas inciertas por parte de los voluntarios con respecto al experimento. La Figura 5.12 muestra las matrices de correlación de REF, respuestas de los sensores y aceptación del consumidor en los distintos experimentos. Los valores de la ma-

triz se calcularon por medio del coeficiente absoluto de correlación de Pearson. No se encontró ninguna correlación importante entre la aceptación y las demás variables. Sin embargo, las características que presentan mayor correlación con la aceptación son las siguientes: en la Figura 5.12a: miedo, felicidad, disgusto, pulso y RGP. En la Figura 5.12b: neutral y felicidad. En la Figura 5.12c: RGP, felicidad y disgusto. En la Figura 5.12d: disgusto y neutral.

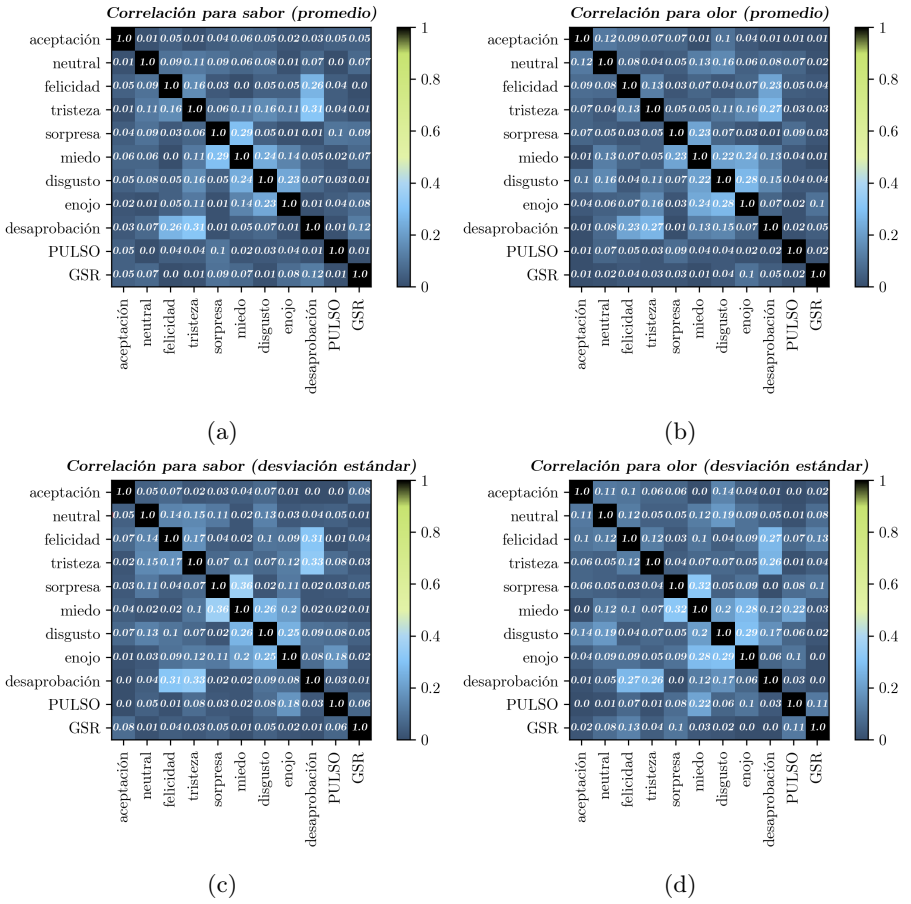
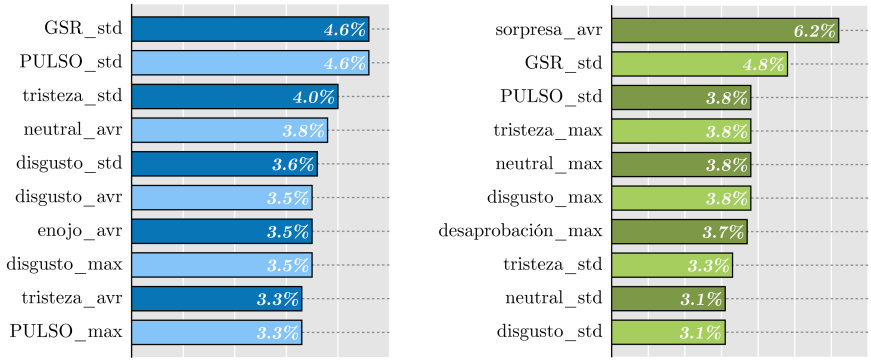


Figura 5.12: Matriz de correlación de REF, respuestas de los sensores y aceptación del consumidor: (a) y (c) exponen las matrices de correlación para las pruebas de sabor, mientras que (b) y (d) corresponden a las pruebas de olor. Las casillas contienen la correlación entre las características marcadas en líneas y columnas.

Podemos ver que felicidad, miedo, disgusto, neutral, pulso y RGP son las características más relacionadas con la aceptación del consumidor. No obstante, la correlación es muy pequeña. El miedo es la emoción más difícil de reconocer en

imágenes estáticas [121] y frecuentemente, humanos y sistemas de reconocimiento la confunden con sorpresa [121, 122]. En la Tabla 5.1 se despliega el EAM de nuestro modelo, tal como se describe en la ecuación (5.4), la cual predice la aceptación de cada muestra en función de los datos obtenidos por el sistema de REF y los sensores. La primera columna describe el tipo de información empleada en el entrenamiento del bosque aleatorio. El modelo obtuvo las mejores predicciones cuando se entrenó solo con las mediciones de RGP. Estos resultados concuerdan con los obtenidos en trabajos anteriores [1, 2].



(a) (b)

Figura 5.13: Importancia de cada variable en los modelos de regresión para (a) sabor y (b) olor.

Como mencionamos en la sección 5.2.2, nuestro modelo de bosque aleatorio calificó la importancia de cada característica en la predicción de la aceptación. Las diez características más relevantes para cada conjunto de experimentos se muestran en la Figura 5.13.

Tabla 5.1: Error medio absoluto (EAM) para el modelo de regresión.

datos	sabor	olor
sensores y emociones	1.8216	1.8593
solo emociones	1.8408	1.8273
solo sensores	1.7896	1.8493
solo RGP	**1.7649**	**1.7817**
solo pulso	1.8173	1.9655

Las desviaciones estándar del pulso y RGP aparecen como las variables más importantes a tomar en cuenta cuando se busca predecir la aceptación. El promedio de las mediciones de sorpresa resulta ser la característica principal en la columna izquierda. Sin embargo, no aparece en la de la derecha. Esto puede deberse al hecho de que el sentido del olfato puede provocar emociones más

intensas que el del gusto. Aún así, las mediciones de emoción son muy similares en ambos experimentos (Figura 5.11). Esto parece indicar que los sensores de pulso y RGP tienen mejor desempeño que la red neuronal en la predicción.

Se generó una gráfica de caja para cada muestra, pero éstas no arrojaron información relevante. Por esta razón, solo mostramos los valores promedio para cada emoción susceptible de ser detectada, como se ve en la Figura 5.11.

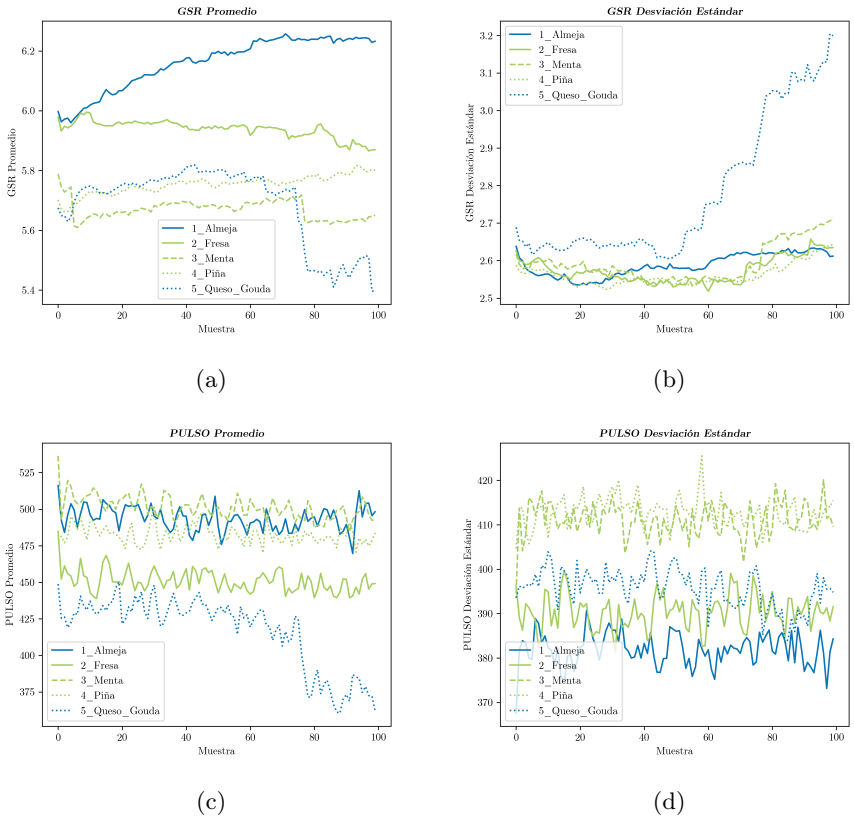


Figura 5.14: Medidas de RGP y pulso para las muestras de sabor.

Calculamos el promedio y desviación estándar para los valores de RGP y pulso de todos los participantes en cada uno de las 100 mediciones obtenidas por los sensores, a razón de 8 por segundo. Las Figuras 5.14 y 5.15 muestran estos resultados: las curvas azules representan los olores y sabores reportados como desagradables, mientras que los agradables aparecen en verde. En la Figura 5.14, dos sabores parecen causar cambios relevantes en las mediciones de los sensores, ambos se reportan claramente como desagradables: el promedio de las muestras aumenta lentamente para 1_Almeja y cae abruptamente para 5_Queso, mien-

tras la desviación estándar de esta última de distingue de las demás gráficas a causa de su marcado aumento. No parece haber disparidad entre las muestras de sabor calificadas como agradables: 2_Fresa, 3_Menta, y 4_Piña. Más aún, permanecen constantes a lo largo del tiempo.

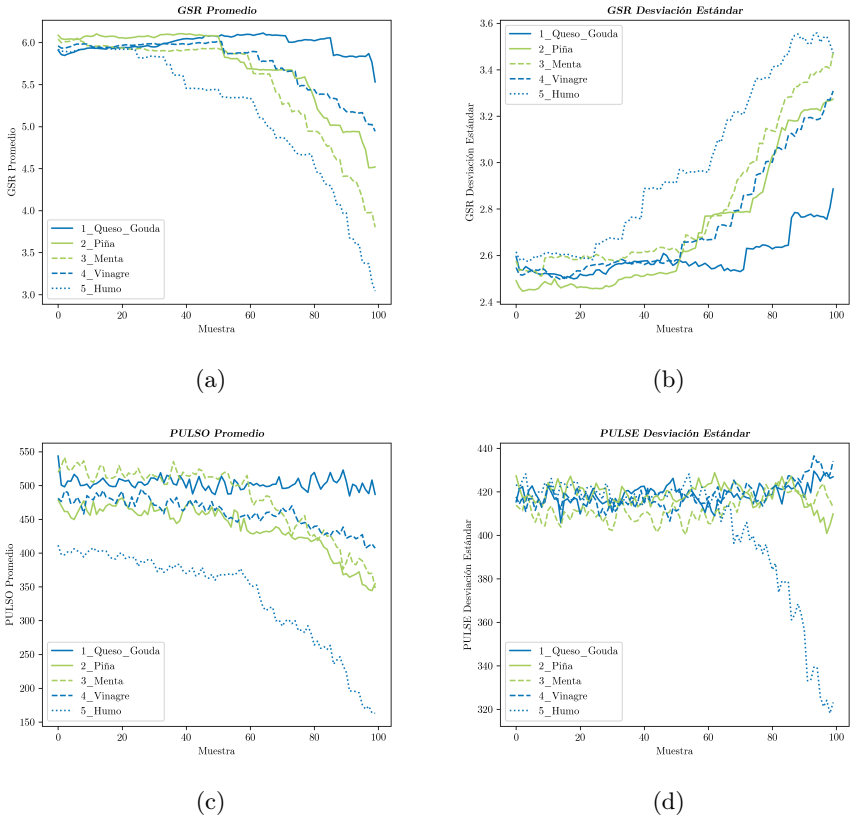


Figura 5.15: Valores de RGP y pulso para las pruebas de olor.

La Figura 5.15 presenta una pauta similar para dos de las curvas asociadas a las muestras desagradables: el valor promedio de RGP en 5_Humo cae, al tiempo que se mantiene sobre los demás en el caso de 1_Queso, con los cambios correspondientes en la desviación estándar. Por otro lado, 4_Vinagre sigue el mismo patrón que las muestras calificadas como altamente agradables: 2_Piña y 3_Menta. El olor del vinagre podría estar provocando reacciones más débiles de lo que reportan las escalas hedónicas. De nuevo, el promedio de lecturas para el pulso despliega curvas características para 1 y 5, pero solamente la desviación estándar de 5 muestra alguna distinción clara. Una vez más, exceptuando 4, los olores reportados como desagradables se separan del resto de alguna manera. Estas características podrían estar relacionadas con reacciones más intensas.

5.4. Discusión

El reconocimiento de expresiones faciales por sí mismo no es un problema que tenga una sola solución claramente marcada, es necesario considerar muchas variables simultáneas. Al día de hoy, la mejor referencia en materia de calificación de emociones observadas en el rostro es la interpretación realizada por humanos, que sigue siendo propensa a errores de clasificación [111], aún después de recibir entrenamiento especializado [123], dado que el reconocimiento de emociones es altamente dependiente del contexto [111, 122], y por ello requiere una comprensión cognitiva de la situación en que una emoción es producida. Además de lo anterior, en experimentos similares al nuestro, [13, 110], los participantes no mostraron prácticamente ninguna alteración facial, incluso al someterse a estímulos intensos [95], y algunas expresiones supuestamente innatas fueron observadas muy rara vez [124]. Todo esto puede ayudar a explicar por qué nuestro sistema de REF, al igual que otros similares encuentran serias dificultades para detectar y clasificar reacciones de tipo emocional.

Aún así, Bredie *et al.* [95] y Crist *et al.* [74] consiguieron producir expresiones de disgusto exitosamente, utilizando soluciones altamente concentradas de cafeína, ácido cítrico y cloruro de sodio. Habrá que considerar la utilización de estímulos similares en trabajos futuros para saber si esto ayuda a la red neuronal a detectar con mayor claridad las expresiones faciales. Gunaratne *et al.* [15] reportan haber encontrado expresiones faciales de tristeza asociadas a la degustación de chocolate salado. Esto puede ser una pista para averiguar por qué nuestro sistema de REF detecta tristeza con tanta frecuencia.

Finalmente, la correlación que se encontró entre emociones y escalas hedónicas es muy baja, al igual que la obtenida por [73]. De aquí es posible concluir que la conexión entre el consumo de alimentos y las emociones experimentadas, así como la que existe entre las emociones reales y las reportadas por el sistema de REF son más débiles de lo esperado, al menos cuando se califican de esta manera.

Por otro lado, el reconocimiento de expresiones faciales para evaluación de productos alimenticios aún no ha sido objeto de muchos estudios, y los algoritmos necesarios siguen en proceso de desarrollo [13]. De cualquier manera, nuestros resultados, como los obtenidos por Samant *et al.* [104], sugieren que la medición de RGP es más confiable para señalar reacciones de tipo emocional. Aunque aún es necesaria una mayor cantidad de investigaciones para confirmar que el agrado producido por un sabor u olor está relacionado con un aumento del ritmo cardiaco según De Wijk *et al.* [125], su proposición de que la intensidad emocional va aunada a una reducción del ritmo cardiaco se ve confirmada hasta cierto punto por las gráficas que aparecen en la Figura 5.15. Estudios posteriores no deberían dejar de lado este tipo de sensores para validar los resultados que aquí presentamos.

Capítulo 6

Conclusiones

En esta tesis hemos presentado un sistema automatizado para analizar las expresiones fisiológicas de la emoción humana por medio de señales biométricas y reconocimiento de expresiones faciales, enfocado principalmente a tratar de determinar la aceptación del consumidor hacia olores y sabores de productos concretos.

Ahora incluimos a manera de conclusión algunas observaciones sobre el trabajo realizado hasta el momento: avances, limitaciones, áreas de oportunidad, además de otras opciones para continuar nuestra investigación en el futuro.

6.1. Avances

- Los experimentos realizados obtuvieron una gran cantidad de información de 120 participantes: video de expresiones faciales, registros electroencefalográficos, cardiacos, etc. Aunque hemos utilizado una buena parte en nuestro análisis, todo este material puede ser útil para otros estudios.

- Se consiguió estructurar y hacer funcionar un sistema de reconocimiento de emociones con algoritmos propios, lo que permite adaptarlo y aplicarlo a otros experimentos similares en el futuro. Su desempeño es cercano al que se puede esperar de otros sistemas comerciales. La Tabla 6.1 contiene los resultados de precisión obtenidos por *Xpress Engine* y *FaceReader* al aplicarlos a imágenes de CK+ [126]. En las columnas de la derecha aparecen los resultados de nuestro sistema, aplicados a imágenes de CK+ y *AffectNet*.

- Después de varias iteraciones, se diseñó una red neuronal convolucional simplificada de buen desempeño y con resultados de exactitud cercanos a los propuestos en el resto de la literatura científica.

Tabla 6.1: Resultados comparativos de la red neuronal

	Otros sistemas		**Nuestra propuesta**	
Emoción	***FaceReader***	**Xpress**	**CK+**	***AffectNet***
Enojo	77.11 %	90.31 %	48.84 %	88.03 %
Disgusto	74.48 %	95.60 %	82.46 %	88.53 %
Felicidad	99.60 %	97.21 %	100.0 %	90.85 %
Neutral	77.42 %	72.84 %	ND	84.07 %
Tristeza	75.97 %	85.71 %	76.92 %	90.14 %
Sorpresa	94.74 %	92.63 %	93.83 %	85.49 %

6.2. Limitaciones

- Los resultados conseguidos parecen sugerir que el reconocimiento de expresiones faciales por sí solo no basta para clasificar satisfactoriamente las emociones del consumidor, al menos dentro del contexto estudiado. La puntuación que el árbol aleatorio asignó a este módulo del sistema parece darle poca importancia en el proceso de clasificación final.
- Los resultados de la clasificación de expresiones faciales varían ampliamente con el tiempo, lo que dificulta su utilización.

6.3. Posibles mejoras

- Experimentar con otras estructuras y metaparámetros en la configuración de la red neuronal podría ayudar a mejorar la exactitud de sus clasificaciones.
- Nuestro sistema de REF clasifica por separado cada una de las imágenes. Dado que se cuenta con una entrada de video, podrían obtenerse mejores resultados utilizando una red neuronal con memoria, que tome en cuenta la relación entre cada imagen del video y las que le preceden en el tiempo.
- En vistas de la importancia que tuvo en el análisis global la medición de la respuesta galvánica de la piel, podría examinarse también por medio de una red neuronal recurrente.
- Varias fases del sistema de análisis se realizaron por separado. Una mejora importante podría consistir en unificarlas para que el sistema trabaje en tiempo real y de manera más independiente del usuario.

6.4. Trabajo futuro

- Debido a su complejidad, los datos obtenidos del electroencefalograma no se incluyeron en el estudio, pero puede ser interesante averiguar qué relación tienen con las emociones clasificadas por el resto del sistema.

- Utilizamos una red neuronal que fusiona la información obtenida de otras dos. En el futuro podríamos diseñar un modo de fusión que incluya también las lecturas de pulso y respuesta galvánica.
- De todas las imágenes clasificadas que contiene la base de datos de *AffectNet*, solamente se utilizó un 10 %, con la intención de que cada una de las categorías tuviera el mismo número de imágenes. Más adelante se podrían volver a entrenar las redes utilizando toda la información disponible en la base de datos.
- No se encontraron grandes diferencias entre algunas de las distintas expresiones faciales mostradas por los participantes en los experimentos, esto hace más difícil que la red neuronal las clasifique convenientemente, pero podría corregirse en experimentos posteriores, dando a los voluntarios muestras con sabores aún más intensos.

Capítulo 7

Lista de publicaciones

[1] **Víctor M. Álvarez** *et al.* «A method for facial emotion recognition based on interest points». En: *2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)*. IEEE. 2018, págs. 1-4.

[2] **Víctor M. Álvarez** *et al.* «Consumer acceptances through facial expressions of encapsulated flavors based on a nanotechnology approach». En: *2018 Nanotechnology for Instrumentation and Measurement (NANOfIM)*. IEEE. 2018, págs. 1-5.

[3] **Víctor M. Álvarez** *et al.* «Facial emotion recognition: a comparison of different landmark-based classifiers». En: *2018 International Conference on Research in Intelligent and Computing in Engineering (RICE)*. IEEE. 2018, págs. 1-4.

[4] **Víctor Manuel Álvarez Pato** y Ramiro Velázquez Guerrero. «Aproximación al reconocimiento de emociones faciales basado en posición de puntos de interés». En: *Pistas Educativas* 39.128 (2018).

[5] Julieta Domínguez Soberanes *et al.* «Food Product Acceptance and Preference Prediction Through Automated Facial Expression Analysis (Medición de Aceptación y Preferencia de Productos Alimenticios Mediante Análisis Automatizado de Expresiones Faciales)». En: *Pistas Educativas* 41.133 (2019).

[6] **Víctor M. Álvarez-Pato** *et al.* «A Multisensor Data Fusion Approach for Predicting Consumer Acceptance of Food Products». En: *Foods* 9.6 (2020), pág. 774.

[7] Julieta Domínguez-Soberanes *et al.* «Determinación de la aceptación de alimentos mediante reacciones fisiológicas del consumidor: un enfoque basado en aprendizaje automático». En: *Revista Ibérica de Sistemas e Tecnologias de Informação* E43 (2021), págs. 418-434.

Referencias

[1] Forbes. *8 de cada 10 productos nuevos fracasan*. https://www.forbes.com.mx/8-de-cada-10-productos-nuevos-fracasan/. Ago. de 2014.

[2] Olusola Ojeh, ed. *The Role of Sensory Analysis in Quality Control*. 2.ª ed. American Society for Testing y Materials (ASTM), 2021.

[3] Kit Yarrow. *Decoding the new consumer mind: how and why we shop and buy*. John Wiley & Sons, 2014.

[4] D Smith. «The role and changing nature of marketing intelligence». En: *Market research handbook* 5 (2007), págs. 3-36.

[5] Morteza Zangeneh Soroush *et al.* «A review on EEG signals based emotion recognition». En: *International Clinical Neuroscience Journal* 4.4 (2017), pág. 118.

[6] Jairo A Rodas y Luz A Montoya-Restrepo. «Medición y Análisis de Anuncios Publicitarios en Televisión con base en las Herramientas Seguidor-de-Visión y Lector-de-Rostro (EyeTracking y FaceReader)». En: *Información tecnológica* 30.2 (2019), págs. 3-10.

[7] Claudia Gonzalez Viejo *et al.* «Integration of non-invasive biometrics with sensory analysis techniques to assess acceptability of beer by consumers». En: *Physiology & behavior* 200 (2019), págs. 139-147.

[8] Raúl Rojas. *Neural networks: a systematic introduction*. Springer Science & Business Media, 2013.

[9] Ali Mollahosseini, David Chan y Mohammad H Mahoor. «Going deeper in facial expression recognition using deep neural networks». En: *2016 IEEE Winter conference on applications of computer vision (WACV)*. IEEE. 2016, págs. 1-10.

[10] Heechul Jung *et al.* «Joint fine-tuning in deep neural networks for facial expression recognition». En: *Proceedings of the IEEE international conference on computer vision*. 2015, págs. 2983-2991.

[11] Tarik A Rashid. «Convolutional neural networks based method for improving facial expression recognition». En: *The international symposium on intelligent systems technologies and applications*. Springer. 2016, págs. 73-84.

[12] Lukas Danner *et al.* «Facial expressions and autonomous nervous system responses elicited by tasting different juices». En: *Food Research International* 64 (2014), págs. 81-90.

[13] Eliza Kostyra *et al.* «Consumer facial expression in relation to smoked ham with the use of face reading technology. The methodological aspects and informative value of research results». En: *Meat science* 119 (2016), págs. 22-31.

[14] Wei He *et al.* «The relation between continuous and discrete emotional responses to food odors with facial expressions and non-verbal reports». En: *Food Quality and Preference* 48 (2016), págs. 130-137.

[15] Thejani M Gunaratne *et al.* «Physiological responses to basic tastes for sensory evaluation of chocolate using biometric techniques». En: *Foods* 8.7 (2019), pág. 243.

[16] Antonio Malo. «Teorías sobre las emociones». En: Philosophica*: Enciclopedia filosófica* on line. Ed. por Francisco Fernández Labastida y Juan Andrés Mercado. 2007. URL: http://www.philosophica.info/archivo/2007/voces/emociones/Emociones.html.

[17] R Plutchik. *The emotions: Facts, theories and a new model. New York, NY, US*. 1962.

[18] Richard S Lazarus y Richard S Lazarus. *Emotion and adaptation*. Oxford University Press on Demand, 1991.

[19] Daniel Goleman. *La Inteligencia Emocional*. Coleccion Edicion Limitada Series. B. Mexico, Ediciones, S.A. de C.V., 2007. ISBN: 9789707102798. URL: https://books.google.com.mx/books?id=bXdq2%5C%5C_rFdlAC.

[20] Martin EP Seligman y Mihaly Csikszentmihalyi. «Positive psychology: An introduction». En: *Flow and the foundations of positive psychology*. Springer, 2014, págs. 279-298.

[21] Brian Parkinson, Agneta Fischer y Antony SR Manstead. *Emotion in social relations: Cultural, group, and interpersonal processes*. Psychology Press, 2005.

[22] Carroll E Izard. «The many meanings/aspects of emotion: Definitions, functions, activation, and regulation». En: *Emotion Review* 2.4 (2010), págs. 363-370.

[23] Kevin Mulligan y Klaus R Scherer. «Toward a working definition of emotion». En: *Emotion Review* 4.4 (2012), págs. 345-357.

[24] *Diccionario de la lengua española*. 23.ª ed. Último acceso 31/07/2020. URL: https://dle.rae.es.

[25] *Merriam-Webster*. Último acceso 31/07/2020. URL: https://www.merriam-webster.com/.

[26] Richard D Lane y Lynn Nadel. *Cognitive neuroscience of emotion*. Oxford University Press, 2002.

[27] Edmund T Rolls. «Precis of the brain and emotion». En: *Behavioral and brain sciences* 23.2 (2000), págs. 177-191.

[28] Charles Darwin. «The expression of the emotions in man and animals». En: *The expression of the emotions in man and animals*. University of Chicago press, 2015.

[29] WILLIAM JAMES. «II.—WHAT IS AN EMOTION ?» En: *Mind* os-IX.34 (abr. de 1884), págs. 188-205. ISSN: 0026-4423. DOI: 10.1093/mind/os-IX.34.188. eprint: https://academic.oup.com/mind/article-pdf/os-IX/34/188/9278514/os-IX_34_188.pdf. URL: https://doi.org/10.1093/mind/os-IX.34.188.

[30] Walter B Cannon. «The James-Lange theory of emotions: a critical examination and an alternative theory». En: *The American journal of psychology* 100.3/4 (1987), págs. 567-586.

[31] Bruce H Friedman. «Feelings and the body: The Jamesian perspective on autonomic specificity of emotion». En: *Biological psychology* 84.3 (2010), págs. 383-393.

[32] Stanley Schachter y Jerome Singer. «Cognitive, social, and physiological determinants of emotional state.» En: *Psychological review* 69.5 (1962), pág. 379.

[33] Richard S Lazarus y Susan Folkman. *Stress, appraisal, and coping*. Springer publishing company, 1984.

[34] Joshua Ian Davis, Ann Senghas y Kevin N Ochsner. «How does facial feedback modulate emotional experience?» En: *Journal of research in personality* 43.5 (2009), págs. 822-829.

[35] Herminia Pasantes. *De neuronas, emociones y motivaciones*. Fondo de cultura económica, 2018.

[36] Yana Suchy. *Clinical neuropsychology of emotion*. Guilford Press, 2011.

[37] Joseph E LeDoux. «Emotion, memory and the brain». En: *Scientific American* 7.1 (1997), págs. 68-75.

[38] Joseph E LeDoux. «The emotional brain: The mysterious underpinnings». En: *New York: Touchstone Book* (1996).

[39] Nico H Frijda *et al. The emotions*. Cambridge University Press, 1986.

[40] Guillaume Chanel, Karim Ansari-Asl y Thierry Pun. «Valence-arousal evaluation using physiological signals in an emotion recall paradigm». En: *2007 IEEE International Conference on Systems, Man and Cybernetics*. IEEE. 2007, págs. 2662-2667.

[41] Joyce HDM Westerink *et al.* «Computing emotion awareness through galvanic skin response and facial electromyography». En: *Probing experience*. Springer, 2008, págs. 149-162.

[42] Paul Ekman. «Emotions revealed». En: *Bmj* 328.Suppl S5 (2004).

[43] Paul Ekman y Wallace V Friesen. «Constants across cultures in the face and emotion.» En: *Journal of personality and social psychology* 17.2 (1971), pág. 124.

[44] Jody Clay-Warner y Dawn T Robinson. «Infrared thermography as a measure of emotion response». En: *Emotion Review* 7.2 (2015), págs. 157-162.

[45] Katarzyna Blinowska y Piotr Durka. «Electroencephalography (eeg)». En: *Wiley encyclopedia of biomedical engineering* (2006).

[46] Jennifer Healey. «Physiological sensing of emotion». En: *The Oxford handbook of affective computing* (2014), pág. 204.

[47] J Marinello Roura y JJ Samsó. «Diagnóstico hemodinámico en angiología y cirugía vascular». En: *Barcelona: Editorial Glosa* (2003), pág. 84.

[48] Kirk Shelley, Stacey Shelley y Carol Lake. «Pulse oximeter waveform: photoelectric plethysmography». En: *Clinical monitoring* (2001), págs. 420-428.

[49] Sergej Lugović, Ivan Dunđer y Marko Horvat. «Techniques and applications of emotion recognition in speech». En: *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE. 2016, págs. 1278-1283.

[50] Andrew T Duchowski y Andrew T Duchowski. *Eye tracking methodology: Theory and practice*. Springer, 2017.

[51] Jakob De Lemos *et al.* «Measuring emotions using eye tracking». En: *Proceedings of measuring behavior*. Vol. 226. 2008, págs. 225-226.

[52] Claudio Aracena *et al.* «Neural networks for emotion recognition based on eye tracking data». En: *2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE. 2015, págs. 2632-2637.

[53] Evgenia Boutsika. «Kinect in education: A proposal for children with autism». En: *Procedia Computer Science* 27 (2014), págs. 123-129.

[54] Tim Beyl *et al.* «Multi kinect people detection for intuitive and safe human robot cooperation in the operating room». En: *2013 16th international conference on advanced robotics (ICAR)*. IEEE. 2013, págs. 1-6.

[55] Zhengyou Zhang. «Microsoft kinect sensor and its effect». En: *IEEE multimedia* 19.2 (2012), págs. 4-10.

[56] Andreas Kaplan y Michael Haenlein. «Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence». En: *Business Horizons* 62.1 (2019), págs. 15-25.

[57] MJ Den Uyl y H Van Kuilenburg. «The FaceReader: Online facial expression recognition». En: *Proceedings of measuring behavior*. Vol. 30. Citeseer. 2005, págs. 589-590.

[58] Bing Liu. *Web data mining: exploring hyperlinks, contents, and usage data*. Springer Science & Business Media, 2007.

[59] Chia-Yin Yu y Chih-Hsiang Ko. «Applying facereader to recognize consumer emotions in graphic styles». En: *Procedia Cirp* 60 (2017), págs. 104-109.

[60] Ariel Ruiz-Garcia *et al.* «A hybrid deep learning neural approach for emotion recognition from facial expressions for socially assistive robots». En: *Neural Computing and Applications* 29.7 (2018), págs. 359-373.

[61] Musaed Alhussein. «Automatic facial emotion recognition using weber local descriptor for e-Healthcare system». En: *Cluster Computing* 19.1 (2016), págs. 99-108.

[62] Henry Candra *et al.* «Classification of facial-emotion expression in the application of psychotherapy using Viola-Jones and Edge-Histogram of Oriented Gradient». En: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2016, págs. 423-426.

[63] Dominika Maison y Beata Pawłowska. «Using the Facereader Method to Detect Emotional Reaction to Controversial Advertising Referring to Sexuality and Homosexuality». En: *Neuroeconomic and Behavioral Aspects of Decision Making*. Springer, 2017, págs. 309-327.

[64] Shlok Gilda *et al.* «Smart music player integrating facial emotion recognition and music mood recommendation». En: *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*. IEEE. 2017, págs. 154-158.

[65] Teena Hassan *et al.* «Automatic detection of pain from facial expressions: a survey». En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).

[66] Agnieszka Landowska y Jakub Miler. «Limitations of emotion recognition in software user experience evaluation context». En: *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE. 2016, págs. 1631-1640.

[67] Kiavash Bahreini, Rob Nadolski y Wim Westera. «Communication skills training exploiting multimodal emotion recognition». En: *Interactive Learning Environments* 25.8 (2017), págs. 1065-1082.

[68] Athanasios Psaltis *et al.* «Multimodal affective state recognition in serious games applications». En: *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE. 2016, págs. 435-439.

[69] Amir Eftekhari *et al. Emotion recognition to match support agents with customers*. US Patent 9,648,171. 59 de 2017.

[70] Victor Shaburov y Yurii Monastyrshyn. *Emotion recognition in video conferencing*. US Patent 9,576,190. 221 de 2017.

[71] Rosalind W Picard. «Affective Computing for HCI.» En: *HCI (1)*. Citeseer. 1999, págs. 829-833.

[72] Chung-Hsien Wu, Ze-Jing Chuang y Yu-Chung Lin. «Emotion recognition from text using semantic labels and separable mixture models». En: *ACM transactions on Asian language information processing (TALIP)* 5.2 (2006), págs. 165-183.

[73] KA Leitch *et al.* «Characterizing consumer emotional response to sweeteners using an emotion terminology questionnaire and facial expression analysis». En: *Food Research International* 76 (2015), págs. 283-292.

[74] CA Crist *et al.* «Automated facial expression analysis for emotional responsivity using an aqueous bitter model». En: *Food Quality and Preference* 68 (2018), págs. 349-359.

[75] Roland P Carpenter, David H Lyon y Terry A Hasdell. *Guidelines for sensory analysis in food product development and quality control*. Springer Science & Business Media, 2000.

[76] Kevin Gurney. *An introduction to neural networks*. CRC press, 2014.

[77] Nikhil Buduma y Nicholas Locascio. *Fundamentals of deep learning: Designing next-generation machine intelligence algorithms*. .º'Reilly Media, Inc.", 2017.

[78] David Kriesel. *A Brief Introduction to Neural Networks*. 2007.

[79] Yves Chauvin y David E Rumelhart. *Backpropagation: theory, architectures, and applications*. Psychology Press, 2013.

[80] Maria MP Petrou y Costas Petrou. *Image processing: the fundamentals*. John Wiley & Sons, 2010.

[81] Rafael C Gonzalez y Richard E Woods. «Digital image processing (preview)». En: (2002).

[82] Stephen M Pizer *et al.* «Adaptive histogram equalization and its variations». En: *Computer vision, graphics, and image processing* 39.3 (1987), págs. 355-368.

[83] Giuseppe Bonaccorso. *Mastering Machine Learning Algorithms: Expert techniques for implementing popular machine learning algorithms, fine-tuning your models, and understanding how they work*. Packt Publishing Ltd, 2020.

[84] Waseem Rawat y Zenghui Wang. «Deep convolutional neural networks for image classification: A comprehensive review». En: *Neural computation* 29.9 (2017), págs. 2352-2449.

[85] Yann LeCun, Yoshua Bengio y Geoffrey Hinton. «Deep learning». En: *nature* 521.7553 (2015), págs. 436-444.

[86] Dominik Scherer, Andreas Müller y Sven Behnke. «Evaluation of pooling operations in convolutional architectures for object recognition». En: *International conference on artificial neural networks*. Springer. 2010, págs. 92-101.

[87] Charu C Aggarwal *et al.* «Neural networks and deep learning». En: *Springer* 10 (2018), págs. 978-3.

[88] Sandro Skansi. *Introduction to Deep Learning: from logical calculus to artificial intelligence*. Springer, 2018.

[89] Forbes. *Heart Rate and Pulse logger sensorNUL-208*. https://neulog.com/heart-rate-pulse/#specs. Ago. de 2017.

[90] OpenCV. *About*. https://opencv.org/about/. 2021.

[91] Davis E. King. «Dlib-ml: A Machine Learning Toolkit». En: *Journal of Machine Learning Research* 10 (2009), págs. 1755-1758.

[92] Ali Mollahosseini, Behzad Hasani y Mohammad H Mahoor. «Affectnet: A database for facial expression, valence, and arousal computing in the wild». En: *IEEE Transactions on Affective Computing* 10.1 (2017), págs. 18-31.

[93] Patrick Lucey *et al.* «The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression». En: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE. 2010, págs. 94-101.

[94] Egon P Köster. «Diversity in the determinants of food choice: A psychological perspective». En: *Food Quality and Preference* 20.2 (2009), págs. 70-82.

[95] Wender LP Bredie, Hui Shan Grace Tan y Karin Wendin. «A comparative study on facially expressed emotions in response to basic tastes». En: *Chemosensory Perception* 7.1 (2014), págs. 1-9.

[96] Shan Li y Weihong Deng. «Deep facial expression recognition: A survey». En: *arXiv preprint arXiv:1804.08348* (2018).

[97] Brais Martinez *et al.* «Automatic analysis of facial actions: A survey». En: *IEEE transactions on affective computing* (2017).

[98] Jie Cai *et al.* «Island loss for learning discriminative features in facial expression recognition». En: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE. 2018, págs. 302-309.

[99] Jianfeng Zhao, Xia Mao y Jian Zhang. «Learning deep facial expression features from image and optical flow sequences using 3D CNN». En: *The Visual Computer* 34.10 (2018), págs. 1461-1475.

[100] Yong Li *et al.* «Occlusion aware facial expression recognition using cnn with attention mechanism». En: *IEEE Transactions on Image Processing* 28.5 (2018), págs. 2439-2450.

[101] Yingying Wang *et al.* «Facial Expression Recognition Based on Auxiliary Models». En: *Algorithms* 12.11 (2019), pág. 227.

[102] Sze-Teng Liong *et al.* «Shallow triple stream three-dimensional cnn (ststnet) for micro-expression recognition». En: *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE. 2019, págs. 1-5.

[103] Benjamin Mahieu *et al.* «Eating chocolate, smelling perfume or watching video advertisement: Does it make any difference on emotional states measured at home using facial expressions?» En: *Food Quality and Preference* 77 (2019), págs. 102-108.

[104] Shilpa S Samant y Han-Seok Seo. «Using both emotional responses and sensory attribute intensities to predict consumer liking and preference toward vegetable juice products». En: *Food Quality and Preference* 73 (2019), págs. 75-85.

[105] Sofie Lagast *et al.* «Consumers' emotions elicited by food: A systematic review of explicit and implicit methods». En: *Trends in food science & technology* 69 (2017), págs. 172-189.

[106] Sylvia D Kreibig. «Autonomic nervous system activity in emotion: A review». En: *Biological psychology* 84.3 (2010), págs. 394-421.

[107] Karin Wendin, Bodil H Allesen-Holm y Wender LP Bredie. «Do facial reactions add new dimensions to measuring sensory responses to basic tastes?» En: *Food Quality and Preference* 22.4 (2011), págs. 346-354.

[108] Wei He *et al.* «Sensory-specific satiety: Added insights from autonomic nervous system responses and facial expressions». En: *Physiology & behavior* 170 (2017), págs. 12-18.

[109] C Beyts *et al.* «A comparison of self-reported emotional and implicit responses to aromas in beer». En: *Food Quality and Preference* 59 (2017), págs. 68-80.

[110] Gaëlle Le Goff y Julien Delarue. «Non-verbal evaluation of acceptance of insect-based products using a simple and holistic analysis of facial expressions». En: *Food Quality and Preference* 56 (2017), págs. 285-293.

[111] Günther Palm y Michael Glodek. «Towards emotion recognition in human computer interaction». En: *Neural nets and surroundings*. Springer, 2013, págs. 323-336.

[112] Hamed Monkaresi *et al.* «Automated detection of engagement using video-based estimation of facial expressions and heart rate». En: *IEEE Transactions on Affective Computing* 8.1 (2016), págs. 15-28.

[113] Navneet Dalal y Bill Triggs. «Histograms of oriented gradients for human detection». En: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. Vol. 1. IEEE. 2005, págs. 886-893.

[114] Vahid Kazemi y Josephine Sullivan. «One millisecond face alignment with an ensemble of regression trees». En: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, págs. 1867-1874.

[115] Karel Zuiderveld. «Contrast limited adaptive histogram equalization». En: *Graphics gems IV*. Academic Press Professional, Inc. 1994, págs. 474-485.

[116] Davis E King. «Dlib-ml: A machine learning toolkit». En: *Journal of Machine Learning Research* 10.Jul (2009), págs. 1755-1758.

[117] G. Bradski. «The OpenCV Library». En: *Dr. Dobb's Journal of Software Tools* (2000).

[118] François Chollet *et al. Keras*. 2015.

[119] Martín Abadi *et al.* «Tensorflow: A system for large-scale machine learning». En: *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*. 2016, págs. 265-283.

[120] Leo Breiman. «Random forests». En: *Machine learning* 45.1 (2001), págs. 5-32.

[121] Helen Rodger *et al.* «Mapping the development of facial expression recognition». En: *Developmental science* 18.6 (2015), págs. 926-939.

[122] Manuel G Calvo y Lauri Nummenmaa. «Perceptual and affective mechanisms in facial expression recognition: An integrative review». En: *Cognition and Emotion* 30.6 (2016), págs. 1081-1106.

[123] Yangyang Du *et al.* «Perceptual learning of facial expressions». En: *Vision research* 128 (2016), págs. 19-29.

[124] Jessica E Armstrong, David G Laing y Anthony L Jinks. «Taste-Elicited Activity in Facial Muscle Regions in 5–8-Week-Old Infants». En: *Chemical senses* 42.5 (2017), págs. 443-453.

[125] René A de Wijk *et al.* «ANS responses and facial expressions differentiate between the taste of commercial breakfast drinks». En: *PloS one* 9.4 (2014), e93823.

[126] Grzegorz Brodny *et al.* «Comparison of selected off-the-shelf solutions for emotion recognition based on facial expressions». En: *2016 9th International Conference on Human System Interactions (HSI)*. IEEE. 2016, págs. 397-404.