



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias de la Computación

Detección de depresión con técnicas de inteligencia artificial.

Tesis para obtener el grado de Licenciatura en Ingeniería en ciencias de la computación

Presenta:

Luis Angel Herrera Maldonado

Director:

Beatriz Beltrán Martínez

Asesor:

Beatriz Beltrán Martínez

24 de mayo de 2022

Agradecimientos

Quisiera agradecer a todas las personas que estuvieron conmigo durante mi estancia en la universidad, mis compañeros de clase, mis amigos que siempre me motivaron a ser mejor persona y mejor estudiante, mis profesores de los cuales obtuve mucho conocimiento y especialmente a mi familia que me ayudo a cumplir este sueño tan importante para mi, cada uno de ellos me apporto el apoyo que necesité para culminar mi carrera universitaria. También quiero agradecer a mi asesora de tesis, la doctora Beatriz Beltrán Martínez que me apporto su conocimiento y su tiempo durante la elaboración de este proyecto. Por ultimo y no por eso menos importante, quiero agradecer a Jessica Olivares por darme su apoyo durante esta nueva etapa tan importante en mi vida.

*Si he visto más lejos, es poniéndome sobre los hombros de Gigantes.
- Isaac Newton, 1675.*

Índice general

1. Introducción	4
1.1. Objetivos	4
1.1.1. Objetivo general	5
1.1.2. Objetivos específicos	5
1.2. Preguntas de investigación	5
1.3. Hipótesis y Justificación	5
1.4. Estructura de la tesis	6
2. Marco Teorico	8
2.1. Emociones	8
2.2. Expresiones faciales	10
2.3. Depresión	12
2.4. Redes neuronales	14
2.4.1. Neurona biológica	15
2.4.2. Modelo neuronal artificial	18
2.5. Redes neuronales convolucionales	19
3. Estado del arte	25
4. Metodología	30
4.1. Definición del problema	30
4.2. Análisis de factibilidad	31
4.2.1. Factibilidad Operativa	31
4.2.2. Factibilidad Técnica	31
4.2.3. Factibilidad Económica	32
4.3. Requerimientos del sistema	33
4.4. Cronograma de actividades	34
4.5. Modelado de software	36
4.5.1. Sistema de software	37
4.5.1.1. Diagrama de casos de uso	38
4.5.1.2. Diagrama de clases	39

4.5.1.3.	Diagramas de secuencia	40
4.5.1.4.	Diseño de interfaz gráfica	42
4.5.2.	Red neuronal	43
4.5.2.1.	Diseño de arquitectura	43
4.5.2.2.	Conjunto de datos	45
4.5.2.3.	Métricas	48
5.	Experimentos y resultados	49
5.1.	Etapa de entrenamiento	49
5.2.	Etapa de pruebas	50
5.3.	Resultados finales	54
6.	Conclusión y trabajo futuro	59
6.1.	Conclusión	59
6.2.	Trabajo futuro	60
Anexo		61

Capítulo 1

Introducción

La depresión es una enfermedad mental que, según la organización mundial de la salud (OMS), afecta a más de 300 millones de personas en todo el mundo [1], algunas de ellas han vivido diariamente con depresión sin darse cuenta, por ello es importante desarrollar tecnologías que ayuden a hacer un diagnóstico rápido, y de esa manera, evitar que la calidad de vida de estas personas se vea afectada severamente. La expresión facial brinda información necesaria para reconocer las emociones en una persona, de acuerdo con el psicólogo Paul Ekman, existen seis emociones con expresiones faciales universales: enojo, tristeza, alegría, disgusto, miedo y sorpresa [2]. Durante varios años se han realizado trabajos de investigación para determinar cuáles son las emociones que están relacionadas con dicha enfermedad. De acuerdo con [3] las unidades de acción (AU) que están presentes en una persona con depresión son: AU12, AU14 y AU10 (ver el apartado del anexo para más información sobre las unidades de acción), estas AU están relacionadas con las emociones felicidad, disgusto y desprecio, sin embargo en [4] se han realizado experimentos de detección de depresión usando las unidades de acción siguientes: AU6, AU12, AU15 Y AU26, las cuales están relacionadas con las emociones felicidad y tristeza. Las expresiones faciales pueden ser buenas para detectar depresión, sin embargo, una técnica más eficiente de detección es el análisis de texto. En este trabajo de investigación, se propone desarrollar una tecnología con capacidad analizar de expresiones faciales para identificar esta enfermedad.

1.1. Objetivos

En este trabajo de investigación se tiene el objetivo de proponer un método que nos permita realizar un diagnóstico rápido de depresión.

1.1.1. Objetivo general

Se establece como objetivo general diseñar una arquitectura de inteligencia artificial con la capacidad de realizar un diagnóstico rápido de depresión utilizando las expresiones faciales de una persona.

1.1.2. Objetivos específicos

A continuación se listarán los objetivos específicos los cuales nos ayudaran a cumplir el objetivo general.

- Seleccionar información general sobre las emociones.
- Explicar la relación entre las emociones y la depresión.
- Justificar la relación entre las emociones y las expresiones faciales.
- Probar métodos que son utilizados en el tratamiento de imágenes.
- Proponer una técnica de inteligencia artificial para la clasificación de emociones por expresiones faciales.
- Experimentar con la inteligencia artificial propuesta.
- Evaluar la inteligencia artificial.

1.2. Preguntas de investigación

A continuación se presentan las preguntas científicas que se responderán a lo largo de este trabajo escrito.

- ¿Una inteligencia artificial puede diagnosticar depresión?
- ¿Existen emociones que están relacionadas con la depresión?
- ¿Las expresiones faciales pueden indicar si una persona tiene depresión?

1.3. Hipótesis y Justificación

Una de las aplicaciones de la Inteligencia Artificial(IA) es la clasificación de imágenes y por consecuente, se puede desarrollar una arquitectura que a través de expresiones faciales reconozca las emociones de una persona y en base a las emociones relacionadas con la depresión se realice un diagnóstico depresivo.

La inteligencia artificial ha tenido aplicaciones en varias áreas como la robótica, ciberseguridad, biología, química y áreas de la salud, sin embargo, hasta este momento no es muy común que los expertos en inteligencia artificial propongan un método que sea enfocado al área psicológica y trastornos mentales. Se sabe que cada año aproximadamente se suicidan 800,000 personas en todo el mundo, esta cifra es preocupante ya que el 95 % tuvo algún síntoma de depresión. Muchas veces las personas con depresión no acuden con un experto debido a que los tratamientos son costosos o algunas veces las personas creen que solo es una etapa de tristeza y no obtienen un diagnóstico por parte de un experto. Generalmente las personas con bajos recursos no tienen la economía para asistir con un experto a realizarse un diagnóstico depresivo y por eso es necesario que se desarrolle un método con el cual una persona pueda realizarse un diagnóstico rápido con una baja probabilidad de error.

1.4. Estructura de la tesis

Este trabajo de tesis esta estructurado por capítulos y secciones. Cada uno de los capítulos que se han establecido cumplen con explicar una parte importante de este trabajo.

La estructura es la siguiente:

- **Introducción:** En este capítulo se establece la introducción de la tesis, objetivos generales y específicos, preguntas de investigación, hipótesis y justificación.
- **Marco Teórico:** En este capítulo se especifican y se explican los temas que son necesarios para entender el proyecto propuesto.
- **Estado del arte:** En este capítulo se mencionan algunos de los trabajos relacionados con este proyecto que se han realizado anteriormente.
- **Metodología:** En este capítulo se detalla el modelo, la estructura y los métodos que serán utilizados en el desarrollo de este proyecto.
- **Resultados:** En este capítulo se muestran los resultados tanto en el software como en la arquitectura de inteligencia artificial.
- **Conclusión y trabajo futuro:** En este capítulo se cerrará este trabajo mencionando las conclusiones finales y el trabajo futuro que se realizará para mejorar este proyecto.

- **Anexo:** En este último apartado se incluye información complementaria sobre el trabajo que se realizó.

Capítulo 2

Marco Teorico

Para entender el funcionamiento del sistema se deben explicar cuales son las bases científicas que han sido consultadas y que han sido de inspiración para llevar acabo esta investigación. Debido a que el proyecto se basa en depresión, primero se debe de conocer sobre las emociones y algunos de los temas con los que tiene relación. Además, debido a que en este proyecto se harán uso de las expresiones faciales de una persona es obligatorio conocer que es lo que los autores especialistas en el tema dicen sobre ellas y como se relacionan con las emociones. En este capítulo se abordarán los temas que son de suma importancia en este proyecto y su explicación debe ser imprescindible.

2.1. Emociones

Las emociones en los seres humanos es algo difícil de explicar, sin embargo todos hemos sentido algunas de las emociones más comunes como lo son: la tristeza, la alegría, ira, etc. Sobre la naturaleza de las emociones todavía existen muchas preguntas sin la adecuada respuesta [5]. Empecemos definiendo que son las emociones y como influyen en nuestro día a día. De acuerdo con el sociólogo Norman Kent Denzin “las emociones son una experiencia corporal viva, veraz, situada y transitoria que es percibida en el interior y recorriendo el cuerpo, y que, durante el transcurso de la vivencia, sume a la persona y a sus acompañantes en una realidad nueva y transformada - la realidad de un mundo constituido por la experiencia emocional” [6]. Algunos otros expertos en el campo de las emociones han dado otras definiciones, por ejemplo el sociólogo Theodore D. Kemper define las emociones como “una compleja y organizada predisposición a participar en ciertas clases de conductas biológicamente adaptativas ... caracterizada por unos peculiares

estados de excitación fisiológica, unos peculiares sentimientos o estados afectivos, un peculiar estado de receptividad y una peculiar pauta de reacciones expresiva” [7]. Por otro lado el psicólogo Edward E. Lawler define las emociones como “estados evaluativos, sean positivos o negativos, relativamente breves, que tienen elementos fisiológicos, neurológicos y cognitivos” [5].

Las emociones a simple vista parecen sencillas de entender, sin embargo, cuando nos adentramos más a ellas se vuelven más complejas y más difíciles de explicar.

Casi todo el mundo piensa que sabe qué es una emoción hasta que intenta definirla. En ese momento prácticamente nadie afirma poder entenderla [8].

Se entiende por emoción como experiencia multidimensional que involucra 3 sistemas de respuesta: cognitivo, conductual y fisiológico. Esto está en sincronía con el modelo tridimensional de la ansiedad que fue propuesto por Lang [9].

Las emociones que siente una persona no solo son respuestas fisiológicas sino que cada persona puede experimentar diferentes experiencias emocionales, esto dependerá de diversos factores, como por ejemplo, sus expectativas ante la situación, la identidad social activa en cada momento o la identificación de la persona con otros grupos sociales. De acuerdo con Leslie Brody en su teoría de la apreciación, los seres humanos no somos mecanismos biológicos sintientes, sino que valoramos cognitivamente los elementos del entorno antes de experimentar o de expresar una emoción [10]. Muchos autores han llegado a la conclusión que las emociones dependen de varios factores. El pedagogo L. Alan Sroufe dice que las emociones tienen varias facetas, lo que implica la consideración de factores cognitivos, sociales y comportamentales [11].

La teoría de atribución de Edward E. Lawler et al [12] menciona que la emoción experimentada por una persona, no dependerá solo de él sino que también de la atribución causal que realice la persona. Si el sujeto cree que otro ha sido el causante de un hecho indeseable, sentirá ira, en caso de que considere que él es el responsable sentirá culpa o pena y por último si considera que el hecho fue causa del destino, las emociones que sentirá serán tristeza o desesperación. Debemos aclarar que la experiencia emocional y la expresión emocional son dos términos completamente diferentes, podemos expresar lo que sentimos mediante palabras, gestos faciales, tonos vocales, conductas y cambios fisiológicos. La expresión emocional no puede quedar reducida a una simple manifestación de un estado interno, pues la expresión, orientada comunicativamente hacia el otro, emerge en el contexto de la interacción social [13].

2.2. Expresiones faciales

En 1872 Darwin publicó en su libro llamado "La expresión de las emociones en hombres y animales"[14] que en algunas emociones las expresiones faciales eran universales. De acuerdo con las observaciones de Darwin, existen tres principios donde explica el origen y el desarrollo de casi todas las expresiones faciales y algunos gestos que son involuntarios al sentir alguna sensación o emoción [14].

- Si se repiten a menudo los movimientos útiles para satisfacer algún deseo o aliviar alguna sensación, llegan a hacerse tan habituales que se ejecutan, sean o no de utilidad, cada vez que se siente el mismo deseo o sensación incluso en un grado muy débil.
- El hábito de ejecutar voluntariamente movimientos opuestos ha llegado a establecerse en nosotros mediante la práctica de toda nuestra vida.
- Es la acción directa del sistema nervioso excitado sobre el cuerpo, con independencia de la voluntad y en gran medida con la independencia del hábito.

La oblicuidad de las cejas y el estiramiento de los ángulos de la boca hacia abajo provienen del esfuerzo para prevenir un arranque de llanto a gritos o terminarlo después de haber surgido. Este movimiento expresivo se deriva del intento de controlar o prevenir algunos otros movimientos expresivos no deseables. En estas situaciones, la conciencia y la voluntad debieron de haber intervenido en un principio con unos fines concretos, como el escapar de algún peligro, el aliviar molestias o satisfacer deseos [14].

El método que Darwin utilizó para realizar el estudio de las expresiones faciales consistió en utilizar fotografías que el Dr. Duchenne tomó a un paciente que no tenía sensibilidad en el rostro y que de esta manera pudo galvanizar ciertos músculos de la cara obteniendo de esta manera ciertas expresiones. Sin embargo, este método solo fue utilizado para culturas europeas, por otro lado, para las culturas no europeas, Darwin realizó un cuestionario de 16 preguntas sobre descripciones de expresiones faciales. Con ayuda de las respuestas de los cuestionarios que Darwin obtuvo de sus contactos, pudo afirmar que la teoría de las expresiones faciales de las emociones eran universales [15]. Darwin concluyó sus investigaciones diciendo que todas las expresiones más importantes exhibidas por el hombre son iguales a los largo de todo el mundo [14].

En el siglo XX el psicólogo Paul Ekman decidió retomar parte de la investigación relacionada con las expresiones faciales. En los años sesenta decidieron hacer sus propios experimentos en culturas letradas y ágrafas para

demostrar la universalidad de las expresiones faciales. El experimento consistió en mostrar 30 imágenes de las expresiones de catorce personas, las cuales expresaban seis emociones que han sido consideradas como primarias: felicidad, miedo, sorpresa, tristeza, ira y asco. Después de realizar los experimentos en las culturas letradas, los investigadores notaron que los resultados de las interpretaciones fueron casi las mismas en la mayoría de las personas encuestadas. Al realizar el mismo experimento en una de las culturas ágrafas, Paul Ekman tuvo ciertas dificultades ya que al encuestar a personas que no leían ningún idioma se les tuvo que enseñar los nombres de las emociones. Los resultados que obtuvo del experimento en la cultura ágrafa se mostraron diferentes comparados con los resultados que se realizaron en las culturas letradas. Después de un año Paul Ekman llevo a cabo un nuevo experimento, pero esta vez lo hicieron con otra cultura ágrafa, en este nuevo experimento la prueba fue modificada, ahora a las personas se les entregaría dos o tres fotografías relatando breves historias y ellos debían seleccionar la foto que fuera más adecuada al relato. Estos resultados fueron muy parecidos a los resultados que se obtuvieron con las culturas letradas. Al igual que Darwin, Paul Ekman concluyó que existen expresiones faciales que mantienen características universales en las especies humanas [16].

Las expresiones faciales de las diferentes emociones universales son causadas debido al movimiento de ciertos músculos del rostro. El psicólogo Paul Ekman definió un sistema de codificación facial *The Facial Action Coding System* (FACS por sus siglas en inglés)[17]. El sistema de codificación FACS se basa en la observación del movimiento de los músculos faciales, estos fueron llamados Unidades de Acción (AU). Ekman propuso un conjunto de AU's de cardinalidad 44. Los investigadores de [18] solo utilizaron 18 de las 44 AU's propuestas por Paul Ekman para describir las seis emociones universales. En la figura 2.1 se pueden observar las seis emociones que se consideran primarias.



Figura 2.1: Imagen en donde se muestran las seis emociones básicas, imagen tomada del sitio de internet [19]

En la tabla 2.1 se listan las AU que se activan dependiendo de la emoción que la persona sienta.

Emociones	Unidades de Acción
Felicidad	6, 12, 16, 25, 26
Sorpresa	1, 2, 5, 26, 27
Disgusto	9, 10, 17, 25, 26
Ira	2, 4, 7, 16, 17, 23, 24, 25, 26
Miedo	1, 4, 5, 7, 20, 25, 26
Tristeza	1, 4, 7, 15, 17, 25, 26

Cuadro 2.1: Unidades de acción de cada una de las seis emociones básicas.

2.3. Depresión

Como se describe en la introducción, la depresión es una enfermedad mental que afecta a muchas personas alrededor de todo el mundo y conlleva a que las personas ya no realicen sus actividades escolares, laborales y familiares. Muchas personas creen que la depresión es algo con lo que puede lidiar uno mismo y que no es grave, sin embargo, según los datos de la Organización

Mundial de la Salud (OMS) [1] menciona que cada año se suicidan aproximadamente 800,000 personas en todo el mundo y además el suicidio es la segunda causa de muerte en el grupo etario de 15 a 29 años.

La mayoría de las personas que sufren de esta enfermedad no han sido atendidos de manera eficaz, uno de los retos que se enfrentan es la falta de recursos en sus países, la falta de seriedad de la enfermedad, además, otro factor que influye son las personas que estigmatizan y juzgan los trastornos mentales. Aunque algunas personas si acudan con un especialista, estas personas suelen ser diagnosticadas erróneamente, eso significaría que en algunos países no existe una preparación adecuada de especialistas en temas de depresión.

La depresión puede ser el resultado de interacciones que han sido complejas en factores sociales, psicológicos y biológicos. Quienes han pasado por circunstancias que han sido adversas por ejemplo desempleo, luto o traumas psicológicos tienen más probabilidad de padecer esta enfermedad.

Aunque en párrafos anteriores ya se ha hablado sobre la depresión, aun no se ha hecho una definición formal. De acuerdo a la literatura que se ha consultado, el Dr. Pedro Ratamal en su libro llamado "Depresión" define a esta enfermedad como una alteración del estado de ánimo y consistente en su disminución, con un grado variable de pérdida de interés o dificultad para experimentar placer en las actividades cotidianas y acompañado de diversos síntomas psíquicos (tristeza, alteraciones de la concentración, memoria, etc.) y físicos (disminución de la libido, anorexia-hiperfagia, etc.) [20].

Los síntomas y los signos de la depresión pueden ser ordenados de la siguiente manera [20]: *Alteraciones emocionales* (tristeza patológica, angustia e irritabilidad). *Alteraciones del pensamiento* (fallas de concentración y memoria, desinterés indecisión, desesperanza, ideación delirante, ideación suicida). *Alteraciones somáticas* (insomnio-hipersomnio, anorexia-hiperfagia, disminución o aumento de peso, disminución de la libido, fatiga, algias). *Alteraciones de los ritmos vitales* (estar peor a ciertas horas del día, habitualmente en las mañanas, cambios menstruales, tendencia a la presentación estacional de los síntomas). *Alteraciones de la conducta* (llanto, agitación, lentificación, aislamiento, mutismo).

Se han demostrado un conjunto de síntomas que es llamado "depresión endógena" [21]. Los síntomas de la depresión endógena son: pérdida de intereses en la comida y en el sexo, dilación psicomotora, así como trastornos del sueño. Este tipo de depresión se define como un estado de depresión en el que la causa primaria precipitante reside en los factores biológicos o somáticos, en contraste a los factores situacionales que acontecen en el ambiente de la persona [22][23].

Por otro lado, existe un tipo de depresión que es llamada "depresión por

desesperanza” y según los expertos, es causada por estresores ambientales. De acuerdo a la teoría propuesta por Abramson et al [24] dice que la depresión por desesperanza es un producto de la combinación de un estilo de atribución negativo con un acontecimiento vital negativo. Algunos de los síntomas que se presentan en una persona con depresión por desesperanza son: apatía, falta de energía y un sentimiento fuerte de desesperanza con respecto a las situaciones cotidianas.

En el Manual Diagnóstico y Estadístico de los Trastornos Mentales 5ta edición (DSM5) [25] menciona dos tipos de depresión, una es llamada “depresión mayor” y la otra es conocida como “distimia” (depresión menor). Como se ha explicado en párrafos anteriores, en ambos tipos de depresión el síntoma más notorio es la disminución del estado de animo. Sin embargo el DSM5 menciona que la depresión mayor tiene un inicio claro y debe perdurar por lo menos 2 semanas y puede ser distinguido de una manera fácil. Por otro lado, la distimia es una depresión que es menor, con síntomas menores y excluye síntomas más graves como la ideación psicótica o suicida y que dura por lo menos dos años y puede ser confundida con el carácter de una persona.

Los profesionales en el área de psicología pueden ofrecer tratamientos basados en diferentes terapias como la terapia cognitivo conductual que fue propuesta por el psiquiatra Aaron Temkin Beck [26], ésta terapia no es la única que los especialistas ofrecen, también se ocupa la terapia de activación conductual y la psicoterapia interpersonal. También suelen recetar medicamentos antidepressivos, como los inhibidores selectivos de la recaptación de la serotonina y los antidepressivos tricíclicos. De acuerdo a la OMS [1] los antidepressivos pueden ser eficaces en la depresión moderada o grave, pero no son el mejor tratamiento en los casos de depresión leve y no se deben utilizar para tratar la depresión en niños ni como tratamiento de primera línea en adolescentes.

Debido a que esta enfermedad es un problema grave que afecta a la sociedad, la OMS ha creado manuales sobre intervenciones breves que ayudan a tratar la depresión y que pueden ser utilizados por personas que no son expertas en el área de la psicología.

2.4. Redes neuronales

A través del tiempo, los científicos han querido imitar el comportamiento biológico del cerebro humano, después de mucho trabajo de investigación y experimentación crearon un sistema que nombraron “redes neuronales” o también “neuroredes”, el trabajo de las redes neuronales es emular la forma en la que el sistema nervioso procesa la información. Después de una basta

y extensa investigación compleja acerca del funcionamiento del cerebro, los científicos descubrieron que el cerebro humano es un sistema complejo no lineal y paralelo, en otras palabras, el cerebro humano puede procesar mucha información compleja simultáneamente, esta es una gran diferencia si el cerebro humano es comparado con una computadora ya que las computadoras son máquinas secuenciales que solo pueden procesar una operación a la vez. Una vez explicado esto, definamos que es una red neuronal. En [27] se menciona que una red neuronal es un procesador de información, de distribución altamente paralela, constituido por unidades de procesamiento llamadas “neuronas”. Una red neuronal tiene las siguientes características.

- Tener una inclinación natural a adquirir el conocimiento a través de la experiencia, el cual es almacenado, al igual que en el cerebro, en el peso relativo de las conexiones neuronales.
- Tienen una alta plasticidad y gran adaptabilidad y son capaces de cambiar dinámicamente junto con el medio.
- Poseen un alto nivel de tolerancia a fallas, es decir, pueden sufrir un daño considerable y continuar teniendo un buen comportamiento, al igual que en los sistemas biológicos.
- Tener un comportamiento altamente no-lineal, lo que permite procesar información de otros fenómenos no lineales.

2.4.1. Neurona biológica

De acuerdo con los expertos, una neurona biológica tiene el aspecto que se muestra en la imagen 2.2

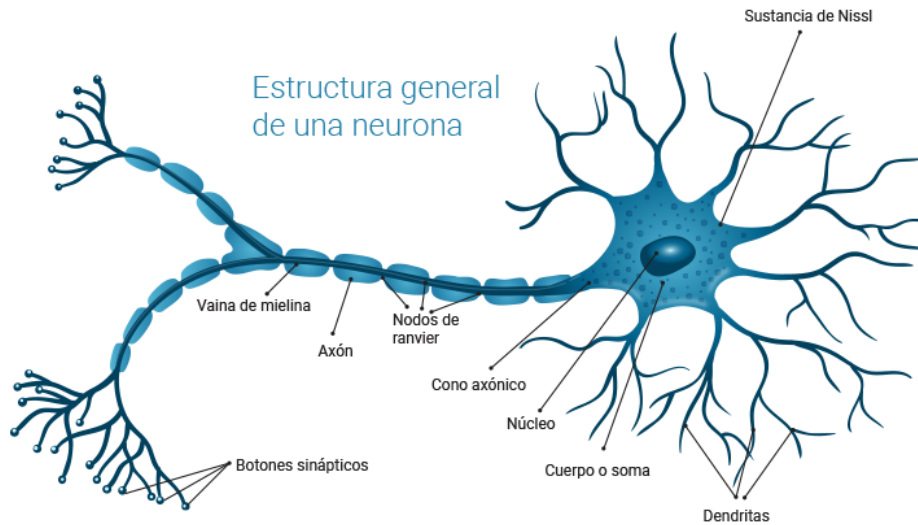


Figura 2.2: Neurona biológica, imagen tomada de [28].

Casi todas las neuronas biológicas codifican el resultado de salida como un conjunto de pulsos eléctricos que son llamados potenciales de acción [27], son creados en el cuerpo celular y de ahí son transportados por el axón, después, estos pulsos eléctricos llegan a la etapa final que es la sinapsis. De acuerdo con [27] una sinapsis es una interconexión entre dos neuronas.

A continuación se definirán las dos tipos de sinapsis que existen, la sinapsis química y la sinapsis eléctrica [29].

- **Sinapsis eléctrica** este tipo de sinapsis utiliza uniones gap (conexiones proteicas entre células adheridas estrechamente) para transportar iones de una neurona a otra, este tipo de sinapsis es bidireccional y es más rápida que la sinapsis química.
- **Sinapsis química** este tipo de sinapsis se efectúa entre neuronas que tiene 20 o 30 nanómetros de distancia y se da mediante la transmisión y recepción de neurotransmisores, fruto de una secreción celular muy veloz, este tipo de sinapsis es unidireccional. En la imagen 2.3 se puede observar una representación gráfica de este proceso sináptico.

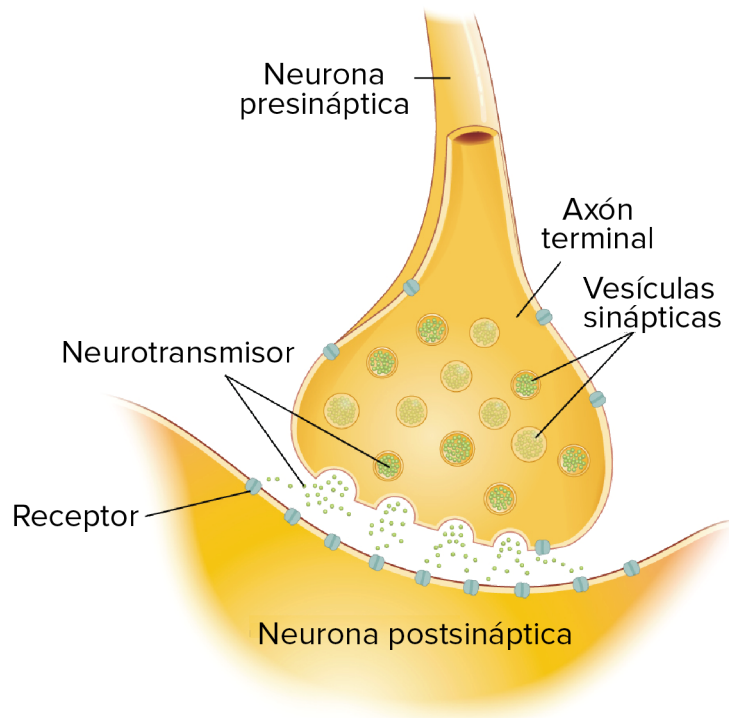


Figura 2.3: Imagen de una sinapsis química. Tomada de [30]

De acuerdo con [27] los comportamientos importantes para nosotros que existen en una neurona son los siguientes:

- El pulso que llega a una sinapsis y el que sale de ella no son iguales en general. El tipo de pulso de salida dependerá sensiblemente de la cantidad de neurotransmisor. Esta cantidad cambia durante el proceso de aprendizaje, es aquí donde se almacena la información. Una sinapsis modifica el pulso, ya sea reforzándolo o debilitándolo
- En el cuerpo celular se suman las entradas de todas las dendritas. Si estas entradas sobrepasan un cierto umbral, entonces se transmitirá un pulso a lo largo del axón, en caso contrario no transmitirá. Después de transmitir un impulso, la neurona no puede transmitir durante un tiempo de entre 0.5 ms (milisegundos) a 2 ms. A este tiempo se le llama periodo refractario

2.4.2. Modelo neuronal artificial

Los investigadores McCulloch et al [31] crearon un modelo que es abstracto y simple de una neurona artificial. En la figura 2.4 se puede observar el modelo propuesto.

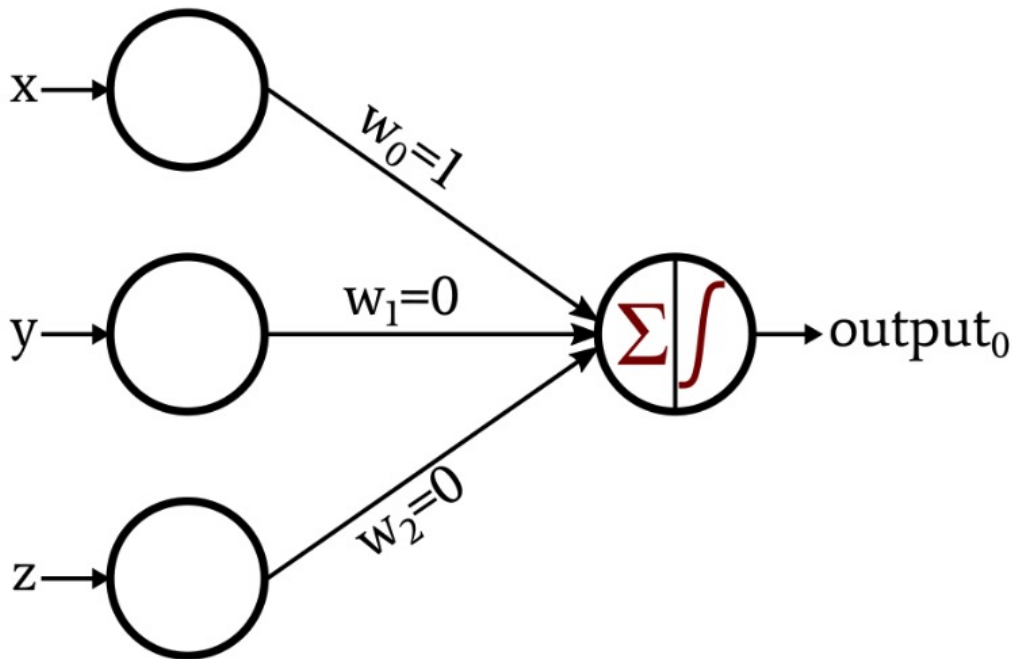


Figura 2.4: Imagen del modelo de una neurona artificial. Tomada de [32]

Este modelo propuesto de una neurona artificial, también llamado perceptrón, tiene características similares a las de una neurona biológica. Como se explicó anteriormente, una de las características importantes es la suma de todas las entradas de las dendritas en el cuerpo celular.

- Un arreglo de entradas x_j con $j = 1, \dots, n$.
- Un arreglo de pesos sinápticos w_{ij} con $i = 1, \dots, n$ en donde i representa el número de la neurona a donde se dirige la información y j representa el número de la neurona de donde proviene la información.
- También existe una regla de propagación llamada h que es definida a partir del conjunto de entradas y los pesos sinápticos. La forma que más se utiliza es una combinación lineal de entradas y pesos sinápticos, así como se muestra en la ecuación 2.1.

$$h_i(x_1, \dots, x_n, w_{i1}, \dots, w_{in}) = \sum_{j=1}^n w_{ij}x_j \quad (2.1)$$

Por lo general existe un parámetro que es añadido al conjunto de pesos sinápticos y se representa con θ , este nuevo peso es considerado un umbral de activación y se acostumbra a restar al potencial post-sináptico así como se muestra en la ecuación 2.2.

$$h_i(x_1, \dots, x_n, w_{i1}, \dots, w_{in}) = \sum_{j=1}^n w_{ij}x_j - \theta_i \quad (2.2)$$

Si los índices i y j comienzan en cero y denotamos $w_{i0} = \theta_i$ y $x_0 = -1$, podemos expresar la regla de propagación como en la ecuación 2.3.

$$h_i(x_1, \dots, x_n, w_{i1}, \dots, w_{in}) = \sum_{j=0}^n w_{ij}x_j = \sum_{i=1}^n w_{ij}x_j - \theta_i \quad (2.3)$$

- Una función de activación que representa la salida de la neurona artificial y su estado de activación.

$$y_i = f_i(h_i) = f_i\left(\sum_{j=0}^n w_{ij}x_j\right) \quad (2.4)$$

Una función de activación que es común en la activación de una neurona es la función escalonada, en este caso se tiene:

$$y_i = \begin{cases} 1 & \text{para } \sum_{j=0}^n w_{ij}x_j \geq 0, \\ 0 & \text{para } \sum_{j=0}^n w_{ij}x_j < 0 \end{cases} \quad (2.5)$$

Esta función de activación es utilizada por el perceptrón original así como en el modelo pionero que fue propuesto por McCulloch et al en [31].

2.5. Redes neuronales convolucionales

Las redes neuronales convolucionales (CNN) han tenido importantes resultados en los últimos años en tareas de reconocimiento de patrones, algunos de esos problemas son el reconocimiento, y clasificación de imágenes y reconocimiento de voz. Una de las características más importantes es la reducción

del número de pesos de las tradicionales redes neuronales artificiales (ANN), aunque esta característica parece ser el aspecto más importante, la detección de características abstractas es otro aspecto que de igual manera es importante, dicho de otra manera, puede ser que en la tarea de clasificación de imágenes, en la primera capa de una CNN tal vez pueda identificar bordes o líneas en las imágenes y en capas posteriores puede reconocer formas más específicas como cuadrados, círculos, rectángulos, etc.

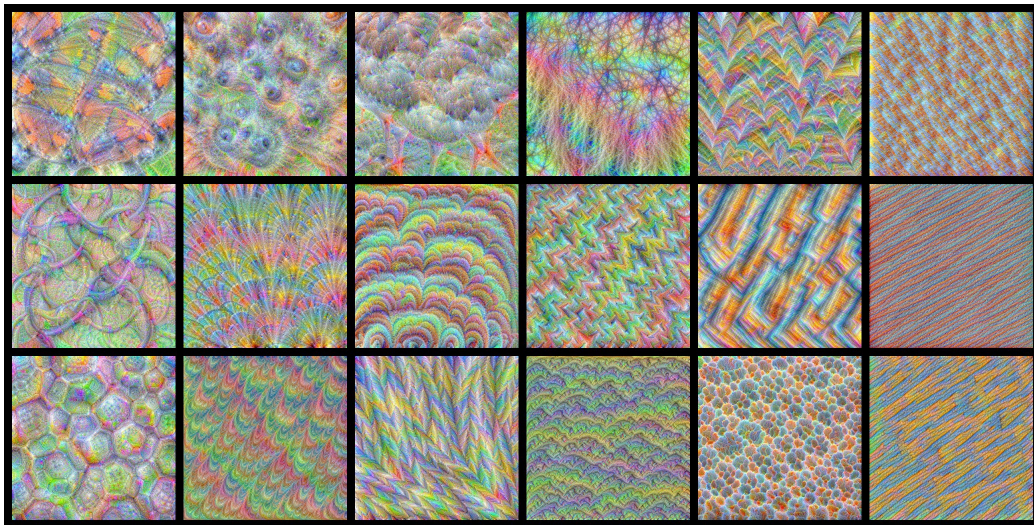


Figura 2.5: Visualización de características de una CNN. Tomada de [33].

De forma general, una convolución es una operación que involucra dos funciones que tienen como argumento valores en tiempo real, por ejemplo, si nosotros queremos saber la posición de un objeto en el espacio, nosotros podemos utilizar un sensor láser, sabemos que la posición del objeto está definida por la ecuación $x(t)$ donde t es el tiempo y la ecuación x retorna la posición del objeto. Sin embargo es posible que el sensor tenga algún tipo de ruido que afecte la estimación de la posición del objeto en el espacio, así que podemos utilizar más de una medida para obtener los valores de la posición más aproximados. Nosotros podemos sacar el promedio de varios datos dados por el sensor (las lecturas del sensor más recientes son más importantes), de esta manera, se tendrá un promedio ponderado que dé más peso a las medidas más recientes, así que tenemos una ecuación ponderada $w(a)$ en donde a es la época¹ del valor de una medida. Si nosotros aplicamos esta ecuación en todo momento, podemos obtener una nueva ecuación s que regresará un valor más

¹época: Cada ciclo de corrección de propagación para reducir la pérdida.

aproximado de la posición del objeto, la operación 2.6 es llamada convolución. [34].

$$s(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a). \quad (2.6)$$

En una red neuronal convolucional, el primer argumento son los datos de entrada y el segundo argumento es la matriz de convolución.

Una CNN puede tener como datos de entrada imágenes de tipo RGB (Red, Blue, Green) o un vídeo en escala de grises en donde cada *frame* representa la profundidad, así como se observa en la figura 2.6. Supongamos que la entrada es una imagen de tipo RGB cuadrada de 50 píxeles, por lo tanto tendríamos una imagen de 50x50x3.

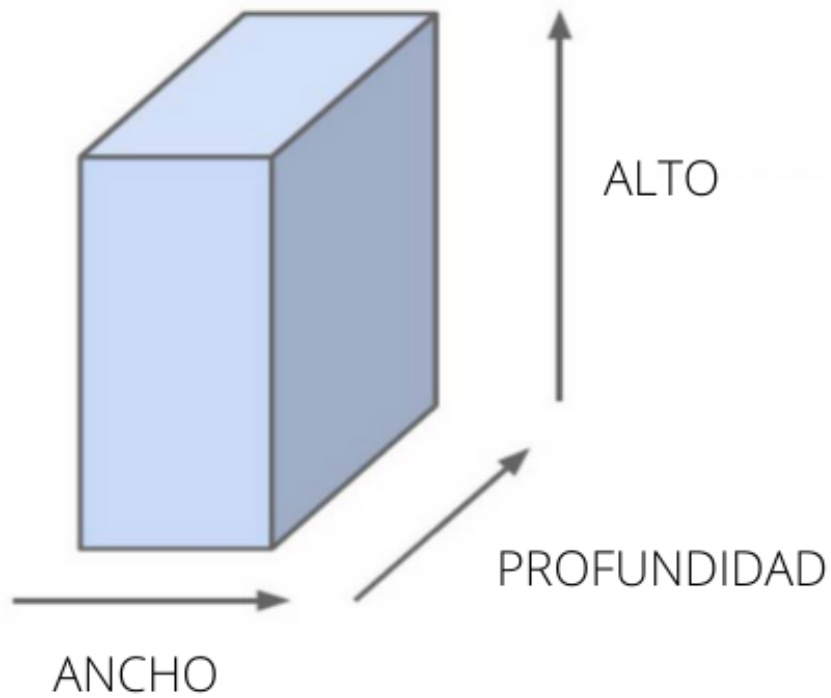


Figura 2.6: Imagen de una capa de una red convolucional. Imagen adaptada de [35]

Como se menciona en párrafos anteriores, una de las ventajas de una CNN es que el número de pesos de la red es menor a los pesos de una red neuronal completamente conectada. Por ejemplo si tenemos como datos de entrada una imagen en RGB cuadrada de 32 píxeles y la conectamos a una sola neurona (perceptrón), el número de conexiones de pesos sería de $32 \times 32 \times 3$, ahora, si decidimos agregar otra neurona a nuestra red, el número total de conexiones de la red con solo dos neuronas sería de $32 \times 32 \times 3 \times 2$ ya que cada neurona que agreguemos tendrá $32 \times 32 \times 3$ conexiones. Aunque en este ejemplo el número total de conexiones en una red neuronal completamente conectada con solo dos neuronas y una sola capa parece no ser tan alto, solo con dos neuronas no

es suficiente para hacer una correcta clasificación de imágenes, por lo tanto, se deberían de agregar más neuronas y más capas, con esto, el número de pesos se elevaría a cantidades muy altas.

Lo que hace una CNN es solo visualizar una pequeña región de una imagen en lugar de ver la imagen completa, como se puede observar en la figura 2.7, de esta manera el número de pesos disminuye considerablemente, ya que las neuronas de la siguiente capa solo reciben la la región local de la capa anterior. Para que quede más claro explicaremos esto con un ejemplo. Supongamos que tenemos como entrada una imagen cuadrada en RGB de 32 píxeles, si nosotros utilizamos una red neuronal completamente conectada con una capa con 32×32 neuronas, el número de conexiones de pesos sería de $32 \times 32 \times 3 \times 32 \times 32$ (3,145,728). Ahora bien, con una convolución el número de pesos se reduce drásticamente, si ahora, en lugar de pasar toda la imagen solo pasamos una región local de 5 píxeles el número de conexiones serían $5 \times 5 \times 3 \times 32 \times 32$ (76,800) como se puede notar, el número de conexiones es menor.

Otra suposición para simplificar el número de conexiones es mantener los pesos de las conexiones locales fijas para todas las neuronas de la siguiente capa. Esto conectaría las neuronas vecinas en la siguiente capa con exactamente el mismo peso de la región local en la capa anterior. Esto reduce el número de pesos a solo $5 \times 5 \times 3$ (75) para conectar $32 \times 32 \times 3$ neuronas a 32×32 en la siguiente capa [35].

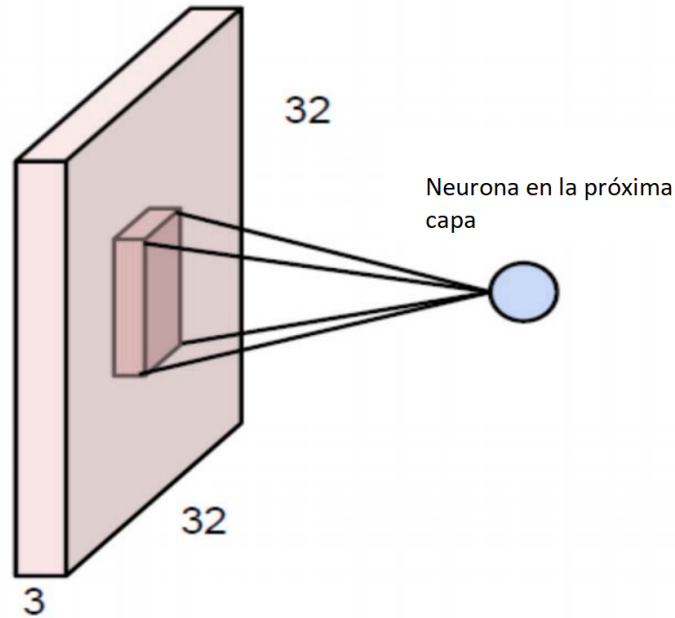


Figura 2.7: Convolución. Imagen adaptada de [35]

En la figura 2.8 se muestra una arquitectura de red neuronal convolucional propuesta por Yann LeCun en su artículo llamado *Gradient Based Learning Applied to Document Recognition* [36].

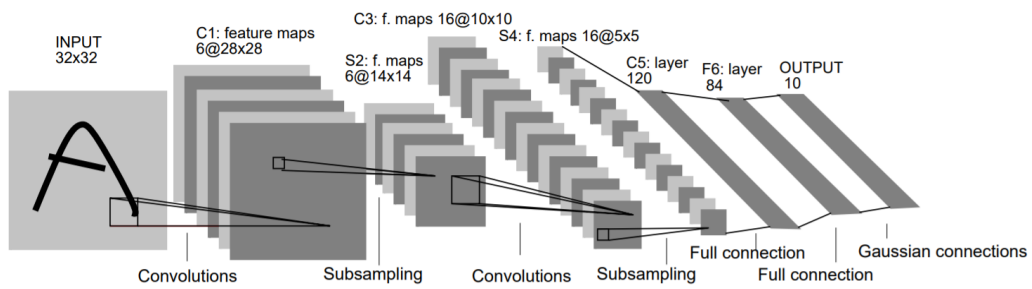


Figura 2.8: Arquitectura de CNN llamada LeNet, imagen tomada de [36]

Capítulo 3

Estado del arte

Debido a que en los últimos años la enfermedad de la depresión se ha tomado con más seriedad, no solo investigadores del área de psicología han decidido realizar trabajos de investigación para detectar depresión sino que también investigadores de otras áreas han presentado artículos en donde presentan técnicas basadas en tecnología como lo es la inteligencia artificial para hacer dicha detección. Durante este capítulo se explicaran algunos de los trabajos que se han realizado sobre este tema en años anteriores.

En el artículo titulado *Extraction of Facial Features for Depression Detection among Students* [3], los autores proponen una técnica de detección de depresión que está basada en técnicas de inteligencia artificial. Este sistema esta diseñado para detectar depresión en estudiantes ya que debido al estrés de la escuela y al cambio de vida que tuvieron al ingresar al nivel superior, los estudiantes tienen una mayor tendencia de padecer de esta enfermedad. Los autores mencionan que este sistema fue entrenado con expresiones faciales de felicidad, desprecio y disgusto.

Para realizar los experimentos correspondientes, los alumnos fueron grabados mientras contestaban un cuestionario para detectar depresión y de la misma manera, fueron grabados mientras contestaban un cuestionario para detectar un nivel de estrés. Estos vídeos fueron convertidos a imágenes para obtener las expresiones faciales de los alumnos, después, las imágenes obtenidas fueron normalizadas para una detección más efectiva. Se hizo uso de un algoritmo de máquinas de soporte vectorial para realizar la clasificación de emociones.

Para realizar el entrenamiento del algoritmo, los autores utilizaron un conjunto de datos que contenía 30 imágenes de cada expresión facial(felicidad, disgusto y deprecio). Algunas de las imágenes fueron obtenidas del conjun-

to de datos llamado JAFFE¹ (Japanese Female Facial Expression) y algunas otras imágenes fueron obtenidas de internet. Se hizo uso del algoritmo Viola-Jones para detectar en que parte de la imagen se encuentra un rostro. Después, a las imágenes se le aplicaron un total de 40 filtros de Gabor. Haciendo uso de operaciones matemáticas, las imágenes resultantes fueron transformadas a vectores para después ingresar los datos al algoritmo.

Por último, los autores llegaron a la conclusión de que si en la mayoría de las imágenes un estudiante no muestra expresiones de felicidad y además muestra expresiones de disgusto o desprecio, el estudiante puede ser clasificado como una persona con depresión severa. Por otro lado, si un estudiante presenta expresiones moderadas de felicidad y expresiones fuertes de disgusto y desprecio, el estudiante puede ser clasificado como una persona con depresión leve, en caso de que un estudiante muestre expresiones altas de felicidad y las expresiones de disgusto y desprecio sean bajas, el estudiante es clasificado como una persona sin depresión.

En otro artículo titulado *IRIT at e-Risk 2018* [37] el cual fue publicado en el año 2018, se presenta otra técnica de detección de depresión que al igual que en el artículo mencionado anteriormente, se hace uso de técnicas de inteligencia artificial. Para esta técnica en lugar de utilizar imágenes de expresiones faciales, los autores decidieron utilizar textos que fueron recuperados del sitio web llamado *Reddit*². Para realizar el conjunto de datos correspondiente, los investigadores clasificaron a los usuarios en dos categorías: depresivo y no depresivo, un usuario era clasificado como depresivo si en algunas de sus publicaciones o comentarios mencionaba explícitamente que había sido diagnosticado con depresión, por otro lado, se clasifica a una persona como no depresiva si en sus publicaciones o comentarios no menciona ninguna palabra relacionada con depresión. El conjunto de datos de entrenamiento contiene un total de 887 usuarios, de los cuales 135 son clasificados como depresivos y 752 como no depresivos. Por otro lado, el conjunto de datos que se utilizó en la etapa de pruebas contiene un total de 820 usuarios, de los cuales 79 son clasificados como depresivos y 741 como no depresivos. Además, en la etapa de entrenamiento se usó otro conjunto de datos adicional llamado *Clpsych 2016 dataset*³.

En este trabajo, los autores mencionan que desarrollaron dos modelos diferentes para realizar la detección de depresión. El primer modelo estuvo basado en características. Este tipo de modelo requiere de ingeniería de características que se basó en un conjunto de estadísticas o características

¹<https://zenodo.org/record/3451524>.

²<https://www.reddit.com/>

³<https://dblp.org/db/conf/naacl/clpsych2016.html>

lingüísticas. Para cada usuario las características fueron calculadas de la siguiente manera. Se calcula el valor de la característica para cada una de las publicaciones o comentarios, después, se promedia el valor sobre sus escritos. Cuando varios fragmentos son utilizados por un mismo usuario, los valores obtenidos son promediados, este modelo fue entrenado con el algoritmo de Bosques Aleatorios. Por otra parte, en el segundo modelo se hizo uso de una técnica llamada vectorización de texto, en donde cada texto obtenido fue transformado a un vector, este modelo fue entrenado con un clasificador de regresión logística. Los resultados de ambos modelos fueron medidos con las métricas de *Early risk detection error metric (ERDE)*, *F1-Score*, *Precision* y *Recall*. El modelo basado en características estuvo por encima de los demás modelos.

Así como los autores anteriores han utilizado conjuntos de datos basados en imágenes y texto, en el artículo publicado en el año 2018 llamado *Detecting depression using an ensemble logistic regression model based on multiple speech features* [38] han utilizado un conjunto de datos que están basados en características vocales. Se menciona que algunos expertos en el área de psicología, psiquiatría y neurociencias han descubierto una correlación entre las características de la voz y la depresión, no obstante, algunos otros expertos de las mismas áreas han encontrado que no existe ninguna correlación importante. Los autores mencionan que todos los participantes de este experimento son hablantes chinos con una edad entre 18 y 55 años. Todos los participantes pasaron por 3 etapas, la primera etapa consistió en llenar una hoja de evaluación en donde se le solicitaron datos como la edad, el historial de salud, género, estado educativo y de empleo. En la segunda etapa los participantes fueron escogidos por psiquiatras basándose en el *Diagnostic and Statistical Manual Of Mental Disorders (DSM-IV)*. Por último, todos los participantes fueron entrevistados por psiquiatras para realizar un diagnóstico depresivo utilizando el cuestionario PHQ-9 el cual se evalúa en una escala del 0 a 27. A continuación se detalla la evaluación de dicho cuestionario.

- **0-4:** mínima existencia o ausencia de síntomas depresivos.
- **5-9:** síntomas depresivos leves.
- **10-14:** síntomas depresivos moderados.
- **15-19:** síntomas depresivos moderados a graves.
- **20-27:** síntomas depresivos graves.

De acuerdo con los resultados del cuestionario, los participantes se dividieron en dos grupos: depresivos y no depresivos. Si el resultado de una

persona es mayor a 5 era etiquetado como depresivos y en caso contrario como no depresivos.

La fase de experimentación en donde los autores recopilaron las características vocales de las personas consistió en 3 etapas. En la primera etapa, los partícipes fueron entrevistados con preguntas tomadas del DSM-IV. En la segunda etapa, los partícipes leyeron una historia y además, leyeron 3 conjuntos de palabras neutras, felices y negativas. En última etapa los participantes detallaron imágenes que los autores les mostraron. Todas las voces se grabaron en un lugar silencioso y a prueba de ruido, los autores mencionan que el nivel de ruido en el laboratorio se mantuvo por debajo de 80 decibeles. En este experimento participaron un total de 170 personas, 85 personas con un estado no depresivo (34 hombres y 51 mujeres). Y 85 personas diagnosticadas con depresión (32 hombres y 53 mujeres). En este artículo se utilizaron Máquinas de Soporte Vectorial (SVM), Regresión Logística (LR) y un Modelo de Mezcla Gaussiana (GMM).

La depresión puede ser causada por pensamientos de desesperanza o depresión endógena, la depresión por desesperanza es un tipo de depresión el cual es producto de la combinación de un estilo de atribución negativo con un acontecimiento vital negativo. En cambio, la depresión endógena ha sido definida como un estado de depresión en el que la causa primaria precipitante reside en los factores biológicos o somáticos, en contraste a los factores situacionales que acontecen en el ambiente de las personas. Por ello, algunos investigadores han querido descubrir que es lo que sucede en el cerebro y como influyen los neurotransmisores cuando una persona tiene depresión y han propuesto un método para realizar una predicción de depresión por medio de resonancias magnéticas. En el artículo publicado en el año 2013 titulado *Towards automated detection of depression from brain structural magnetic resonance images* [39]. Una resonancia magnética es un sistema que es altamente utilizado en áreas de investigación y en prácticas clínicas. Las imágenes cerebrales tienen un gran potencial para diagnosticar y tratar enfermedades psiquiátricas y es por ello que algunos investigadores interesados en descubrir más sobre el tema han demostrado que la depresión puede estar asociada con algún daño en regiones del cerebro.

Según lo mencionan Kuryati Kipli et al. En la mayoría de trabajos que se han realizado sobre este tema, se han utilizado conjuntos de datos privados y que fueron construidos con resonancias de pacientes de hospitales y clínicas. Los únicos conjuntos de datos públicos que existen son ELUDE⁴ que contiene 281 resonancias de personas con depresión y 154 controles, y MIRIAD⁵ que

⁴<https://www.nitrc.org/projects/elude/>

⁵<https://www.nitrc.org/projects/miriad/>

contiene 50 resonancias de personas con depresión y 50 controles. Todas las imágenes obtenidas pasan por un preprocesamiento para mejorar la calidad de las imágenes y algunos de los procesos que son usados comúnmente son realinear una imagen, normalizar el tamaño de las imágenes, estandarizar la orientación y la posición. Estas imágenes también son segmentadas y constan de dos etapas: la extracción de tejido cerebral y aislamiento de zonas de interés del cerebro. La primera etapa consiste en remover el tejido no cerebral como lo es el cráneo. La etapa de aislamiento de zonas de interés consta de segmentar la región del cerebro que puede ser usada para el proceso de la extracción de características. Para realizar la clasificación, los autores mencionan que se han utilizado diferentes algoritmos de aprendizaje automático en diferentes estudios, se han utilizado SVM y también se ha utilizado el clasificador K vecinos más cercanos para la remisión del tratamiento en pacientes con depresión mayor. Algunos otros investigadores han propuesto un método de clasificación basado en probabilidad general para predecir depresión.

Después explicar algunos de los trabajos de investigación relacionados con IA y con depresión podemos decir que en los trabajos revisados aun no se han hecho experimentos con CNN, expresiones faciales y depresión, así que en este trabajo de investigación se expondrá si una CNN podría ser opción para una predicción de dicha enfermedad y en base a los resultados que se obtengan, quizá se pueda desarrollar un sistema que pueda ser accesible para cualquier persona con acceso a internet.

Capítulo 4

Metodología

Como se ha comentado en capítulos anteriores, el objetivo general de esta tesis es proponer un sistema basado en alguna técnica de inteligencia artificial que sea capaz de realizar una detección de depresión usando solo las expresiones faciales de una persona.

Para cumplir con el objetivo general, se plantea desarrollar una aplicación de escritorio la cual se desarrollara en el lenguaje Python¹ y con la biblioteca Tkinter[40], en esta interfaz gráfica el usuario podrá tomarse una fotografía para que la red neuronal pueda procesarla y realizar el diagnóstico correspondiente. Para el desarrollo de la inteligencia artificial se hará uso de un técnica de aprendizaje profundo específicamente con una red neuronal convolucional ya que una de las características más importantes de esta red neuronal es el análisis de imágenes. Esta red neuronal sera desarrollada con el lenguaje Python y el framework de código libre Tensorflow[41]. Una red neuronal convolucional solo analizará la imagen correspondiente, sin embargo es necesario usar una arquitectura de inteligencia artificial especializada en la clasificación, para este proyecto se utilizara una red neuronal completamente conectada.

4.1. Definición del problema

La pandemia de la enfermedad COVID-19 [42] nos ha obligado a permanecer en nuestras casas interrumpiendo nuestras actividades cotidianas y como consecuencia la salud mental de la mayoría de las personas se ha visto afectada. En un estudio realizado en los meses de Enero y Febrero del año 2020[43], en el que participaron 1210 personas de 194 ciudades de China, se encontró que el 54 % sufrió un impacto psicológico moderado o severo, el 29 %

¹<http://www.python.org>

tuvo síntomas moderados de ansiedad, mientras que el 17% tuvo síntomas moderados de depresión. Esta disrupción nos ha obligado a crear e innovar herramientas digitales para realizar actividades que antes eran presenciales por ejemplo las clases escolares, conciertos musicales, etc. Esto nos indica que se deben crear más herramientas tecnológicas para que las personas puedan obtener un diagnóstico rápido de depresión sin tener que salir de sus hogares.

4.2. Análisis de factibilidad

Los análisis de factibilidad y viabilidad son herramientas que nos ayudan a determinar si un proyecto puede ser realizado o no. Éstos análisis deben ser realizados antes de planificar el proyecto ya que se puede dar el caso en el que después de semanas de trabajo de planificación se llegue a conclusión de que el proyecto que se ha empezado es factible pero no es viable, o en el peor de los casos, el proyecto no es factible ni viable. Un análisis de factibilidad evalúa la capacidad económica y operativa que se tiene para el desarrollo del proyecto, el aprovechamiento del talento humano o los recursos estratégicos y aspectos legales que podrían generar riesgos en la implementación del proyecto. Para este proyecto se analizarán 3 tipos de factibilidad.

- Factibilidad Operativa.
- Factibilidad Técnica.
- Factibilidad Económica.

4.2.1. Factibilidad Operativa

Este proyecto es factible operativamente debido a que el sistema que se desarrollará tendrá un diseño amigable e intuitivo y podrá ser usado por cualquier persona que tenga conocimientos básicos en computación. Aunque este sistema esta pensado para usuarios comunes, este sistema es escalable y puede ser utilizado masivamente en las empresas o en las escuelas para determinar el estado de la salud mental de los asistentes, también puede ser utilizado por clínicas psicológicas para un diagnóstico rápido.

4.2.2. Factibilidad Técnica

Este proyecto es factible técnicamente debido a que se poseen los conocimientos en ingeniería de software, en el lenguaje Python el cual es necesario

para el desarrollo del software y el desarrollo de la arquitectura de inteligencia artificial. Todas estas herramientas y conocimientos son fundamentales para realizar las funcionalidades necesarias del sistema que se presentan en la sección 4.3. El programa es sencillo y fácil de entender para los usuarios por lo que no habrá ningún inconveniente en la utilización del sistema. Para los programadores que se unan en el desarrollo del sistema, existirá documentación para comprender el modelado del software, esta documentación incluirá diagramas de casos de uso, diagramas de clases y diagramas de interacción así como el diagrama de la arquitectura de inteligencia artificial.

4.2.3. Factibilidad Económica

Como se menciona en párrafos anteriores, para el desarrollo del sistema solo se ocuparán herramientas, lenguajes y software de código libre por lo que no es necesario una inversión inicial para la compra de software o licencias. El sistema se ejecutará localmente por lo que no se deberá rentar un dominio ni un servicio de alojamiento para la aplicación, solo será necesario que el usuario lo instale en su computadora personal. En caso de que el número de usuarios aumente, será necesario desarrollar más funcionalidades, migrar a la nube y por lo tanto, se ocuparán más recursos humanos, un dominio y un servicio de alojamiento. En el cuadro 4.1 se definen cuáles serían los costos de los recursos técnicos y recursos humanos para una aplicación local.

RUBRO	COSTO MENSUAL	CANTIDAD	TOTAL
Salario de programador	\$11,000	2	\$22,000
Servicio de luz	\$350	1	\$350
Servicio de Internet	\$500	1	\$500
Total mensual			\$22,850

Cuadro 4.1: Costo del sistema versión local.

En el cuadro 4.2 se definen cuáles serían los costos de los recursos técnicos y recursos humanos en caso de que la aplicación local se quisiera migrar a una aplicación web.

RUBRO	COSTO MENSUAL	CANTIDAD	TOTAL
Salario de programador	\$11,000	3	\$33,000
Renta de dominio	\$200	1	\$200
Renta de alojamiento	\$400	1	\$400
Servicio de luz	\$350	1	\$350
Servicio de Internet	\$500	1	\$500
Total mensual			\$34,350

Cuadro 4.2: Costo del sistema versión web.

4.3. Requerimientos del sistema

Definir de un conjunto de requisitos, ya sea funcionales o no funcionales es una tarea importante cuando se empieza el desarrollo de un proyecto. La recolección de requisitos es un proceso en el cual se definen cuales son las necesidades del usuario, además, nos proporciona una base para definir el alcance del proyecto. Los requisitos nos proporcionan una visión más clara del problema a resolver antes de empezar a diseñar la solución. No sería agradable que el sistema ya esté desarrollado y este un resuelva un problema equivocado. Como se menciona en párrafos anteriores, el principal objetivo de este sistema es la detección de depresión en una persona, para cumplir con este objetivo es necesario que se establezcan todos los requerimientos necesarios de la solución. En el cuadro 4.3 se muestran los requisitos funcionales y no funcionales del sistema que se propone. Estos requisitos servirán como guía en el proceso de diseño y modelado del sistema.

Requisitos funcionales		
Código	Prioridad	Descripción
RE1	Muy alta	Opción para subir una imagen o tomar una fotografía.
RE2	Muy alta	Tomar una fotografía con la cámara de la computadora.
RE3	Muy alta	Subir una fotografía desde la galería de imágenes.
RE4	Alta	Prevista de la cámara.
RE5	Alta	Detección de rostro.
RE6	Alta	Tratamiento de la imagen.
RE7	Alta	Realizar diagnóstico de depresión.
RE8	Alta	Mostrar los resultados en pantalla.
RE9	Media	Opción para realizar nuevamente el diagnóstico.
Requisitos no funcionales		
Código	Prioridad	Descripción
RE9	Media	El diseño de la interfaz debe ser agradable para el usuario.
RE10	Media	El sistema tiene que ser portable.
RE11	Alta	Las fotografías no se deben almacenar.
RE12	Alta	Los resultados no deben almacenarse

Cuadro 4.3: Requisitos funcionales y no funcionales del sistema.

4.4. Cronograma de actividades

Antes de empezar a construir cualquier proyecto, ya sea la creación de un nuevo automóvil eléctrico o el desarrollo de un nuevo sistema de software, es necesario organizar el tiempo que se tiene de acuerdo a la fecha de entrega y designar las tareas al personal capacitado. En la etapa de construcción de un proyecto, generalmente muchas de las tareas se desarrollan paralelamente y el resultado de una tarea ya desarrollada puede influir en el comienzo de una nueva tarea. Las relaciones entre tareas suelen ser difíciles de comprender y eso conlleva a un retraso en el proyecto, la mejor manera de evitar este riesgo es creando un calendario o cronograma de actividades. La creación de un cronograma es un proceso en el cual se analiza la secuencia de actividades, el tiempo de duración de cada actividad, requisitos de recursos y restricciones. Para calendarizar todas las actividades que se realizarán en este proyecto, se utilizará un diagrama de Gantt, en el se listan todas las actividades que son necesarias para concluir el proyecto así como los tiempos definidos para cada tarea y la relación de tareas entre ellas. En la imagen 4.1 se muestra el calendario de actividades y las fechas establecidas para la realización de cada una de las actividades que se llevaran a cabo durante este proyecto, por ejemplo, el entrenamiento de la red neuronal, el diseño del sistema, etc.

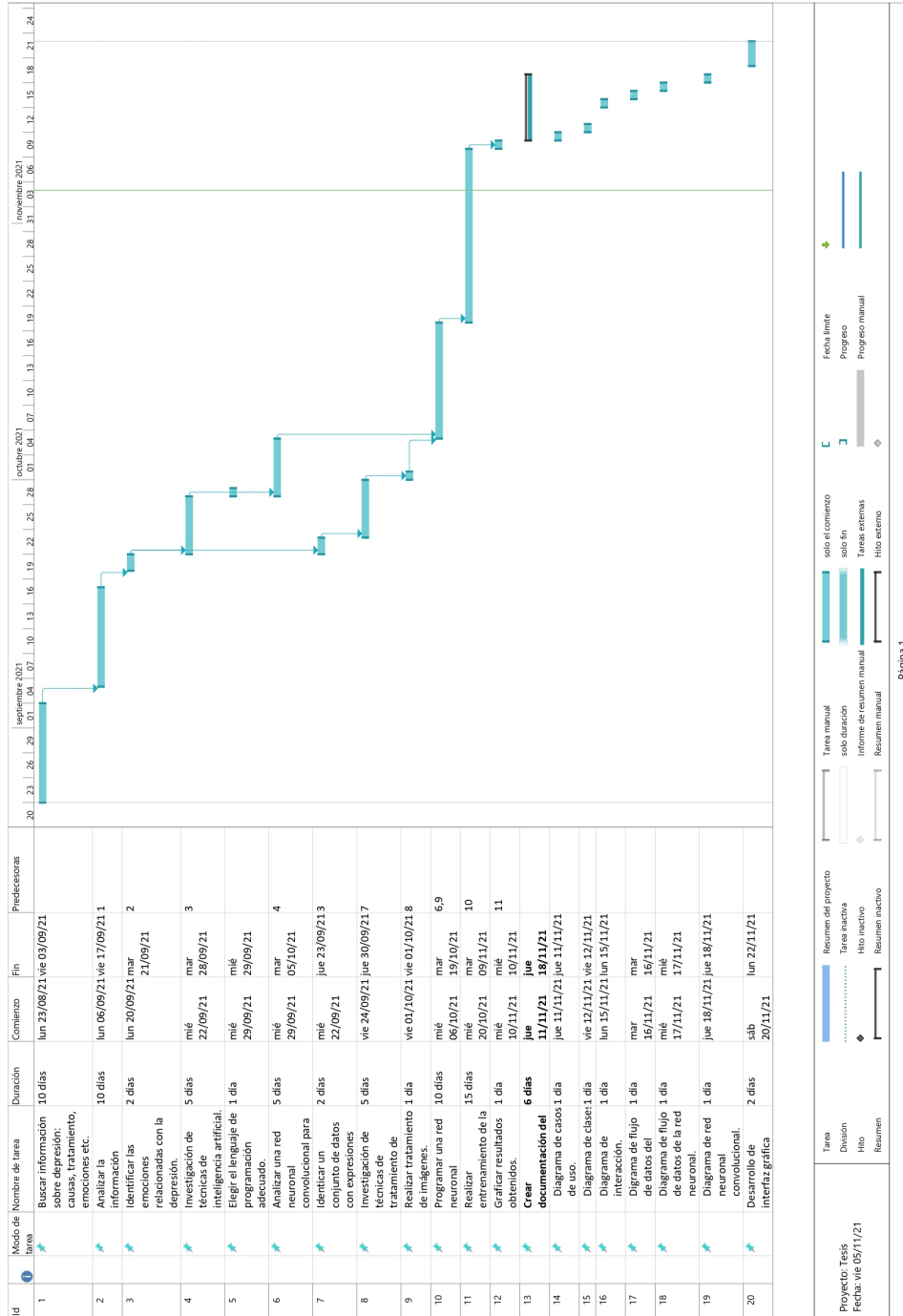


Figura 4.1: Cronograma de actividades del proyecto

4.5. Modelado de software

El modelo de software es una buena practica que debe realizarse antes de empezar a escribir montones y montones de lineas de código. Cuando se empieza a construir un edificio no podemos solo comprar algunos materiales y algunas herramientas básicas y empezar a construir, ya que lo más probable es que la construcción del edificio fracase, para que esto no suceda es necesario que expertos realicen el diseño de planos detallados en donde se especifique los tamaños de las oficinas, las vías de electricidad, tubería de drenaje, etc.

Al igual que en la arquitectura, en el desarrollo de software es algo muy parecido, el modelado de software proporciona planos detallados de un sistema, estos planos pueden ser generales o muy específicos(entre más específicos mejor).

Un modelo puede ser estructural, destacando la organización del sistema o puede ser de comportamiento, destacando su dinámica. En este capítulo se modelará el sistema de software, por un parte se utilizarán diagramas UML² para el modelado del sistema de software y por otra parte se modelará la arquitectura de la red neuronal.

Para determinar el diagnóstico de depresión de un usuario, se tomara como base el método que el Dr. Venkataraman et al propuso en [3].

- Si la probabilidad de que las expresiones faciales de un usuario representen felicidad es baja y la probabilidad de que representen emociones como tristeza, disgusto y deprecio es alta entonces el diagnóstico será **Depresión alta**.
- Si la probabilidad de que las expresiones faciales de un usuario representen felicidad es un nivel medio y la probabilidad de que representen emociones como tristeza, disgusto y deprecio es alta entonces el diagnóstico será **Depresión media**.
- Si la probabilidad de que las expresiones faciales de un usuario representen felicidad es alta y la probabilidad de que representen emociones como tristeza, disgusto y deprecio es media entonces el diagnóstico será **Depresión baja**.
- Si la probabilidad de que las expresiones faciales de un usuario representen felicidad es alta y la probabilidad de que representen emociones como tristeza, disgusto y deprecio es baja entonces el diagnóstico será **No tiene depresión**.

²Lenguaje Unificado de Modelado

4.5.1. Sistema de software

Como se ha mencionado antes, para el modelado del sistema, se utilizarán diagramas UML, ya que este lenguaje fue creado con la finalidad de definir visualmente la arquitectura de un sistema.

El diseño de la interfaz gráfica estará representado por medio de un prototipo o un boceto. Estos bocetos también son llamados *Wireframes*.

Los diagramas UML que se utilizarán para el modelado del software serán los siguientes:

- Diagrama de casos de uso.
- Diagrama de clases.
- Diagrama de secuencia.
- Diseño de interfaz gráfica.

4.5.1.1. Diagrama de casos de uso

Un diagrama de casos de uso indica un conjunto de secuencias, donde cada secuencia representa la interacción de los elementos externos del sistema. Un caso de uso representa un requisito funcional. En este diagrama se involucra la interacción entre los actores externos y el sistema.

En la figura 4.2 se pueden observar los casos de uso de nuestro sistema. En el sistema existe una interacción de dos actores, por una parte el usuario puede subir una fotografía o una imagen ya existente de su rostro y la inteligencia artificial realizará un tratamiento a la imagen, seguido de esto, la IA hará el diagnóstico de depresión y por ultimo, la IA muestra los resultados correspondientes.

Cabe aclarar que para el caso de uso de detección del rostro en una imagen, se hará uso de un clasificador que ya ha sido previamente entrenado. En este trabajo no se explicará a detalle el funcionamiento del clasificador pues excede los objetivos establecidos, este método de detección de rostros fue propuesto en [44].

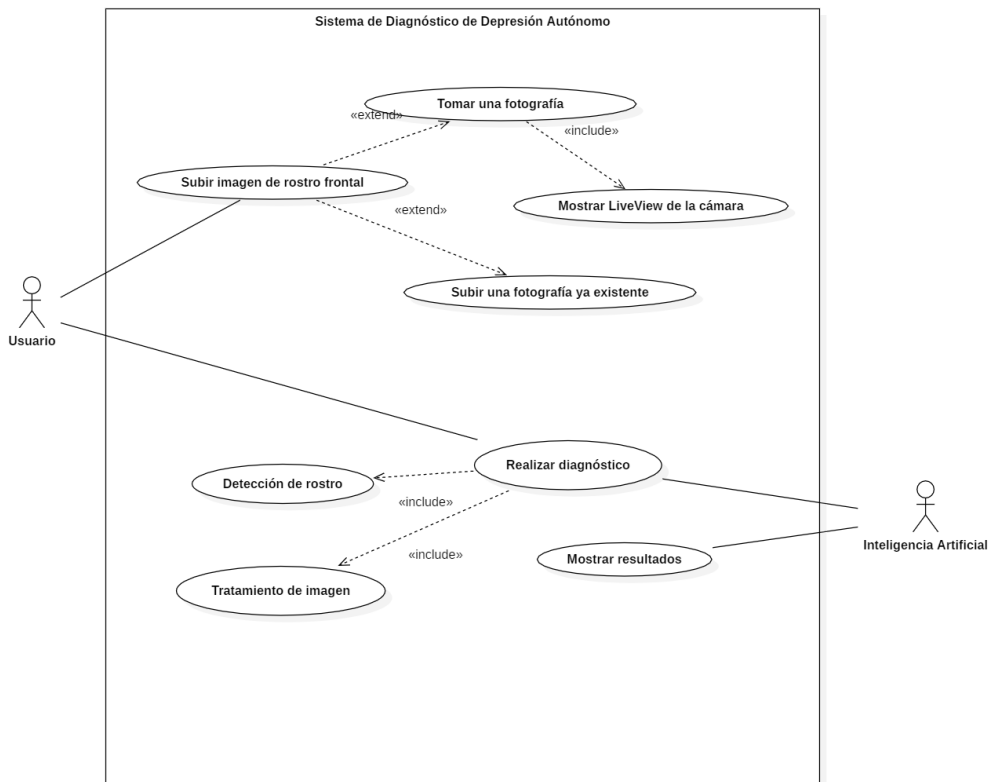


Figura 4.2: Diagrama de casos de uso del sistema.

4.5.1.2. Diagrama de clases

Las clases son los bloques más importantes al momento de realizar el modelado de cualquier sistema sin importar si es software, hardware o simplemente es algo conceptual. Una clase se puede definir como una descripción de un conjunto de objetos que comparten las mismas propiedades, por ejemplo, atributos, operaciones o relaciones.

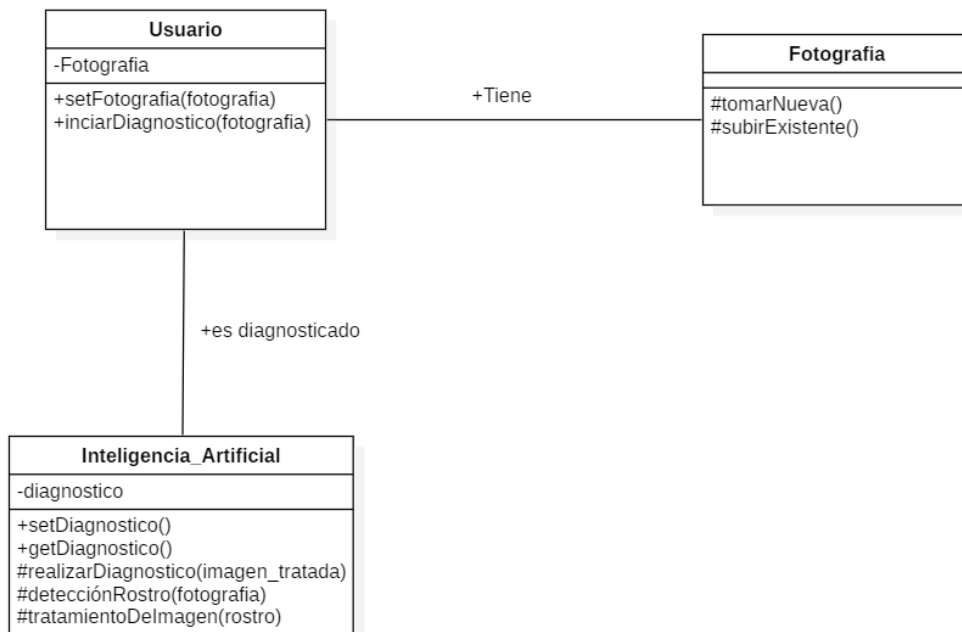


Figura 4.3: Diagrama de clases del sistema.

Para el sistema que se está desarrollando, se contemplan cinco clases diferentes que están relacionadas entre si, es decir, cada una de las clases colabora directamente con otra clase. Las clases que se han previsto en el modelado son:

- **Usuario:** La clase usuario tiene un solo atributo el cual es la fotografía del usuario. El usuario puede agregar una fotografía o iniciar con el diagnóstico.
- **Fotografia:** La clase fotografia no tiene ningún atributo, tiene dos operaciones las cuales son `tomarNueva` y `subirExistente`. Esta clase esta compuesta de dos clases especializadas, la clase cámara y la clase `archivos_existentes`.

- **Camara:** La clase camara es una clase especializada y no tiene atributos. Esta clase solo tiene una operacion la cual es tomarFotografia.
- **Archivos_Existentes:** Esta clase es una clase especializada y no tiene atributos. Esta clase solo tiene una operación la cual es subirFotografiaExistente.
- **Inteligencia_Artificial:** Esta clase es la más importante para el desarrollo del sistema que se propone, pues es en esta clase en donde se realizará el tratamiento de la imagen, la detección del rostro y el diagnóstico. Esta clase tiene un atributo llamado “diagnostico” y cinco operaciones, dos de ellas son para obtener y definir el resultado y las otras tres son para realizar la detección de rostro, tratamiento de imagen y el diagnóstico correspondiente.

En la figura 4.3 se puede observar el diagrama completo de las clases que se han definido.

4.5.1.3. Diagramas de secuencia

Los diagramas de interacción se utilizan para modelar los aspectos dinámicos que existen en un sistema además, con estos diagramas se puede visualizar y construir la dinámica de los objetos definidos o incluso puede ocuparse para modelar un flujo de control para un caso de uso. Para nuestro sistema utilizaremos los diagramas de secuencia, estos diagramas de interacción son usados para la ordenación temporal de los mensajes que se envían y se reciben entre clases. Al momento de realizar los diagramas de secuencia se tomó en cuenta que cada caso de uso que se definió anteriormente involucra diferentes clases y por lo tanto se envían y se reciben diferentes mensajes. Por este motivo, se decidió realizar un diagrama de secuencia por cada caso de uso.

En el primer diagrama que se muestra en la figura 4.4 se representa la secuencia que se tiene que seguir para que el usuario suba una fotografía de su rostro. Para este caso de uso se involucran 3 clases, la clase de usuario, la clase de Fotografia y la clase Camara o la clase Archivos_Existentes dependiendo el tipo de fotografía que el usuario decida subir.

En el segundo diagrama que se muestra en la figura 4.5 se representa la secuencia que se tiene que seguir para que la inteligencia artificial realice el diagnóstico. Para este caso de uso se involucran 2 clases, la clase de usuario, la clase Inteligencia Artificial.

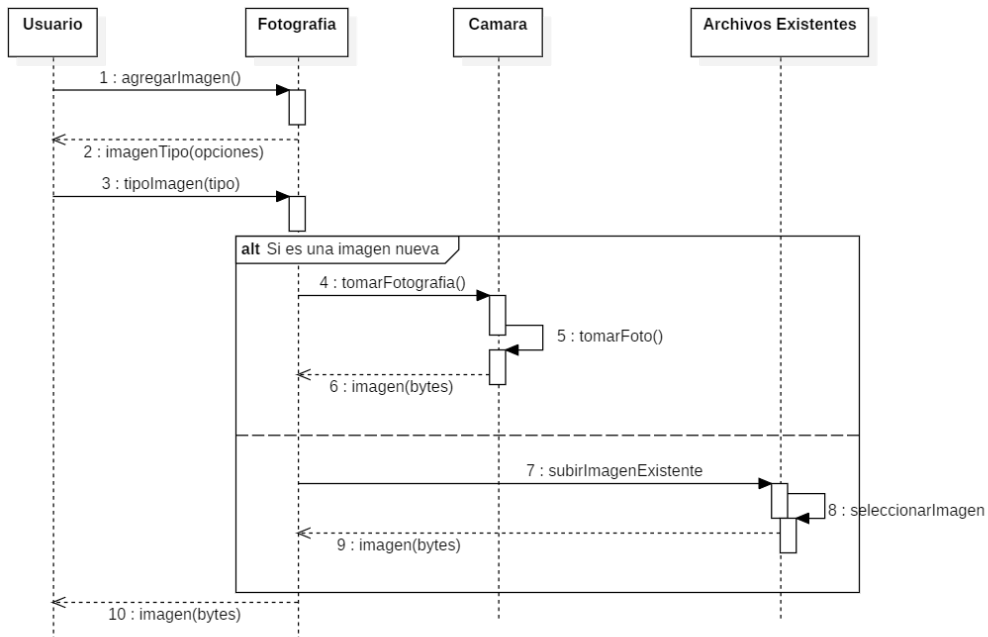


Figura 4.4: Diagrama de secuencia para el caso de uso Subir imagen.

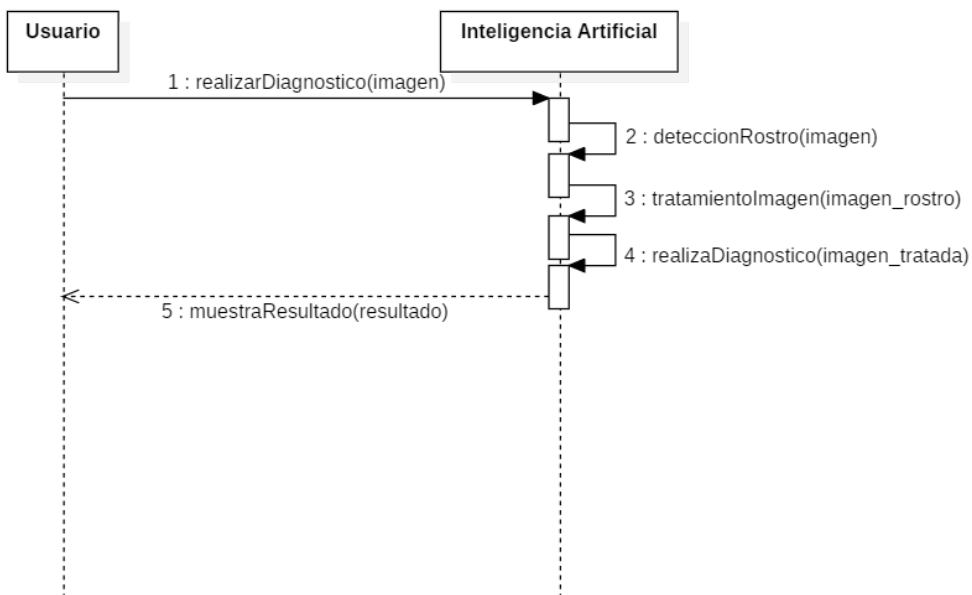


Figura 4.5: Diagrama de secuencia para el caso de uso Realizar Diagnóstico.

4.5.1.4. Diseño de interfaz gráfica

La interfaz gráfica de un sistema es una parte fundamental debido a que a través de esta se realiza la interacción humano computadora, por lo tanto, esta interfaz debe ser clara y entendible para el usuario. La interfaz gráfica del sistema que se propone, solo contiene una ventana y en esta ventana se encuentran los elementos necesarios que el usuario necesita para realizar un diagnóstico.

En la figura 4.6 se puede observar la interfaz gráfica propuesta para este sistema en particular. En la parte inferior de la ventana se encuentran dos botones que indican las opciones que el usuario tiene para realizar un diagnóstico, como se comentó en los diagramas anteriores, el usuario puede subir una fotografía existente o tomar una nueva. También se encuentra un recuadro en donde se muestra la prevista de la cámara o la imagen existente específica que el usuario decida subir. Por último, se encuentra el apartado de resultados en donde se muestran las probabilidades que de que las expresiones faciales del usuario represente algunas de las emociones relacionadas con la depresión y un diagnóstico de depresión leve, alta o severa.

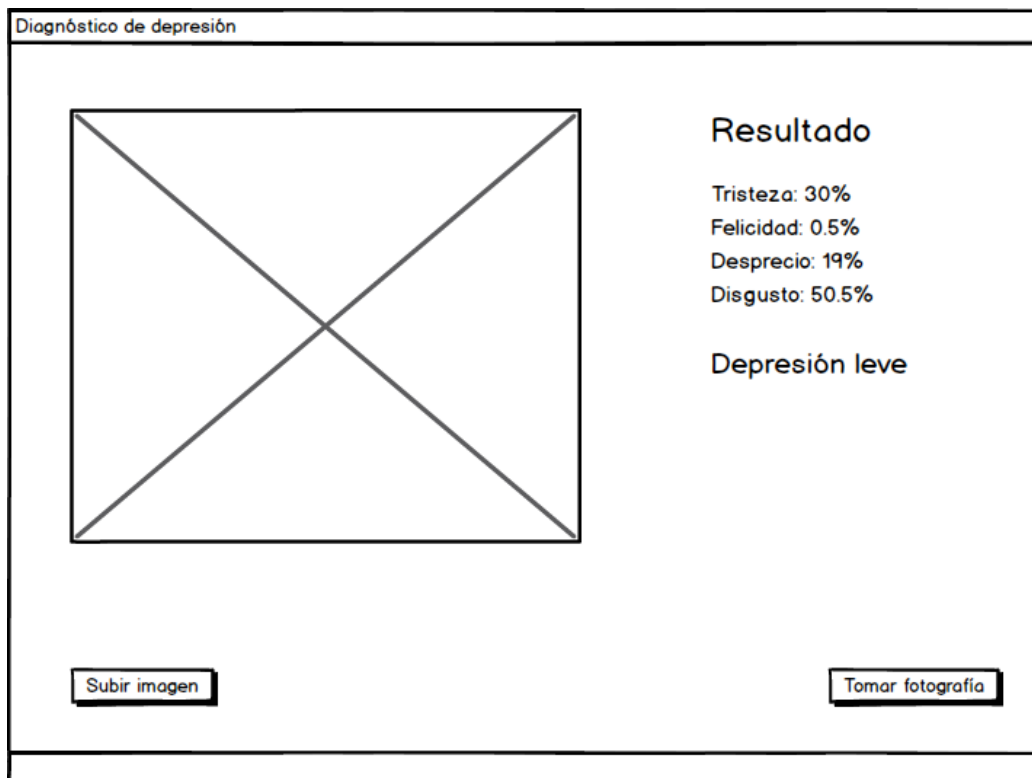


Figura 4.6: Wireframe de la interfaz gráfica del sistema.

4.5.2. Red neuronal

Como se ha mencionado en el capítulo uno, el objetivo general de este trabajo es proponer una arquitectura de inteligencia artificial que tenga la capacidad de realizar un diagnóstico de depresión usando solo las expresiones faciales de una persona. En esta sección se describirá la arquitectura propuesta así como el tratamiento de imágenes que se realizó, el conjunto de datos que se utilizó para el entrenamiento y las pruebas de la red neuronal.

4.5.2.1. Diseño de arquitectura

Para realizar un diagnóstico de depresión por medio de inteligencia artificial se han utilizado diferentes técnicas y diferentes conjuntos de datos, en algunos de los métodos se han utilizado conjunto de datos de audio, texto, e imágenes. En el capítulo tres se han mencionado algunos de los trabajos que se han realizado en años anteriores. La arquitectura de inteligencia artificial que se propone para este proyecto esta basada en redes neuronales convolucionales. El propósito de la arquitectura propuesta es realizar una clasificación de emociones por medio de imágenes que contienen rostros de los usuarios.

En el capítulo dos, específicamente en la sección cinco se ejemplifica una arquitectura de una red neuronal convolucional que fue propuesta por primera vez por Yann LeCun et al en [36] (ver figura 4.7). Esta arquitectura se tomó como base para el desarrollo de la arquitectura que se observa en la figura 4.8 a la cual se nombró como DEPRECNN.

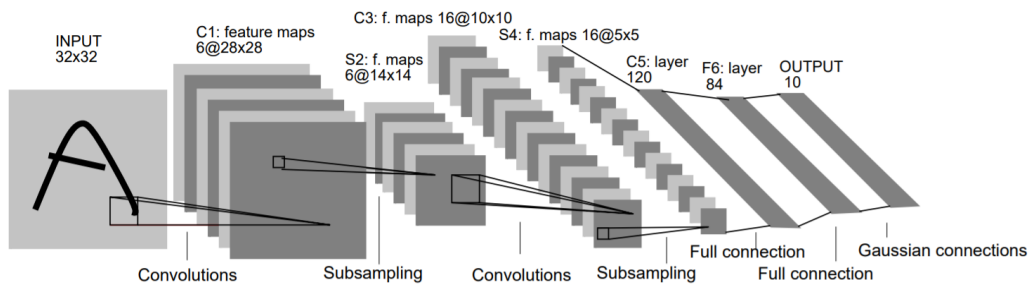


Figura 4.7: Arquitectura base, imagen tomada de [36]

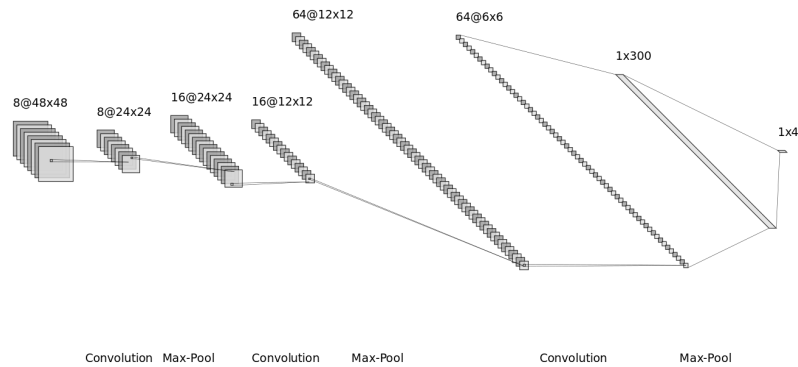


Figura 4.8: Arquitectura propuesta.

De la arquitectura base solo se tomó la estructura general, es decir, las capas convolucionales y las capas totalmente conectadas, sin embargo se hicieron algunas modificaciones que se listan a continuación.

- Agregación de más capas convolucionales.
- Agregación de más canales.
- Agregación de más capas max pooling.
- Disminución de capas totalmente conectadas.

La red neuronal DEPRECNN esta compuesta por 3 capas convoluciones y 3 capas de max pooling para la extracción de características. La operación max pooling toma el píxel más alto del tamaño del kernel que se ha definido y de esta manera se reduce la varianza y el nivel de complejidad de computación. Esta CNN esta conectada a dos capas totalmente conectadas que serán las encargadas de realizar la clasificación de emociones. La primera capa tiene una dimensión de 48 píxeles por 48 píxeles con una profundidad de 8 canales y la segunda capa es una capa max pooling. En el cuadro 4.4 se detallan las propiedades de cada capa de la red neuronal.

	C1	C2	C3	C4	C5	C6	C7	C8
Tipo	Convolución	Max Pooling	Convolución	Max Pooling	Convolución	Max Pooling	Totalmente Conectada	Totalmente Conectada
Kernel	(3x3)	(2x2)	(3x3)	(2x2)	(3x3)	(2x2)		
Función Activación	ReLu		ReLu		ReLu		Sigmoid	Softmax
Dimensiones	(48x48)	(24x24)	(24x24)	(12x12)	(12x12)	(6x6)		
Profundidad	8	8	16	16	64	64	300	4
Número de parámetros	800		1168		9280		691500	1204

Cuadro 4.4: Propiedades de cada una de las capas de la red.

4.5.2.2. Conjunto de datos

Una red neuronal de cualquier tipo siempre debe ser entrenada con conjuntos de datos relacionados con el objetivo específico de una arquitectura, pues es en la etapa de entrenamiento en donde la red aprende los patrones que existen en los datos. Por ejemplo, cuando el objetivo es reconocer marcas de autos, la red debe ser entrenada con imágenes de diversos tipos de autos. Debido a que el objetivo de esta red es clasificar las emociones de una persona por medio de sus expresiones faciales, el conjunto de datos con el que debe entrenarse la arquitectura debe contener las expresiones faciales de las 6 emociones universales que se describieron en el capítulo dos específicamente en la sección dos. Principalmente se utilizó el conjunto de datos llamado CK+³ el cual contiene un total de 987 imágenes de 7 diferentes emociones, estas imágenes están en escala de grises y tienen una dimensión de 48 píxeles de ancho y de alto, cabe aclarar que en el conjunto de datos existen imágenes que están repetidas dos veces.

- **Enojo:** 135 imágenes.
- **Desprecio:** 54 imágenes.
- **Disgusto:** 177 imágenes.
- **Miedo:** 75 imágenes.
- **Felicidad:** 207 imágenes.
- **Tristeza:** 84 imágenes.

En capítulos anteriores se explico que en trabajos de investigación que se han realizado durante los últimos años, algunos autores habían experimentado con las emociones de disgusto, desprecio y felicidad, es por eso que en el entrenamiento, validación y pruebas de la red se utilizaron las imágenes de las 3 emociones mencionadas anteriormente y se agregaron las imágenes de una emoción mas, la emoción de tristeza. El número de imágenes de cada emoción en el conjunto de datos CK+ no es exactamente igual, algunas emociones tienen más de 100 imágenes y algunas otras tiene menos así que se decidió estandarizar el número de imágenes a 54 debido a que la emoción de desprecio solo contiene 54 imágenes. Finalmente el conjunto de datos tiene un total de 216 imágenes de las cuales el 80 % se utilizó para el entrenamiento, 10 % para validación y 10 % para las pruebas.

³Disponible en: <https://www.kaggle.com/shawon10/ckplus>

El conjunto de datos que se tenía pensado utilizar contenía muy pocos datos lo cual perjudicaba los resultados finales, por lo que se agregaron más imágenes, estas imágenes fueron tomadas del conjunto de datos llamado Extended and Augmented Google FER⁴ el cual contiene un total de 40800 imágenes de 7 diferentes emociones, estas imágenes están en escala de grises y tiene una dimensión de 48 píxeles de alto y de ancho, por lo que no existe ningún problema al realizar esta conjunción. De este conjunto de datos se tomaron 204 imágenes de las mismas 4 emociones.



Figura 4.9: Ejemplos del conjunto de datos de entrenamiento.

⁴Disponible en: <https://www.kaggle.com/prajwalsood/google-fer-image-format>



Figura 4.10: Ejemplos del conjunto de datos de validación.



Figura 4.11: Ejemplos del conjunto de datos de pruebas.

El conjunto de datos final contiene un total de 420 imágenes, de este total se tomaron el 80 % para entrenamiento, 10 % para validación y 10 % para las pruebas. En las figuras 4.9, 4.10 y 4.11 se muestran algunos ejemplos del conjunto de datos utilizado.

4.5.2.3. Métricas

Después de realizar el entrenamiento de un modelo supervisado de inteligencia artificial, es necesario que se evalúe el desempeño del algoritmo propuesto, en la literatura se han definido diversas métricas que nos indican si el rendimiento es aceptable o no es aceptable. Para la evaluación de la arquitectura DEPRECNN se utilizaron las siguientes métricas.

- Accuracy
- Recall
- Precision

La métrica accuracy(exactitud) proporciona la cantidad de datos clasificados correctamente entre la cantidad de datos totales. El accuracy esta definido por la ecuación 4.1, en donde VP= Verdaderos Positivos, VN=Verdaderos Negativos, FP=Falsos Positivos y FN=Falsos Negativos.

$$\frac{(VP + VN)}{(VP + FP + VN + FN)}. \quad (4.1)$$

La métrica precision(precisión) nos indica la cantidad positivos predichos que realmente son positivos. Esta métrica está dada por la ecuación 4.2, en donde VP= Verdaderos Positivos y VN=Verdaderos Negativos.

$$\frac{(VP)}{(VP + FP)}. \quad (4.2)$$

La métrica recall(exhaustividad) nos indica la cantidad que positivos predichos que fueron clasificados correctamente y esta definida por la ecuación 4.3, en donde en donde VP= Verdaderos Positivos y FN=Falsos Negativos.

$$\frac{(VP)}{(VP + FN)}. \quad (4.3)$$

Capítulo 5

Experimentos y resultados

Después de revisar la metodología propuesta, se procederá a mostrar los resultados de los experimentos realizados. En este capítulo se explicará el proceso del entrenamiento de la red neuronal, la etapa de pruebas y el resultado final con la interfaz gráfica desarrollada.

5.1. Etapa de entrenamiento

Antes de realizar el entrenamiento de la red neuronal fue necesario aplicar un tratamiento a las imágenes del conjunto de datos. Con este tratamiento realizado, las expresiones faciales de las personas están más marcadas y esto nos conduce a mejores resultados que en los experimentos realizados con imágenes crudas. En la figura 5.1 se pueden observar algunos ejemplos de imágenes con dicho tratamiento.

El conjunto de datos para la etapa de entrenamiento esta compuesto por 336 imágenes de 4 emociones diferentes, por cada emoción se tienen 84 imágenes.

- **Felicidad:** 84 imágenes.
- **Desprecio:** 84 imágenes.
- **Disgusto:** 84 imágenes.
- **Tristeza:** 84 imágenes.

El entrenamiento de la red que se propuso en el capítulo anterior se realizó por 30 épocas, este numero se determinó después de realizar diversos experimentos en dondes los resultados no fueron los deseados. Para evaluar el

rendimiento de esta arquitectura se utilizaron las métricas que se detallaron en el capítulo anterior.

En el cuadro 5.1 se muestran los parámetros que se utilizaron en la etapa del entrenamiento así como las diferentes métricas que se utilizaron para medir el rendimiento de la red.



Figura 5.1: Ejemplos de imágenes ecualizadas.

	Epocas	Función de costo	Optimizador	Tasa de aprendizaje	Métricas
DEPRECNN	30	Categorical Cross-Entropy	Adam	0.001	Accuracy Recall Precision

Cuadro 5.1: Parámetros del entrenamiento de la red.

5.2. Etapa de pruebas

En la etapa de experimentación, se entrenó la misma arquitectura con imágenes crudas y con imágenes que tienen un histograma ecualizado, se decidió aplicar esta técnica de tratamiento de imágenes al conjunto de entrenamiento debido a que con esta técnica se reduce la varianza de los datos. En

el cuadro 5.2 se puede observar el rendimiento de cada uno de los modelos de inteligencia artificial que se desarrollaron. La arquitectura que fue entrenada con imágenes con un histograma ecualizado es mejor en las tres métricas.

	Ecualización	Accuracy	Precision	Recall
DEPRECNN	Si	0.7750	0.7750	0.7750
DEPRECNN	No	0.7250	0.7250	0.7250

Cuadro 5.2: Resultados de las arquitecturas entrenadas.

Para entender un poco mejor los resultados de las métricas que se obtuvieron con el conjunto de prueba, se recomienda observar la evolución de las métricas a través de las épocas de entrenamiento. En la figura 5.2 se muestran las gráficas de la evolución de nuestra red neuronal a través de las épocas y se puede observar que entre la época uno y la época cinco fue en donde las métricas alcanzaron un máximo y de ahí las métricas oscilaron hasta llegar a un punto de estabilización.

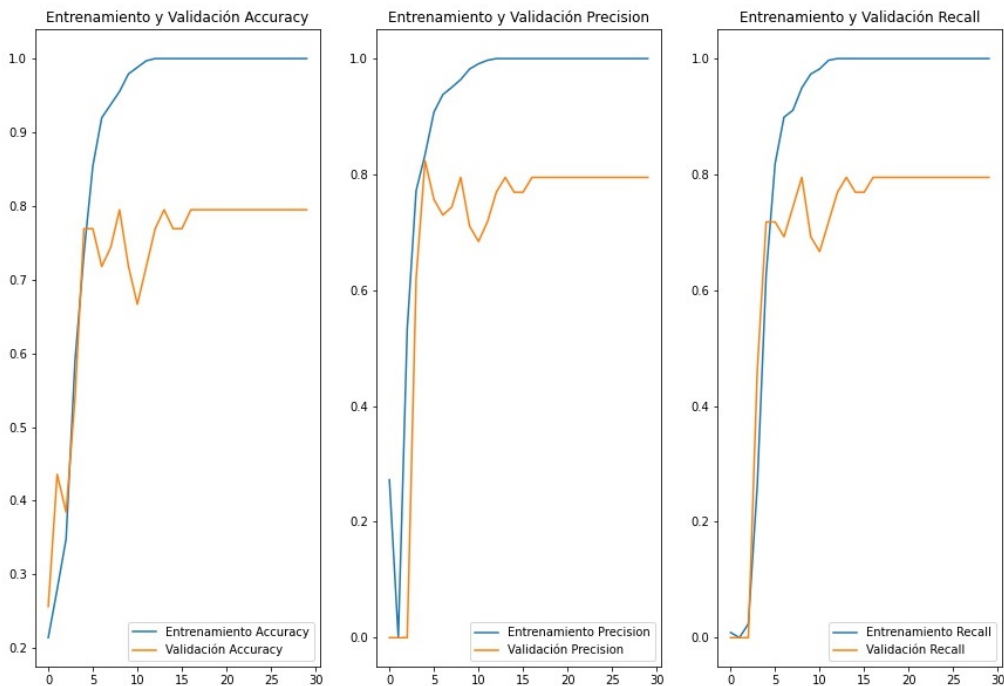


Figura 5.2: Métricas a través del entrenamiento.

Cuando se desarrolla y entrena una red neuronal para realizar una clasificación ya sea de imágenes, datos numéricos o datos de texto, es recomendable visualizar el número de ejemplos que se clasificaron correctamente así como los falsos positivos y los falsos negativos. La matriz de confusión es una herramienta que es común para visualizar estos números.

En la imagen 5.3 se pueden observar cuantas imágenes se clasificaron con la emoción correcta y cuantas imágenes se clasificaron con una emoción incorrecta. Por ejemplo, las diez imágenes que contienen expresiones faciales relacionadas con la felicidad, fueron clasificadas correctamente, mientras que con la emoción de disgusto, cuatro fueron clasificadas como feliz y seis con la emoción correcta.

Matriz de confusión

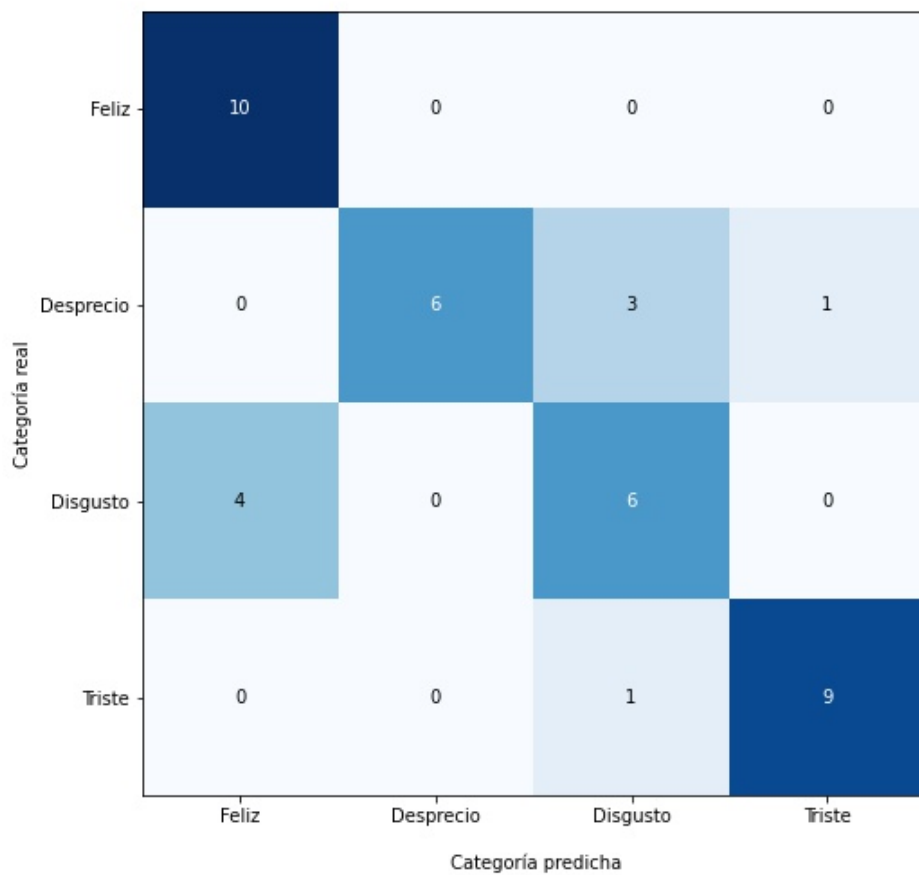


Figura 5.3: Matriz de confusión.

Por último, en la imagen 5.4 se pueden observar algunas de las imágenes del conjunto de prueba que fueron clasificadas correctamente e incorrectamente por el modelo de Inteligencia Artificial.



Figura 5.4: Emociones predichas por la red neuronal.

5.3. Resultados finales

Como resultado final, se desarrolló una interfaz gráfica en donde se incrustó el modelo entrenado. Como se describió en el capítulo cuatro, específicamente en la sección 4.5.1.4 la interfaz gráfica contiene un recuadro en donde se visualiza la prevista de la cámara o la imagen existente que el usuario decida subir. Además, en la interfaz contiene botones para realizar las acciones de subir fotografía o tomar fotografía. En la figura 5.5 se muestra la interfaz desarrollada.

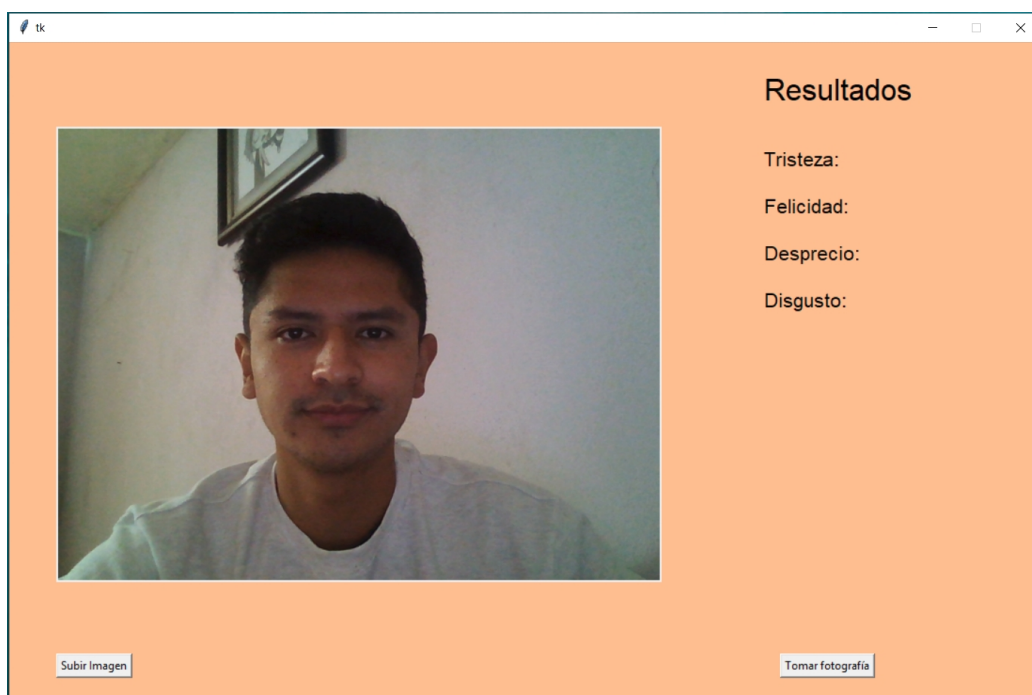


Figura 5.5: Interfaz gráfica desarrollada.

Para realizar la detección del rostro en una fotografía se utiliza el algoritmo de Viola-Jones, se aplica un tratamiento a la imagen y por ultimo, el modelo de inteligencia artificial realiza la clasificación de emociones y realiza el posible diagnóstico.

El flujo de datos del funcionamiento del sistema sería el siguiente.

1. El usuario toma o sube una fotografía.
2. Se detecta el rostro en la imagen.
3. Se recorta el rostro de la imagen.

4. Se reescala la imagen a 48x48 píxeles.
5. Se aplica una ecualización de histograma.
6. Se realiza la clasificación.
7. Se realiza el diagnóstico.
8. Se muestran los resultados en pantalla.

Como este diagnóstico está dado por una inteligencia artificial, el resultado puede ser erróneo, es por ello que el sistema muestra una ventana emergente avisando al usuario que el resultado puede ser erróneo y que es mejor que acuda a un especialista. En la imagen 5.6 se puede observar la ventana emergente.

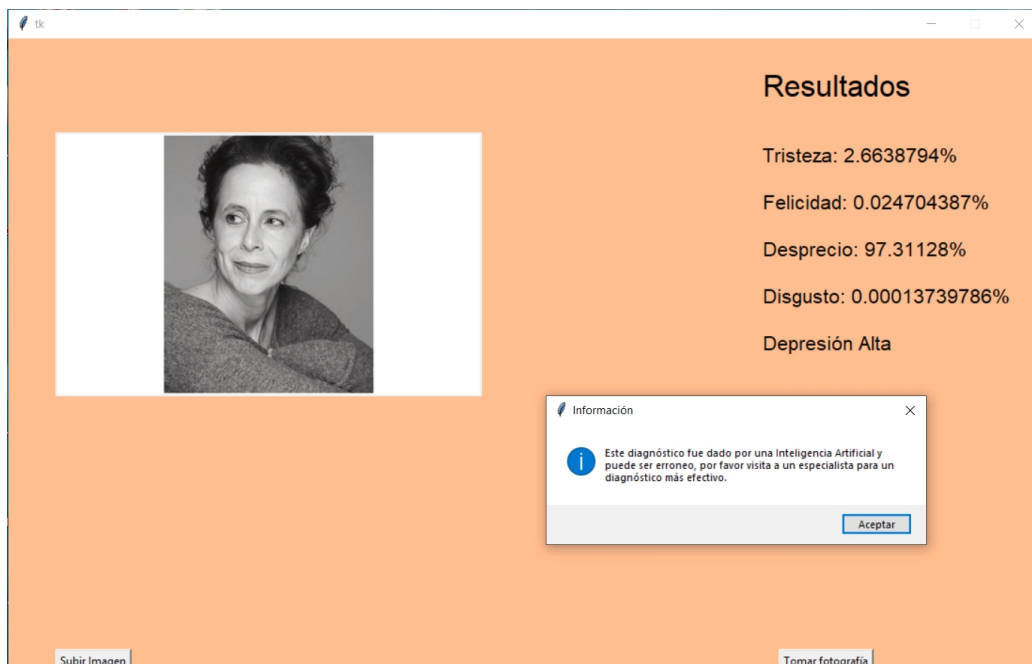


Figura 5.6: Aviso del diagnóstico.



Figura 5.7: Ejemplo de diagnóstico.

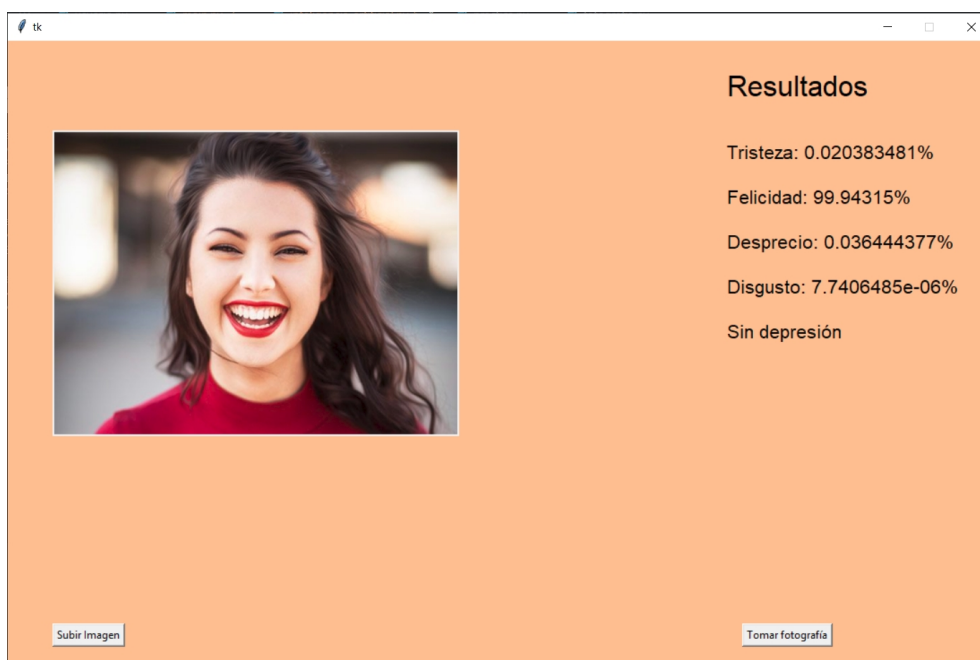


Figura 5.8: Ejemplo de diagnóstico.

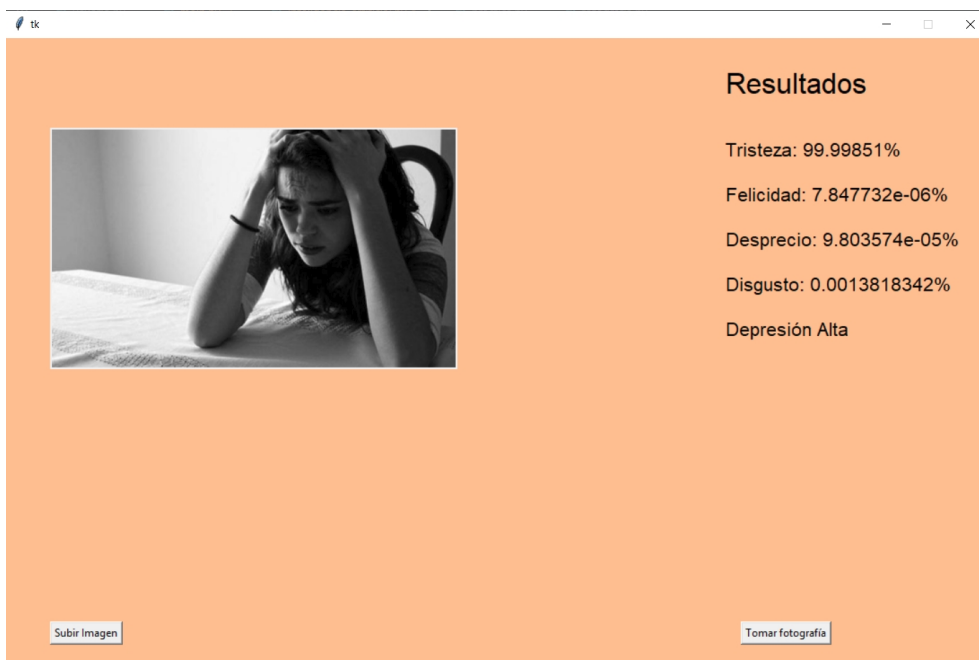


Figura 5.9: Ejemplo de diagnóstico.

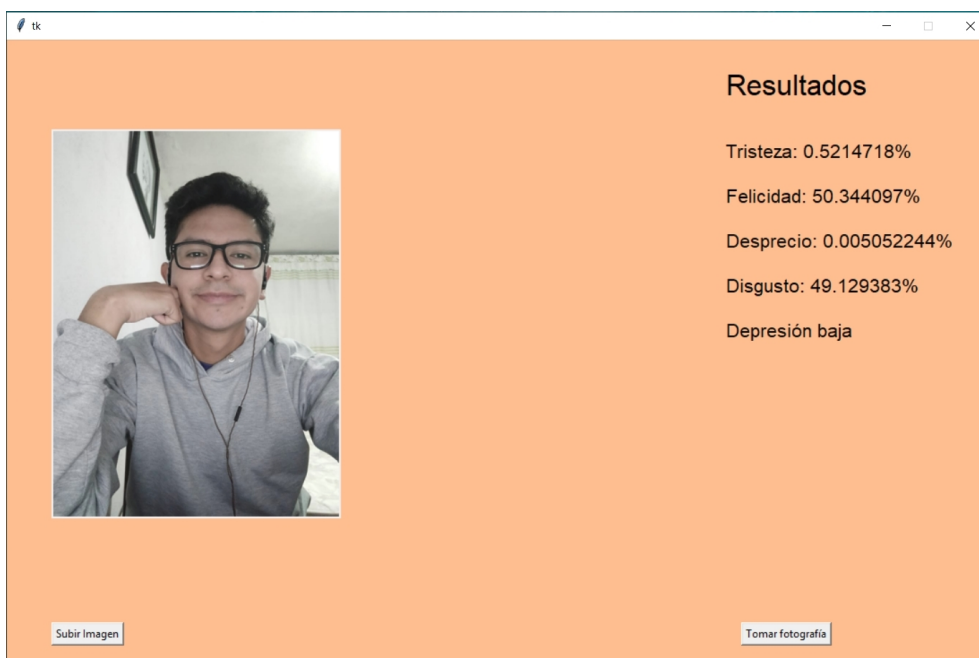


Figura 5.10: Ejemplo de diagnóstico.

En las imágenes 5.7,5.8, 5.9, 5.10, se muestran algunos diagnósticos que se realizaron con el sistema terminado. En las pruebas que se realizaron con el sistema completo, se utilizaron imágenes de Google¹ así como imágenes propias tomadas con la webcam y con el celular.

¹www.google.com

Capítulo 6

Conclusión y trabajo futuro

En este capítulo se dará un cierre escrito a este trabajo de experimentación e investigación y comentaremos los aprendizajes obtenidos así como la conclusión. Se comentará el trabajo que se realizará a futuro para mejorar la metodología de desarrollo y la etapa de experimentación.

6.1. Conclusión

El diagnóstico de la enfermedad de depresión es una tarea complicada que hasta el momento solo los expertos en el área de psicología han podido realizar con una precisión considerablemente alta. De acuerdo con la investigación realizada, no ha sido posible obtener un diagnóstico automático con una alta precisión. Esto es debido a que la depresión esta asociada a emociones humanas y la tecnología actual aun no ha sido capaz de comprender las emociones de una persona.

La detección de depresión por expresiones faciales con una simple fotografía tiene un nivel de error considerablemente alto debido a que el usuario puede experimentar una emoción de tristeza, disgusto o desprecio momentánea y de acuerdo con [25], uno de los síntomas de la depresión es que el paciente se sienta triste o sin esperanza en un periodo de dos o más semanas. Una de las funcionalidades de este sistema es detectar la emoción de una persona por medio de sus expresiones faciales entonces, si el usuario usa esta funcionalidad por un periodo de dos o más semanas el diagnóstico de depresión puede ser más asertivo. Incluso este método de diagnóstico puede ser combinado con otra técnica.

Las redes neuronales convolucionales son consideradas como una herramienta especializada en visión computacional para realizar clasificación de imágenes o para detectar objetos en una imagen. En este trabajo los resul-

tados obtenidos con la clasificación de emociones son aceptables, principalmente en felicidad y tristeza. Está claro que estos resultados pueden mejorar en un trabajo futuro pero esto se conversará en mayor profundidad en la siguiente sección.

6.2. Trabajo futuro

La arquitectura propuesta para este proyecto se puede considerar como aceptable, sin embargo, como en todo proyecto, los resultados se pueden mejorar en un futuro.

El tratamiento de imágenes es una área bastante amplia que no se ha cubierto en su totalidad en este trabajo, por lo que una las tareas que se seguirán trabajando serán las de investigar, analizar y experimentar con nuevos métodos de tratamiento de imágenes, por ejemplo, los filtros de Gabor. La investigación de temas en psicología relacionados con la depresión es una tarea que no se ha explorado al 100 % por lo que se seguirá trabajando en la investigación acompañado de un experto en el área.

La arquitectura puede ser mejorada para aumentar los valores que se obtuvieron en las métricas utilizando algunas alternativas de redes neuronales convolucionales por, una red neuronal convolucional recurrente(RCNN).

Anexo

En el capítulo dos sección dos se habló sobre las 44 unidades de acción que Paul Ekman propuso, sin embargo solo se presentaron el número de unidades de acción que están presentes en las seis emociones universales. En el cuadro 6.1 se especifican los nombres y el número de unidad de acción de las 44 unidades que Paul Ekman definió.

Unidad de acción	Nombre de la acción
1	Levantamiento interior de ceja
2	Levantamiento exterior de ceja
4	Bajar cejas
5	Levantamiento del párpado superior
6	Levantamiento de mejillas
7	Apretar párpado(s)
8	Labios encimados uno de otro
9	Arrugar la nariz
10	Levantamiento del labio superior
11	Profundidad nasolabial
12	Tiramiento labial esquinial
13	Tiramiento labial frontal
14	Hoyuelo facial
15	Depresión labial esquinial
16	Depresión labial frontal
17	Levantamiento de barbilla
18	Arruga labial
19	Muestrero de lengua
20	Apretar los labios
21	Apretamiento de cuello
22	Embudo labial
23	Morder labios
24	Presión Labial
25	Deslizamiento labial

Unidad de acción	Nombre de la acción
26	Caída de la mandíbula
27	Apretamiento bucal
28	Lamido labial
29	Tracción de la mandíbula
30	Deslizamiento de mandíbula
31	Contracción mandibular
32	Mordida labial
33	Succión de mejillas
34	Inflar mejillas
35	Soplido de mejillas
36	Protuberancia de lengua
37	Limpieza labial
38	Dilatado nasal
39	Compresión nasal
41	Bajada de párpados
42	Contracción retinal
43	Ojos cerrados
44	Recolector retinal
45	Parpadeo
46	Guiño

Cuadro 6.1: Unidades de acción.

Bibliografía

- [1] O. M. de la Salud, “Depresión.”
- [2] P. Ekman and D. Keltner, “Universal facial expressions of emotion,” *Se-gerstrale U, P. Molnar P, eds. Nonverbal communication: Where nature meets culture*, vol. 27, p. 46, 1997.
- [3] D. Venkataraman and N. S. Parameswaran, “Extraction of facial features for depression detection among students,” *Int. J. Pure Appl. Math*, vol. 118, no. 7, pp. 455–463, 2018.
- [4] M. Gavrilescu and N. Vizireanu, “Predicting depression, anxiety, and stress levels from videos using the facial action coding system,” *Sensors*, vol. 19, no. 17, p. 3693, 2019.
- [5] E. J. Lawler and S. R. Thye, “Bringing emotions into social exchange theory,” *Annual review of sociology*, vol. 25, no. 1, pp. 217–244, 1999.
- [6] N. K. Denzin, “On understanding emotion. new brunswick,” 2009.
- [7] T. D. Kemper, “How many emotions are there? wedding the social and the autonomic components,” *American journal of Sociology*, vol. 93, no. 2, pp. 263–289, 1987.
- [8] M. A. Wenger, F. N. Jones, and M. H. Jones, “Emotional behavior,” *Emotion: bodily change. Princeton, NJ: van Nostrand*, 1962.
- [9] P. J. Lang, “Fear reduction and fear behavior: Problems in treating a construct.,” in *Research in psychotherapy conference, 3rd, May-Jun, 1966, Chicago, IL, US*, American Psychological Association, 1968.
- [10] L. Brody, *Gender, emotion, and the family*. Harvard University Press, 2009.

- [11] L. A. Sroufe, M. S. tr Donís Galindo, and J. rev. téc Huerta Ibarra, *Desarrollo emocional: la organización de la vida emocional en los primeros años*. 2000.
- [12] E. J. Lawler, S. R. Thye, and J. Yoon, “Social exchange and micro social order,” *American sociological review*, vol. 73, no. 4, pp. 519–542, 2008.
- [13] C. Marinetti, P. Moore, P. Lucas, and B. Parkinson, “Emotions in social interactions: Unfolding emotional experience,” in *Emotion-oriented systems*, pp. 31–46, Springer, 2011.
- [14] C. Darwin and T. R. F. Rodríguez, *La expresión de las emociones en los animales y en el hombre*. Alianza Madrid, 1984.
- [15] H. A. de Arcaya Ajuria, “La comunicación no verbal. interrelaciones entre las expresiones faciales innatas y las aprendidas,” *Gazeta de antropología*, vol. 19, 2003.
- [16] P. Ekman, “Cross-cultural studies of facial expression,” *Darwin and facial expression: A century of research in review*, vol. 169222, no. 1, 1973.
- [17] P. Ekman, “Facial action coding system,” 1977.
- [18] Y. Zhang and Q. Ji, “Active and dynamic information fusion for facial expression understanding from image sequences,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 5, pp. 699–714, 2005.
- [19] C. D. Lastreto, “Nuestras 6 emociones básicas. somos seres emocionales que razonan.”
- [20] P. Retamal, *Depresión*. Editorial Universitaria, 1998.
- [21] P. M. Lewinsohn, A. M. Zeiss, R. A. Zeiss, and R. Haller, “Endogeneity and reactivity as orthogonal dimensions in depression.,” *Journal of Nervous and Mental Disease*, 1977.
- [22] D. O’Leary, “The endogenous subtype and naturalistic course in depression,” *Journal of affective disorders*, vol. 41, no. 2, pp. 117–123, 1996.
- [23] G. Parker, D. Hadzi-Pavlovic, K. Wilhelm, I. Hickie, H. Brodaty, P. Boyce, P. Mitchell, and K. Eyers, “Defining melancholia: properties of a refined sign-based measure,” *The British Journal of Psychiatry*, vol. 164, no. 3, pp. 316–326, 1994.

- [24] L. B. Alloy, L. Y. Abramson, G. I. Metalsky, and S. Hartlage, “The hopelessness theory of depression: Attributional aspects,” *British Journal of Clinical Psychology*, vol. 27, no. 1, pp. 5–21, 1988.
- [25] A. P. Association. and A. P. Association., *Diagnostic and statistical manual of mental disorders : DSM-5*. American Psychiatric Association Arlington, VA, 5th ed. ed., 2013.
- [26] J. S. Beck and A. T. Beck, *Cognitive therapy: Basics and beyond*. No. Siri) i9780898628470, Guilford press New York, 1995.
- [27] F. Izaurieta and C. Saavedra, “Redes neuronales artificiales,” *Departamento de Física, Universidad de Concepción Chile*, 2000.
- [28] M. C. Soberanis, “Inspiración biológica de las redes neuronales artificiales.”
- [29] M. E. Raffino, “Sinapsis.”
- [30] K. Academy, “Neurotransmisores y receptores.”
- [31] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.
- [32] R. Keim, “How to train a basic perceptron neural network.”
- [33] “How to visualize convolutional features in 40 lines of code | by fabio m. graetz | towards data science.” <https://towardsdatascience.com/how-to-visualize-convolutional-features-in-40-lines-of-code-70b7d87b0030>. (Accessed on 08/03/2021).
- [34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [35] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in *2017 International Conference on Engineering and Technology (ICET)*, pp. 1–6, Ieee, 2017.
- [36] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [37] F. Ramiandrisoa, J. Mothe, F. Benamara, and V. Moriceau, “Irit at e-risk 2018,” in *9th Conference and Labs of the Evaluation Forum, Living Labs (CLEF 2018)*, pp. 1–12, 2018.

- [38] H. Jiang, B. Hu, Z. Liu, G. Wang, L. Zhang, X. Li, and H. Kang, “Detecting depression using an ensemble logistic regression model based on multiple speech features,” *Computational and mathematical methods in medicine*, vol. 2018, 2018.
- [39] K. Kipli, A. Z. Kouzani, and L. J. Williams, “Towards automated detection of depression from brain structural magnetic resonance images,” *Neuroradiology*, vol. 55, no. 5, pp. 567–584, 2013.
- [40] F. Lundh, “An introduction to tkinter,” *URL: www.pythonware.com/library/tkinter/introduction/index.htm*, 1999.
- [41] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattemberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015. Software available from tensorflow.org.
- [42] T. P. Velavan and C. G. Meyer, “The covid-19 epidemic,” *Tropical medicine & international health*, vol. 25, no. 3, p. 278, 2020.
- [43] W. Cullen, G. Gulati, and B. Kelly, “Mental health in the covid-19 pandemic,” *QJM: An International Journal of Medicine*, vol. 113, no. 5, pp. 311–312, 2020.
- [44] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1, pp. I–I, Ieee, 2001.