Facultad de Ingeniería

Magíster en Gestión de Tecnologías de la Información y Telecomunicaciones

RECONOCIMIENTO DE ESTRÉS EN IMÁGENES DE ROSTROS MEDIANTE DEEP LEARNING CON TRANSFER LEARNING

Trabajo final para optar al grado de Magíster en Gestión de Tecnologías de la Información y Telecomunicaciones.

Autor:

Cristian Muñoz Contreras

Profesor guía:

Romina Torres Torres

Profesor co-guía:

Rodrigo Salas Fuentes

Santiago, Chile

2021

RESUMEN

Actualmente las personas se ven afectadas en su vida cotidiana y laboral por una gran carga mental denominada estrés, la que provoca serios daños a la salud llegando a imposibilitar la continuidad en sus actividades y la ocurrencia de accidentes. Su detección es realizada por especialistas del área de la salud, quienes aplican herramientas, procedimientos y técnicas que dado los tiempos que involucran, no permiten la toma de decisiones de forma oportuna, ya que el estrés se presenta en el corto plazo por lo que se requiere que su detección sea rápida (semi-automática o automática), eliminando los errores debido a la subjetividad humana. Este estudio utiliza 791 imágenes de rostros de jugadores de futbol, clasificando como estrés a las que se encuentran en la situación de patear un penal, que es definida como estresante por las investigaciones en este deporte. Para su reconocimiento semiautomático, se desarrolló un método basado en aprendizaje por transferencia y supervisado mediante una red neuronal convolucional VGG16, se establecieron tres experimentos con veinticinco pruebas cada uno, donde el mejor resultado obtiene una exactitud (accuracy) de 97,5% (pérdidas de 5,7%), superando en un 12,1% al método existente que cumple con los requisitos.

Palabras clave: Detección de estrés, red neuronal convolucional, VGG16, reconocimiento semiautomático, aprendizaje por transferencia y supervisado, exactitud (accuracy).

TABLA DE CONTENIDOS

I. CAPITULO I: INTRODUCCIÓN

1.1. Fundamentación

En la actualidad las personas se ven afectadas en su vida cotidiana y en especial en la laboral por una gran carga mental denominada estrés, la que provoca serios daños a la salud llegando a imposibilitar la continuidad en sus actividades. Esta condición puede ser producto de factores personales y/o del entorno, siendo los primeros los relacionados con las capacidades o recursos propios de la persona para sobrellevar esta carga, y los segundos se explican en las dificultades que el medio, en especial el trabajo, le exigen enfrentar de forma eficaz.

En Chile existe la Ley 16.744/68 de Accidentes Laborales y Enfermedades Profesionales [1], siendo un seguro de carácter obligatorio pagado por el empleador que contempla tanto prestaciones médicas como compensaciones económicas, y en su Decreto Supremo Nº 109/68 [2] define Neurosis profesionales incapacitantes, a las que pueden adquirir distintas formas de presentación clínica, tales como: trastorno de adaptación, depresión reactiva, trastorno por somatización y por dolor crónico y establece como fase crónica e irreversible de la enfermedad cuando provoca entre un 40% a 65% de incapacidad temporal. Además, se indica que esta es causada por trabajos que expongan al riesgo de tensión psíquica siempre y cuando se compruebe relación de causa a efecto con el trabajo.

La Superintendencia de Seguridad Social de Chile (SUSESO) realiza todos los años un reporte con las estadísticas de accidentabilidad y enfermedades profesionales, en su reporte del año 2021, indica que los casos de salud mental siguen siendo importantes dentro de las enfermedades laborales representando el 35% del total [3].

En estudio realizado por SUSESO [4], cuyo objetivo fue encontrar los principales determinantes de los accidentes laborales sufridos por trabajadores cubiertos por la ley, utilizando una base de datos con las respuestas de un cuestionario aplicado en los años 2017 y 2018, determinaron que las variables más importantes para explicar los accidentes laborales están relacionadas con la auto percepción en la salud de los trabajadores como son la salud general, salud mental, vitalidad y, sobre todo, el estrés.

Estos factores personales junto con los factores del trabajo son considerados como las causas básicas que explican la ocurrencia de un accidente, por tanto, para tratar el riesgo es fundamental el reconocimiento y control de estas causas de manera oportuna.

El detectar cuando una persona se encuentra bajo condición de estrés de manera oportuna, permitiría tomar decisiones de manera de evitar con ello la ocurrencia de accidentes por factores personales. Actualmente, la detección queda sujeta a especialistas del área de la salud, quienes aplican herramientas, procedimientos y técnicas que dado los tiempos que involucran (desde horas hasta días), no permiten la toma de decisiones de forma oportuna, ya que el estrés se presenta en el corto plazo por lo que se requiere que su detección sea rápida y automática, abordando también errores debido a la subjetividad humana.

Por tanto, se plantea como pregunta de investigación si existe en la literatura maneras de poder **detectar automáticamente o semi-automáticamente el estrés**.

1.2. Discusión bibliográfica

Russell y Zhandong [5] desarrollaron un método para la detección de estrés mediante redes neuronales profundas, el cual considera investigaciones anteriores, las que han demostrado que el análisis de señales fisiológicas es un predictor fiable del estrés. Tales señales (pulso del volumen de sangre, aceleración, actividad electrodérmica, ritmo cardiaco, temperatura de la piel, actividad eléctrica de los músculos esqueléticos, respiración) se obtienen de los sensores que están conectados al cuerpo humano. Los investigadores han intentado detectar el estrés mediante métodos tradicionales de aprendizaje automático para analizar señales fisiológicas, con resultados que oscilan entre el 50 y el 90% de exactitud (accuracy). Una limitación de los algoritmos tradicionales de aprendizaje automático es el requisito de características, la exactitud disminuye si las características se identifican incorrectamente. Para abordar esta deficiencia desarrollaron dos redes neuronales: una red neuronal convolucional unidimensional (1D) y una red neuronal perceptrón multicapa.

Las redes neuronales profundas analizan los datos fisiológicos recopilados desde los sensores de muñeca y de pecho para realizar dos tareas, adaptaron cada red neuronal para analizar los datos de los sensores de pecho (red neuronal convolucional 1D) o de los sensores de muñeca (red neuronal perceptrón multicapa). La primera tarea fue la clasificación binaria para la detección de estrés, en la que las redes diferencian entre estresado y no estresado. La segunda tarea fue la clasificación de 3 clases de emociones, en la que las redes diferencian entre estados de línea de base, estresados y relajados. Las redes fueron entrenadas y probadas sobre la base de datos recopilados y disponibles públicamente en estudios anteriores.

Los resultados obtenidos en la red neuronal convolucional profunda alcanzaron tasas de exactitud del 99,80% y 99,55% para clasificación binaria y de 3 clases, respectivamente. La red neuronal profunda del perceptrón multicapa alcanzó un 99,65% y un 98,38% de tasas de exactitud para clasificación binaria y de 3 clases, respectivamente. Los autores señalan que el desempeño de las redes exhibió una mejora significativa sobre métodos

anteriores que analizaban señales fisiológicas, tanto para la detección de estrés binario como para la clasificación de emociones en 3 clases.

La propuesta de Kang, Shin, Jung y Kim [6] consiste en un algoritmo de conjunto que puede determinar con exactitud los estados de estrés mental utilizando una red neuronal convolucional modificada (CNN) con una arquitectura de memoria a corto y a largo plazo (LSTM) y la señal de un electrocardiograma (ECG). Es posible clasificar las señales de estrés analizando las señales de ECG y extracción de parámetros específicos. Para maximizar el rendimiento del algoritmo de clasificación de tensiones propuesto, se utilizaron la transformada rápida de Fourier (FFT) y espectrogramas para preprocesar señales de ECG y producir señales por separado en los dominios del tiempo y de la frecuencia.

El modelo de conjunto propuesto CNN-LSTM logró una clasificación de estrés con una exactitud del 98,3%. Estos resultados exhiben una aproximación 14,7% de mejora en la exactitud en comparación con estudios anteriores que clasifican lo existente bajo estrés y sin estrés. En el futuro, planean mejorar el método de preprocesamiento, como una sutil eliminación de ruido de señales biológicas, y para mejorar la precisión aplicando un filtro de transformación portátil que eliminará las fluctuaciones de la línea de base y el ruido utilizando Transformadas de Fourier. El clasificador de estrés propuesto se espera que sea útil en el manejo de la salud mental, ya que puede clasificar de forma rápida y precisa el estrés experimentado por las personas en la actualidad. También se espera que ayude a prevenir diversas enfermedades como depresión, hipertensión arterial y diabetes a través del manejo periódico del estrés.

Zhang, Feng, Li, Jin y Cao [7] desarrollaron una red de detección de estrés de dos niveles basada en video TSDNet (Two-leveled Stress Detection Network), que integra un detector de nivel facial y un detector de nivel de acción para comprender las expresiones faciales y movimientos de acción para la identificación del estrés. Diseñaron una agrupación de múltiples escalas a nivel de la cara con mecanismo de atención y un mecanismo de atención marco a nivel de acción. El primero empleó la agrupación promedio de múltiples escalas con diferentes tamaños de puntos para comprender los rasgos faciales relacionados con el estrés, y el último se centró en marcos de movimiento corporales clave relacionados con estados estresados. Un integrador ponderado de flujo con atención local y global se utilizó para fusionar los resultados de los detectores de nivel de acción y de rostro.

El estudio consideró a 122 voluntarios (58 hombres and 64 mujeres con edades entre 18 y 26 años), se utilizaron cámaras infrarrojas para registrar las reacciones afectivas (neutrales, relajadas y estresadas) de los participantes cuando vieron tres tipos diferentes de videoclips de 2 minutos. Los videoclips neutrales eran sobre paisajes o preparación de comidas, los relajados fueron lo más destacado de programas de entretención y los estresados fueron los programas de ciencia, donde cada programa científico fue seguido de una prueba de respuestas a preguntas. En resumen, cada video obtenido duró 2 minutos y etiquetados como "no estresado" y "estresado".

Con lo anterior, crearon un conjunto de datos de video que contiene 2.092 videoclips etiquetados y evaluaron el desempeño de TSDNet en el conjunto de datos. Los resultados experimentales muestran que TSDNet superó a los existentes hechos manualmente con estrategias de ingeniería de características, y la integración de detectores de nivel facial y de nivel de acción podría mejorar la exactitud de detección obtenida que fue de 85,42%. En el trabajo futuro, planean agregar la transmisión de audio al marco para explorar el audio y el video en los métodos para la detección del estrés.

Con respecto a los métodos existentes para el reconocimiento de estrés mediante inteligencia artificial (*machine learning y deep learning*), se han encontrado investigaciones que utilizan lecturas de varias señales biológicas capturadas por sensores adheridos al cuerpo de la persona, las que serán procesadas por las redes neuronales convolucionales (CNN) en una o dos dimensiones, obteniendo exactitudes en la predicción de estrés del orden del 99%. Por otra parte, el método encontrado que no utiliza sensores adheridos al cuerpo y que utiliza imágenes de rostros obtenidas mediante cámara infrarroja, en las cuales las personas son sometidas a visualizar videos con el fin de provocar reacciones, ya sean neutras, relajadas o estresadas, dichas imágenes son etiquetadas como estresado y no estresado. Posteriormente esta base de datos de imágenes es procesada por una CNN en tres dimensiones para luego ser integradas sus detecciones de rostros y acciones, obteniendo una exactitud de la predicción de un 85,42%, siendo este el valor a utilizar como referencia para el método propuesto.

1.3. Propuesta

El objetivo general del proyecto es desarrollar un método basado en inteligencia artificial para la detección semi-automática de estrés en una persona, basada en la expresión de su rostro, con una exactitud (accuracy) igual o superior al 85%.

Los objetivos específicos del proyecto son:

1. Adaptar una base de datos y configuraciones con imágenes de rostros en situaciones con estrés y sin estrés.
2. Implementar un modelo de Deep Learning para la detección de estrés en imágenes de rostros.
3. Implementar un modelo de Deep Learning con transfer learning para mejorar el desempeño en la detección de estrés en imágenes de rostros.

1.4. Metodología y plan de trabajo

La metodología a utilizar en el proyecto es el descubrimiento de conocimiento en bases de datos KDD (Knowledge Discovery in Databases), es básicamente un proceso automático en el que se combinan descubrimiento y análisis. El proceso consiste en extraer patrones en forma de reglas o funciones, a partir de los datos, para que el usuario los analice. Esta tarea implica generalmente preprocesar los datos, hacer minería de datos (data mining) y presentar resultados [8]. Las etapas del proceso KDD son las siguientes:

- Selección.
- Preprocesamiento/limpieza.
- Transformación/reducción.
- Minería de datos (data mining).
- Interpretación/evaluación.

Para alcanzar el primer objetivo específico es necesario seleccionar imágenes de rostros de personas y que representen reacciones de estrés y no estrés en situaciones en las cuales los estudios de comportamiento humano las validen objetivamente. Dichas imágenes se deben preprocesar para dar cumplimiento a las exigencias de calidad, iluminación y resolución para permitir la eliminación de posibles errores que no den cuenta del mundo real en el cual están inmersas las personas. En la etapa de transformación, las imágenes se deben estandarizar a los requerimientos técnicos de las etapas siguientes, los cuales determinaran por ejemplo su tamaño y resolución. Para realizar la minería de datos que corresponde a la búsqueda y descubrimiento de patrones en los rostros de las personas, y a la creación de modelos predictivos que requerirán de un algoritmo que permita la clasificación de las condiciones establecidas.

Para alcanzar el segundo objetivo específico, se establecerán los experimentos necesarios para obtener las métricas mediante la interpretación y evaluación de los resultados.

El plan de trabajo considera el desarrollo de tareas en función de los objetivos específicos propuestos y bajo metodología KDD, ver Tabla 1.

Tabla 1: Carta Gantt proyecto.

Objetivos	Nombre de tarea	Duración
Objetivos específicos 1, 2 y 3	**Implementación de un modelo de Deep Learning para el reconocimiento de estrés en imágenes de rostros con transfer learning**	**37 días**
	Implementar método	11 días
	Prueba	**14 días**
	Definir los casos de prueba	3 días
	Definir los procedimientos de prueba	3 días
	Probar	3 días
	Documentar las pruebas realizadas	5 días
	Resultados del método desarrollado para el reconocimiento de estrés	**24 días**
	Evaluación de resultados	**5 días**
	Evaluar los resultados obtenidos con respecto a los esperados	5 días
	Documentación de los resultados	**5 días**

Fuente: Elaboración propia.

1.5. Contribución del trabajo

Esta investigación aplicada propone, un método semi-automático mediante redes neuronales convolucionales, que permita la detección de estrés en las personas mediante imágenes de sus rostros. Es fundamental la validez en la clasificación de la condición de estrés y no estrés de las imágenes a utilizar para el aprendizaje del modelo, obteniendo con ello una óptima generalización y altos niveles de exactitud de la predicción. Esto en el futuro podría ser utilizado para la toma de decisiones en la asignación de personal o

para la verificación de dicha condición de salud mental, el cual se basa en las características más relevantes de expresión del rostro que permitan determinar dicha condición. Los métodos encontrados en el estado del arte, en su mayoría utilizan instrumentación para obtener señales desde el cuerpo para predecir el estrés, en términos comparativos el método propio supera en un 12,1% de exactitud al que cumple con los requisitos, pero es superado en 2,3% de exactitud al mejor de los que utilizan instrumentación adherida al cuerpo.

1.6. Organización y presentación del trabajo

Este trabajo se divide en dos capítulos principales: el capıtulo I describe la introducción al trabajo compuesta por las secciones que presentan la fundamentación, la discusión bibliográfica, la propuesta que incluye los objetivos generales y específicos, la metodología y plan de trabajo, y finalmente la contribución del trabajo.

El capítulo II describe el artículo propuesto considerando los siguientes ítems: en la sección 2.1 se realiza la introducción a la propuesta. En la sección 2.2 se desarrolla el marco teórico que sustenta la propuesta, en la sección 2.3 se describen los materiales y métodos utilizados. En la sección 2.4 los resultados obtenidos al aplicar el método propuesto, en la sección 2.5 se realiza una discusión de los resultados enfocado en el desempeño del modelo en base a sus curvas de aprendizaje y métricas. En la sección 2.6 las conclusiones del trabajo junto a observaciones respecto a futuras investigaciones aplicadas derivadas de los resultados obtenidos.

Finalmente se encuentran las referencias bibliográficas y anexo que menciona algunos instrumentos para evaluar estrés.

II. CAPITULO II: ARTICULO PROPUESTO

Resumen: Actualmente las personas se ven afectadas en su vida cotidiana y laboral por una gran carga mental denominada estrés, la que provoca serios daños a la salud llegando a imposibilitar la continuidad en sus actividades y la ocurrencia de accidentes. Su detección es realizada por especialistas del área de la salud, quienes aplican herramientas, procedimientos y técnicas que dado los tiempos que involucran, no permiten la toma de decisiones de forma oportuna, ya que el estrés se presenta en el corto plazo por lo que se requiere que su detección sea rápida (semi-automática o automática), eliminando los errores debido a la subjetividad humana. Este estudio utiliza 791 imágenes de rostros de jugadores de futbol, clasificando como estrés a las que se encuentran en la situación de patear un penal, que es definida como estresante por las investigaciones en este deporte. Para su reconocimiento semiautomático, se desarrolló un método basado en aprendizaje por transferencia y supervisado mediante una red neuronal convolucional VGG16 pre-entrenada, se establecieron tres experimentos con veinticinco pruebas cada uno, donde el mejor resultado obtiene una exactitud (accuracy) de 97,5% (pérdidas de 5,7%), superando en un 12,1% al método existente que cumple con los requisitos.

Palabras clave: Detección de estrés, red neuronal convolucional, VGG16, reconocimiento semi-automático, aprendizaje por transferencia y supervisado, exactitud (accuracy).

2.1. Introducción

Actualmente, la detección de estrés queda sujeta a especialistas del área de la salud, quienes aplican herramientas, procedimientos y técnicas que dado los tiempos que involucran, no permiten la toma de decisiones de forma oportuna, ya que el estrés se presenta en el corto plazo por lo que se requiere que su detección sea rápida y automática o semi-automática, abordando también errores debido a la subjetividad humana. En estudio realizado por SUSESO [4], cuyo objetivo fue encontrar los principales determinantes de los accidentes laborales sufridos por trabajadores cubiertos por la Ley 16.744/68 de Accidentes Laborales y Enfermedades Profesionales [1], utilizando una base de datos con las respuestas de un cuestionario aplicado en los años 2017 y 2018, determinaron que las variables más importantes para explicar los accidentes laborales están relacionadas con la auto percepción en la salud de los trabajadores como son la salud general, salud mental, vitalidad y, sobre todo, el estrés. SUSESO en reporte del año 2021, indica que los casos de salud mental siguen siendo importantes dentro de las enfermedades laborales representando el 35% del total [3].

Existen dos maneras semi-automáticas de abordar el reconocimiento de estrés en las personas, estas pueden ser con captura de señales fisiológicas mediante instrumentación para luego ser procesadas [5], y las que utilizan imágenes de rostros [6]. Por otra parte, existen las manuales donde la detección puede ser mediante cuestionarios, observaciones, introducción de reactivos y toma de exámenes para evaluar el comportamiento del cuerpo.

Los métodos de aprendizaje profundo utilizan múltiples niveles de representación, obtenidos por composición simple, a través de módulos no lineales para transformar la representación en un nivel con la entrada sin procesar hasta uno de representación superior ligeramente más abstracto. Las composiciones de varias transformaciones mediante funciones complejas permiten el aprendizaje. Las tareas de clasificación, son realizadas por capas superiores de representación, ampliando aspectos de la entrada que son importantes para la discriminación y supresión de las variaciones irrelevantes. El

aspecto clave del aprendizaje profundo es que estas capas de características no están diseñadas por humanos, aprenden de los datos mediante un procedimiento de aprendizaje de propósito general [9].

Lo relevante de las redes neuronales convolucionales es que los datos de entrada son imágenes, están constituidas por las capas de obtención de características (convolución), reducción y selección de características comunes (pooling), y clasificación de las entradas [11]. Estas redes aprenden en la instancia de entrenamiento y generalizan sus resultados en la validación, en la etapa de prueba se verifica el desempeño final, el comportamiento en sus formas dinámicas (sub ajustado, bien ajustado y sobre ajustado) se obtiene en mediante la métrica de pérdidas en entrenamiento y validación [12].

Para acelerar el tiempo de desarrollo es beneficioso el utilizar el aprendizaje por transferencia, el cual permite traspasar el conocimiento de un modelo otro basándose en pesos y sesgos actualizados obtenidos en cada capa. En definitiva, el usar módulos o partes de modelos ya desarrollados, aceleran el proceso de entrenamiento del modelo y mejoran los resultados [15].

El aprendizaje por transferencia y supervisado son las estrategias a utilizar, ya que el primero permite compartir los conocimientos adquiridos en generalidad (pesos, características, etc.) con el aprendizaje específico a obtener, optimizando todos los recursos involucrados y con resultados confiables.

El aporte de este trabajo es un método semi-automático que mediante una red neuronal convolucional (utilizando aprendizaje por transferencia y supervisado basado en red VGG16 pre-entrenada), permita la detección de estrés en las personas mediante imágenes de sus rostros. Este método se podría utilizar para la toma de decisiones en la asignación de personal o para la verificación de dicha condición de salud mental, considerando las características más relevantes de expresión del rostro para determinar dicha condición. Los resultados obtenidos en la métrica de exactitud (accuracy) fue de 97,5% (pérdidas de 5,7%), superando en un 12,1% al método existente que cumple con los requisitos.

Este trabajo se estructura según los siguientes ítems. En la sección 2.2 se desarrolla el marco teórico que sustenta la propuesta, en la sección 2.3 se describen los materiales y métodos utilizados. En la sección 2.4 los resultados obtenidos al aplicar el método propuesto, en la sección 2.5 se realiza una discusión de los resultados enfocado en el desempeño del modelo en base a sus curvas de aprendizaje y métricas. En la sección 2.6 las conclusiones del trabajo junto a observaciones respecto a futuras investigaciones aplicadas derivadas de los resultados obtenidos.

Finalmente se encuentran las referencias bibliográficas y anexo que menciona algunos instrumentos para evaluar estrés.

2.2. Marco teórico

Redes neuronales

Las redes neuronales (NN) son un modelo computacional que consta de un grupo interconectado de nodos (neuronas), cada uno de las cuales se encuentra conectado a muchos otros, intentando emular así el funcionamiento de una red neuronal biológica. Las NN son ampliamente utilizadas para resolver problemas de clasificación de datos.

La neurona es la estructura básica de una red neuronal. Cada neurona al recibir información, de una neurona anterior o del exterior, realiza una función de suma de todos los valores de entrada xi multiplicados por su peso wi. El resultado de este cálculo es ingresado a una función de activación, de modo que la señal debe sobrepasar un límite antes de propagarse [9]. En la Figura 1 se observa el funcionamiento básico de una neurona.

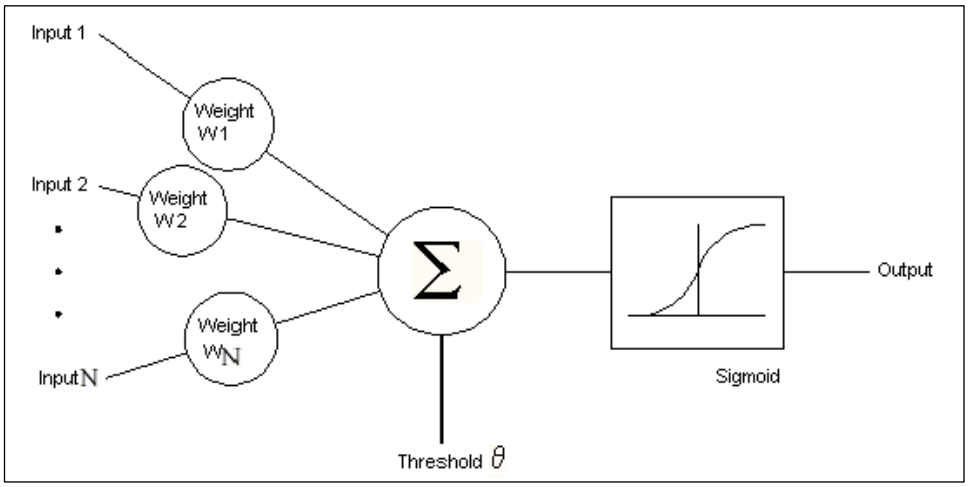


Figura 1: Funcionamiento básico de una neurona artificial.

Fuente: Deep learning [9].

Arquitectura de una red neuronal

La arquitectura básica de una red neuronal consta de tres capas. La primera de ellas es la capa de entrada (*Input Layer*), en la cual se recibe la información que se desea procesar y se realiza una primera evaluación de esta. Posteriormente, las neuronas de la capa de entrada envían su información a cada una de las neuronas de la segunda capa o también llamada capa oculta (*Hidden Layer*), de las cuales pueden existir más de una dependiendo de la complejidad del problema que se desea abordar. Finalmente, luego de la o las capas ocultas se encuentra una capa final (*Output Layer*), la cual entrega los resultados de la clasificación realizada por la red [9]. En la Figura 2 se muestra la arquitectura básica de una red neuronal totalmente conectada.

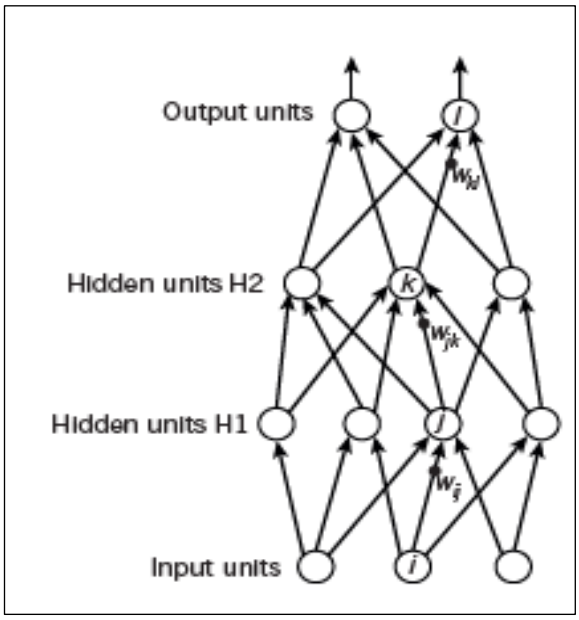


Figura 2: Arquitectura básica de una red neuronal conectada.

Fuente: Deep learning [9].

Función de activación

El cálculo realizado por cada neurona se puede escribir de manera abreviada como se muestra en Ecuación (1):

$$f(x_i, w_i) = \phi(\sum_{\mathrm{i}\epsilon\mathrm{I}}(x_i * w_i)) \qquad (1)$$

Donde ϕ representa la función de activación. Existen diversos tipos de funciones de activación [10], sin embargo, las más comunes son las que se indican en Tabla 2 correspondiendo a las Ecuaciones (2)-(3)-(4)-(5)-(6).

Tabla 2: Funciones de activación.

Lineal	$\phi(x) = \mathrm{X}$	(2)
Binaria	$\phi(x) = \begin{cases} x^2, & x < 0 \\ x^3, & x \geq 0 \end{cases}$	(3)
Sigmoidea	$\phi(x) = \frac{1}{1 + e^{-x}}$	(4)
Tangente hiperbólica	$\phi(x) = \tanh x$	(5)
Rectificación de unidad lineal (ReLU)	$\phi(x) = \max(0, x)$	(6)

Fuente: Pattern Recognition and Machine Learning [10].

Aprendizaje supervisado de una red neuronal

La etapa de aprendizaje de una red neuronal utiliza el método de *backpropagation*. En este se ingresa un conjunto de datos (*input*) categorizados según su clase. De esta manera, este *input* se propaga desde la primera capa hasta la capa final, pasando por cada una de las capas ocultas. El *output* obtenido por la capa final es comparado con la salida deseada (definida por las clases), obteniendo así un error para cada una de las salidas, el cual se puede expresar como se indica en Ecuación (7).

$$J(w) = -\sum_{\mathrm{i}} (y_i log\ (p_i)) \qquad (7)$$

Donde y_i corresponde al *output* i-ésimo deseado, p_i a la predicción i-ésima y la función (w) se representa sobre un espacio multidimensional de dimensión igual al número de pesos de la red [11].

Posteriormente, el error es propagado hacia atrás (*backpropagation*), partiendo desde la capa de salida, pasando por cada una de las capas ocultas hasta llegar a la capa de entrada. La búsqueda del conjunto de pesos que minimiza el error $J(w)$ se realiza mediante la Ecuación (8).

$$\Delta w_j = -\alpha \frac{\partial \mathrm{J}(w)}{\partial w_j} \qquad (8)$$

Con w el peso sináptico w_{ij} que define la fuerza de una conexión sináptica entre dos neuronas (la neurona presináptica i y la neurona postsináptica j) que se busca optimizar, α el hiper parámetro de la red correspondiente a la velocidad de aprendizaje.

Al realizar esto, cada una de las capas intermedias recibe solo una porción del error total, el cual corresponde a la contribución relativa de la neurona correspondiente.

Este proceso es repetido hasta que el peso de cada neurona corresponde al peso que produce el menor error de salida, el cual es obtenido mediante la optimización de la función de error.

Redes neuronales convolucionales

Las redes neuronales convolucionales (CNN) son muy similares a las redes neuronales convencionales (NN). Éstas están compuestas por neuronas, cada una de las cuales posee pesos, y es activada mediante una función de activación, además de poseer una función de error a la salida de la última capa.

La principal diferencia de este tipo de red neuronal es que asume explícitamente que los datos de entrada (*input*) son imágenes, lo que permite codificar ciertas propiedades en la arquitectura; permitiendo mejorar en eficiencia y reducir la cantidad de parámetros en la red.

Generalmente, las CNN están construidas mediante tres diferentes tipos de capas, las cuales son:

- Capa convolucional: Obtiene un mapeo de las características de la imagen.
- Capa de *pooling*: Reduce la dimensionalidad al seleccionar las características más comunes.
- Capa clasificadora: Actúa de manera idéntica a la de una red neuronal convencional, clasificando así los *inputs* entregados.

Capa convolucional

En esta primera fase, se realiza una convolución entre la imagen de entrada (*input*) y una función de filtro (también llamado *kernel*), la cual busca extraer características específicas de la imagen. De esta manera se logra filtrar la imagen con un filtro previamente entrenado. Esta técnica permite que ciertas características se vuelvan más dominantes en la imagen de salida, debido a que poseen un peso más elevado en los pixeles que los representan.

La convolución realizada entre las funciones mencionadas es abordada numéricamente como un producto matricial, lo cual permite trabajar con entradas de tamaño variable [11]. En la Figura 3 se muestra el funcionamiento de una capa convolucional.

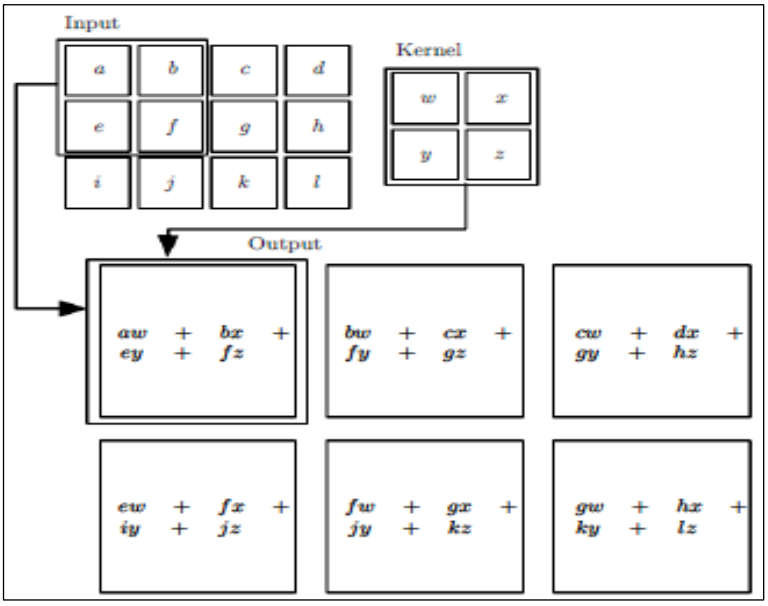


Figura 3: Convolución para una imagen 2D.

Fuente: Deep learning [11].

Capa de pooling

Generalmente la capa de *pooling* se encuentra después de la capa de convolución, y la función de ésta es realizar una reducción de muestreo, lo cual implica paralelamente una pérdida de información. Sin embargo, está perdida de información es beneficiosa para la red, debido a que conduce a una menor sobrecarga de cálculo para las capas sucesoras de la red.

Usualmente se suele utilizar la función *max-pooling* en esta capa. Esta función encuentra el valor máximo entre una ventana de muestra, guardando solo este valor para la siguiente capa [11]. En la Figura 4 se muestra gráficamente la operación que realiza la función *max-pooling*.

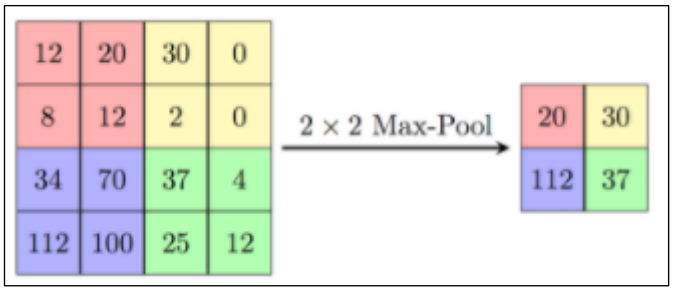


Figura 4: Diagrama de la función max-pooling.
Fuente: Deep learning [11].

Capa clasificadora

Luego de las capas de convolución y de *pooling*, se utilizan capas de neuronas completamente conectadas, en las que cada pixel de la imagen asocia una o varias neuronas. Esta capa funciona de manera idéntica a la de una red neuronal convencional (NN), sin embargo, gracias a la reducción de dimensionalidad y extracción de características predominantes realizadas por las dos capas previas, esta capa es capaz de realizar una clasificación más exacta.

Entrenamiento de una red neuronal

Para realizar el entrenamiento de una red neuronal en primer lugar se debe separar el conjunto de datos a utilizar en dos grupos, los cuales son entrenamiento (*train*) y prueba (*test*). Los datos de entrenamiento son utilizados para lograr que la red aprenda a clasificar la tarea deseada, lo cual es realizado mediante la asignación de pesos (w_k) a cada una de las neuronas que componen la red mediante la técnica de *backpropagation*. De esta manera la red en primera instancia aprende a clasificar al conjunto de datos de entrenamiento. Posteriormente se debe alimentar la red con el conjunto de datos de prueba, los cuales son desconocidos por el grafo, de modo de evaluar el desempeño presentado por el algoritmo.

Un problema recurrente al realizar el entrenamiento de una red es el sobreajuste (*overfitting*) de esta a los datos de entrenamiento, es decir, la red aprende a clasificar de manera acertada los datos con los que fue entrenada, sin embargo, pierde generalidad a la hora de evaluar datos desconocidos (datos de prueba). La consecuencia más obvia del sobreajuste es el desempeño deficiente presentado por la red (lo que equivale a un mayor error para los datos de prueba o una menor exactitud de clasificación). Sin embargo, este problema también implica un mayor requerimiento de información innecesaria sobre cada dato, lo cual se traduce en un mayor costo y tiempo de entrenamiento [12].

Para evitar este problema existen principalmente dos métodos, *Dropout* y regularización L2.

Dropout

Esta técnica consiste en “apagar” cada uno de los nodos de la red de forma individual en cada una de las etapas de entrenamiento con probabilidad $1-p$ (con p la probabilidad de mantener encendido el nodo). De esta manera, en cada etapa del entrenamiento, la red cuenta con una cantidad reducida de nodos (determinada por p). Al entrenar la red en cada iteración con solo una fracción de los nodos totales, la red se ve forzada a generar una forma más robusta de realizar la clasificación, logrando una mayor generalización para datos nuevos, y así evitando el sobreajuste [13].

Cuando se evalúa el desempeño de la red con los datos de prueba no se aplica *Dropout* a los nodos, buscando obtener así el máximo desempeño de la red. El valor estándar de p que suele utilizarse es de 0.5, sin embargo, este puede variar. En la Figura 5 se observa gráficamente la consecuencia del *Dropout* en una red neuronal artificial.

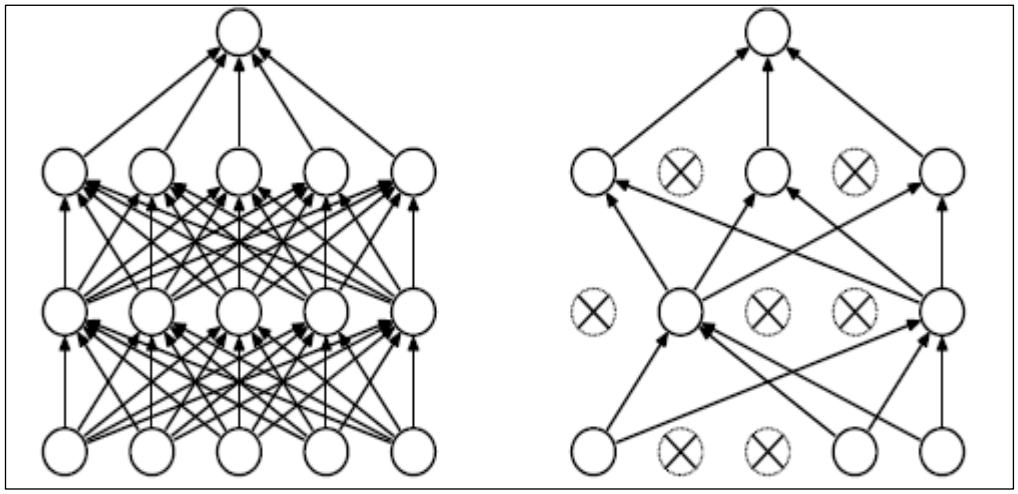


Figura 5: Izquierda: Red neuronal estándar. Derecha: Red neuronal convencional (NN) luego de aplicar Dropout.

Fuente: Dropout: A Simple Way to Prevent Neural Networks from Overfitting [13].

Regularización L2

La regularización L2 busca evitar el sobreajuste de la red incorporando una penalización a la función de error que busca optimizar el algoritmo, de esta manera se previene que pesos individuales puedan tener demasiada influencia sobre la clasificación final. La penalización impuesta corresponde a la sumatoria de los cuadrados de cada uno de los pesos de la red multiplicados por un factor β de ajuste, el cual suele tener un orden entre 0,1 y 0,00001. Matemáticamente la regularización puede escribirse como se ve en Ecuación (9) [14].

$$J'(w) = J(w) + \beta \sum_{j=i}^{n} w_j^2 \qquad (9)$$

Con $J'(w)$ la nueva función de costo, $J(w)$ la función de costo original, β el factor de ajuste de la regularización y w los pesos de cada una de las neuronas de la red.

Aprendizaje por transferencia

Los modelos de aprendizaje automático no están capacitados para funcionar transfiriendo el conocimiento de un modelo a otro. El conocimiento se basa en los pesos y sesgos actualizados obtenidos en cada capa. Entonces, el aprendizaje por transferencia tiene por objetivo el superar el aislamiento y utilizar el conocimiento que se obtiene de una tarea para resolver otras relacionadas. El principal beneficio de usar el aprendizaje por transferencia es que puede acelerar el tiempo para desarrollar y entrenar un modelo, al usar los módulos o partes de modelos ya desarrollados. Estos no solo aceleran el proceso de entrenamiento del modelo, sino que también mejoran los resultados.

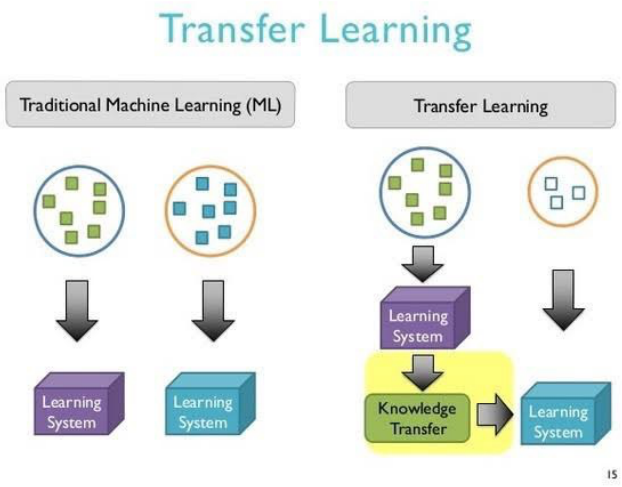


Figura 6: Diferencia entre aprendizaje automático y aprendizaje por transferencia. Fuente: A Quick Overview to the Transfer Learning and it's Significance in Real World Applications [15].

El aprendizaje de automático (ML) tradicional se denomina aislado porque el conocimiento no se acumula. Además, es un aprendizaje de una sola tarea porque el aprendizaje se realiza sin considerar el conocimiento pasado en ninguna otra. A partir de la Figura 6 [15], se infiere que los datos se pasan directamente al sistema de aprendizaje o los modelos de entrenamiento y cada modelo se entrena sin la transferencia de conocimiento de otro sistema de aprendizaje.

En el aprendizaje por transferencia, el concepto es completamente opuesto al del ML tradicional porque el aprendizaje de la nueva tarea depende de las tareas aprendidas previamente. Con este acto, la precisión de la salida será mucho mejor en comparación con su contraparte y se mejorará el proceso, el modelo que se entrena transfiere o comparte conocimientos con el otro sistema de aprendizaje [15], ver Figura 7.

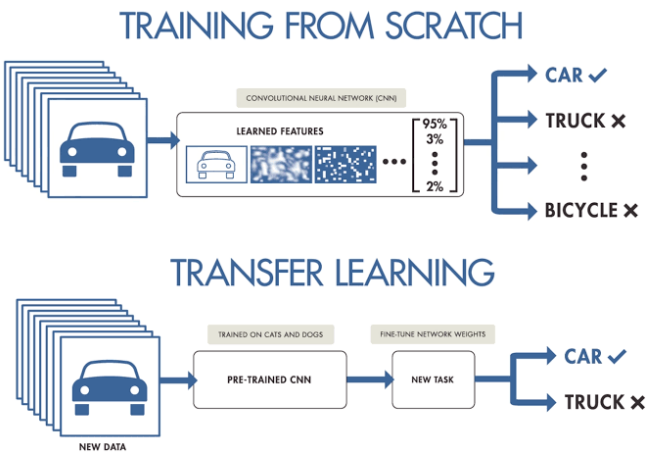


Figura 7: Funcionamiento de aprendizaje automático y aprendizaje por transferencia. Fuente: A Quick Overview to the Transfer Learning and it's Significance in Real World Applications [15].

2.3. Materiales y métodos

Para dar cumplimiento a los objetivos específicos planteados, es necesario establecer las condiciones, las cuales se especifican a continuación:

- Definir las situaciones en las cuales se encuentren expuestas las personas a condiciones de estrés y no estrés, pero de forma natural, con el fin de obtener imágenes de sus rostros, para ello se plantea capturar estas imágenes desde competencias de fútbol de clase mundial, específicamente en una situación altamente estresante para cualquier futbolista, que es patear un penal [16] y otras situaciones dentro del mismo encuentro consideradas como no estresantes.
- Definir desde la literatura y el marco teórico la utilización de redes neuronales convolucionales (CNN) en 3 dimensiones, con aprendizaje supervisado y por transferencia.

Materiales

Se dispone de tres bases de datos donde las dos primeras contienen 791 imágenes de rostros de futbolistas profesionales de diversos tipos de razas, con rango de edades entre 18 a 30 años, sin objetos en el rostro, ojos abiertos y cerrados, capturadas en competencias internacionales de futbol, de las cuales 560 corresponden a hombres (71%) y 231 a mujeres (29%). La tercera base de datos contiene 711 imágenes de rostros de futbolistas profesionales de diversos tipos de razas, con rango de edades entre 18 a 30 años, sin objetos en el rostro y ojos abiertos, capturadas en competencias internacionales de futbol, de las cuales 500 corresponden a hombres (70%) y 211 a mujeres (30%).

En las bases de datos 1 y 2, contienen 218 imágenes en situaciones de desempeños no estresantes dentro de un encuentro de futbol y 573 imágenes en desempeños estresantes como es la situación de patear un penal en un encuentro de futbol internacional [16]. La base de datos 3 contiene 218 imágenes en situaciones de desempeños no estresantes y 493 imágenes en desempeños estresantes. Las diferencias entre las bases de datos 1 y 2 es la distribución de las imágenes en sus categorías para entrenamiento, validación y pruebas. La diferencia con la base de datos 3, es que esta no contiene imágenes de rostros con ojos cerrados.

Otro asunto relevante son las especificaciones técnicas de las imágenes, las cuales se detallan en Tabla 3.

Tabla 3: Especificaciones técnicas de imágenes.

Ancho	Alto	Resolución horizontal	Resolución vertical	Profundidad	Extensión	Tipo
250 píxeles	250 píxeles	96 píxeles por pulgada	96 píxeles por pulgada	24 bits	JPG	Color RGB

Fuente: Elaboración propia.

Por otra parte, se deben distribuir las imágenes y clasificar en grupos para entrenamiento, validación y pruebas para el método a utilizar, para la base de datos 1 la distribución será en los porcentajes de 54%, 23% y 23%, respectivamente, ver Tabla 4.

Tabla 4: Distribución y clasificación de imágenes base de datos 1.

Condición	Entrenamiento	Validación	Pruebas	Total
Estresado	333	120	120	573
No estresado	94	62	62	218
Total	427	182	182	791

Fuente: Elaboración propia.

Para el caso de la base de datos 2, la distribución y clasificación de las imágenes en grupos para entrenamiento, validación y prueba será en los porcentajes de 60%, 20% y 20%, respectivamente, ver Tabla 5.

Tabla 5: Distribución y clasificación de imágenes base de datos 2.

Condición	Entrenamiento	Validación	Pruebas	Total
Estresado	381	96	96	573
No estresado	94	62	62	218
Total	475	158	158	791

Fuente: Elaboración propia.

Para la base de datos 3, la distribución y clasificación de las imágenes en grupos para entrenamiento, validación y prueba será en los porcentajes de 54%, 23% y 23%, respectivamente, ver Tabla 6.

Tabla 6: Distribución y clasificación de imágenes base de datos 3.

Condición	Entrenamiento	Validación	Pruebas	Total
Estresado	293	100	100	493
No estresado	94	62	62	218
Total	387	162	162	711

Fuente: Elaboración propia.

Es importante destacar que las imágenes a utilizar en el grupo de prueba deben ser distintas a los grupos de entrenamiento y validación, ya que se debe garantizar que el algoritmo no las haya visto antes, de manera de asegurar que lo que se detecte sean las características del estrés y no los rostros de los futbolistas, ver Figuras 8 y 9.

Figura 8: Imágenes clasificadas como estrés.

Fuente: Captura propia desde encuentros de futbol entre selecciones del mundo.

Figura 9: Imágenes clasificadas como no estrés.

Fuente: Captura propia desde encuentros de futbol entre selecciones del mundo.

Métodos

Se utilizará como método la CNN VGG16 [17], Visual Geometry Group, University of Oxford, la cual posee 13 capas convolucionales y 3 capas completamente conectadas, ver Figura 10 [18], pre-entrenada mediante BD ImageNet, la cual es parte de la librería de Keras.

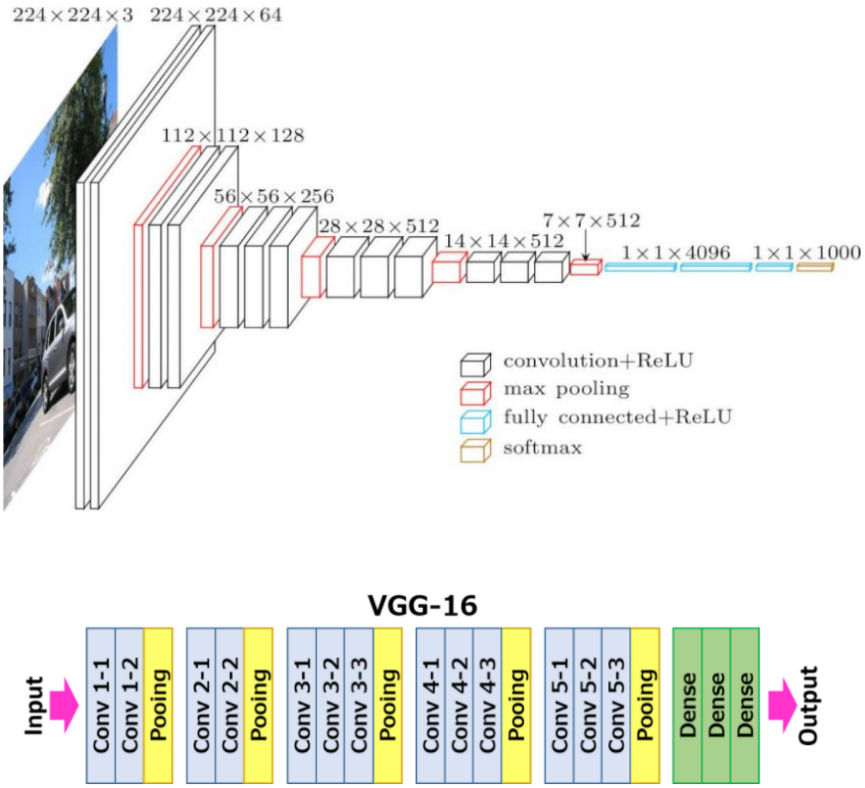


Figura 10: Estructura VGG16.

Fuente: VGG16 – Convolutional Network for Classification and Detection [18].

Para el desarrollo del método se utilizará un diagrama con las etapas genéricas para una CNN, la extracción de características desde las imágenes en la entrada se obtienen mediante las capas de convolución (con función de activación de rectificación de unidad lineal (ReLU)), agrupación/reducción (max-pooling) y una capa final clasificadora

completamente conectada (con función de activación sigmodea), la cual entrega la salida de una neurona binaria, donde es cero (0) cuando identifica estrés y uno (1) cuando identifica no estrés, ver Figura 11.

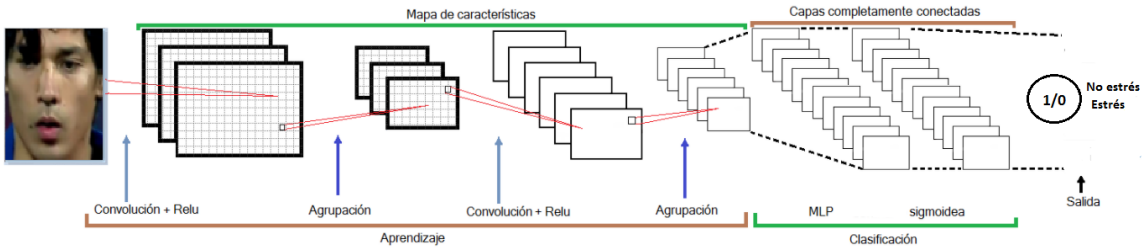


Figura 11: Diagrama genérico para el desarrollo del método.

Fuente: Adaptación propia - Capas CNN.

A continuación, se muestra el diagrama del método especifico utilizando la CNN VGG16:

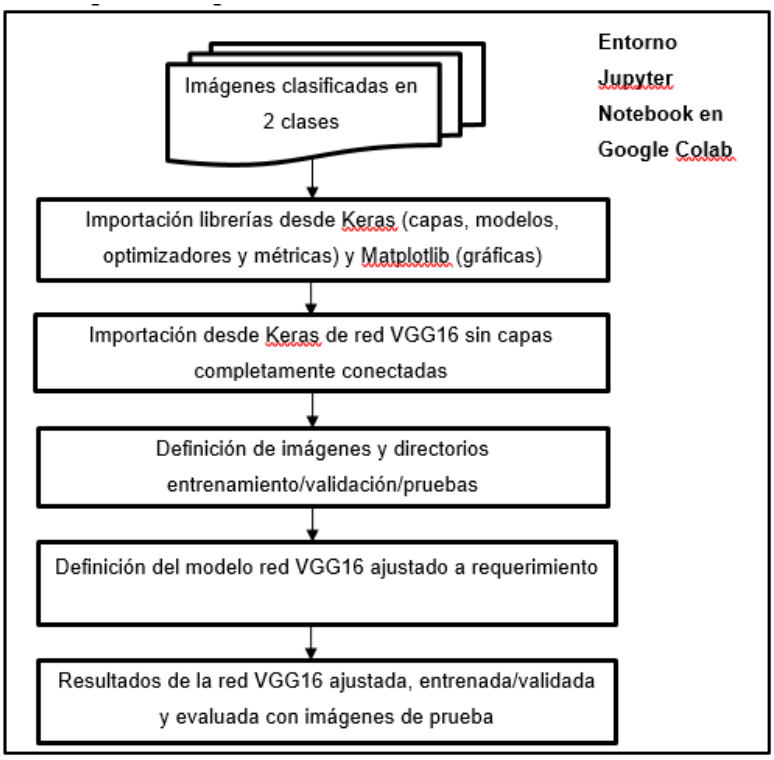


Figura 12: Diagrama desarrollo del método con red VGG16.

Fuente: Elaboración propia.

Con las imágenes de futbolistas y clasificadas de acuerdo a la condición de estrés y no estrés, se entrenará una CNN VGG16 con aprendizaje supervisado y por transferencia [19], con el fin de obtener con este método una exactitud (accuracy) de la predicción de estrés estimada igual o superior al 85%.

Las métricas de la CNN a utilizar, serán la exactitud (accuracy) y precisión, las cuales se determinan a partir de la matriz de confusión [20], que permite reconocer los aciertos y desaciertos del clasificador, los que a continuación se describen:

- Verdaderos positivos (VP): elementos clasificados como verdaderos y que realmente son verdaderos.
- Verdaderos negativos (VN): elementos clasificados como negativos y que realmente son negativos.
- Falsos positivos (FP): elemento clasificado como positivo y que realmente es negativo.
- Falsos negativos (FN): elemento clasificado como negativo y que realmente es positivo.

En la Tabla 7 se muestra la estructura de una matriz de confusión para una clasificación binaria.

Tabla 7: Matriz de confusión.

	Clasificados como positivos	**Clasificados como negativos**
Reales positivos	Verdaderos Positivos	Falsos Negativos
Reales negativos	Falsos Positivos	Verdaderos Negativos

Fuente: A systematic analysis of performance measures for classification tasks [20].

A partir de la matriz de confusión se pueden definir las siguientes métricas de desempeño:

- Precisión: Se define como la tasa entre los verdaderos positivos y el número de muestras predichas como positivas, y da cuenta de la coincidencia entre las etiquetas efectivas con las muestras positivas entregadas por el clasificador, ver Ecuación (10):

$$Precisión = \frac{VP}{VP+FP} \qquad (10)$$

- Exactitud (Accuracy): se define como la tasa de aciertos totales que obtiene un clasificador, y da cuenta de la eficacia general del clasificador, ver Ecuación (11):

$$Exactitud = \frac{VP+VN}{VP+VN+FP+FN} \quad (11)$$

A continuación, en la Tabla 8 se presenta la métrica para el objetivo específico definido.

Tabla 8: Matriz de métrica del objetivo específico medible.

Objetivo especifico	Métrica	Valor actual de la métrica	Criterio de éxito de la métrica
2	Matriz de confusión: exactitud (accuracy) de la predicción	85 % (obtenido desde el estado del arte)	≥ 85 %

Fuente: Elaboración propia.

Para el desarrollo del método se utilizan las siguientes herramientas:

- Biblioteca de código abierto Keras.
- Biblioteca de código abierto TensorFlow.
- Lenguaje de programación interpretado Python.
- Entorno de desarrollo en la nube Jupyter Notebook de Google Colab.
- Biblioteca para generación de gráficos Matplotlib.

2.4. Resultados

Definición de casos y procedimientos de pruebas

Para el método en el que se utiliza la CNN VGG16 pre-entrenada, los parámetros a modificar para encontrar la respuesta optima de la red ajustada son:

- Capas (layers): Corresponde a las últimas 3 capas de la red VGG16, solo en la penúltima capa (2) se puede indicar el número de neuronas a utilizar, ya que la capa 1 corresponde a la conversión de matriz a arreglo plano y la capa 3 corresponde a la salida binaria con una neurona.
- Épocas (epoch): Cantidad de veces o ciclo en que se ejecutaran las instrucciones con todo el conjunto de datos de entrenamiento.
- Tamaño del lote (batch size): Cantidad de imágenes por iteración de un ciclo. Su valor puede estar entre 1 hasta la cantidad de imágenes para entrenamiento.
- Pasos por épocas (steps per epoch): Es la cantidad de imágenes para el entrenamiento sobre el tamaño del lote.
- Pasos de validación (validation steps): Es la cantidad de imágenes para validación sobre el tamaño del lote.

La Tabla 9 muestra las métricas de pérdida (loss), exactitud (accuracy) y precisión obtenidas en las pruebas con la base de datos 1, luego del entrenamiento y validación de la red VGG16 pre-entrenada, considerando diferentes parámetros.

Tabla 9: Pruebas y resultados de métricas con base de datos 1.

Cantidad de imágenes para entrenamiento = 427								
Cantidad de imágenes para validación = 182								
Cantidad de imágenes para pruebas = 182								
Prueba	**Epoch**	**Batch size**	**Steps per epoch**	**Validation steps**	**Capa densa 2**	**Pérdida (loss) (%)**	**Exactitud (accuracy) (%)**	**Precisión (%)**
1	**36**	**10**	**43**	**19**	**32**	**6,5**	**96,7**	**95,2**
2	36	10	43	19	64	15,0	97,2	100
3	36	10	43	19	128	10,0	97,3	98,3
4	36	10	43	19	256	26,6	95,6	96,6
5	36	10	43	19	512	12,0	97,8	98,3
6	18	25	18	8	32	24,6	90,1	88,1
7	18	25	18	8	64	16,5	96,7	100
8	18	25	18	8	128	10,5	96,7	96,7
9	18	25	18	8	256	17,4	94,0	89,2
10	18	25	18	8	512	18,4	96,7	98,3
11	18	10	43	19	32	13,6	95,1	94,9
12	18	10	43	19	64	21,8	95,1	96,5
13	18	10	43	19	128	14,6	96,2	92,3
14	18	10	43	19	256	20,6	95,1	89,6
15	18	10	43	19	512	13,3	96,7	98,3
16	10	25	18	8	32	13,0	95,1	88,4
17	10	25	18	8	64	28,9	93,4	90,3
18	10	25	18	8	128	16,9	95,6	96,6
19	10	25	18	8	256	37,5	89,6	82,1
20	10	25	18	8	512	24,9	95,1	98,2
21	10	10	43	19	32	43,8	91,8	92,7
22	10	10	43	19	64	25,6	90,7	80,8
23	10	10	43	19	128	28,8	87,9	78,6
24	10	10	43	19	256	13,3	95,6	96,6
25	10	10	43	19	512	17,0	96,7	98,3

Fuente: Elaboración propia.

A continuación, se muestran los resultados de la prueba 1 contenidos en Tabla 9 para la primera y última época en entrenamiento y validación, así como su evaluación con imágenes de pruebas, ver Figuras 13, 14 y 15.

Epoch 1/36

43/43 - 6s 102ms/step - loss: 1.9036 - accuracy: 0.7025 - precision: 0.3575 - val_loss: 1.1759 - val_accuracy: 0.7418 - val_precision: 0.8571

Epoch 36/36

43/43 - 4s 86ms/step - loss: 0.0055 - accuracy: 0.9993 - precision: 0.9969 - val_loss: 0.6205 - val_accuracy: 0.9451 - val_precision: 1.0000

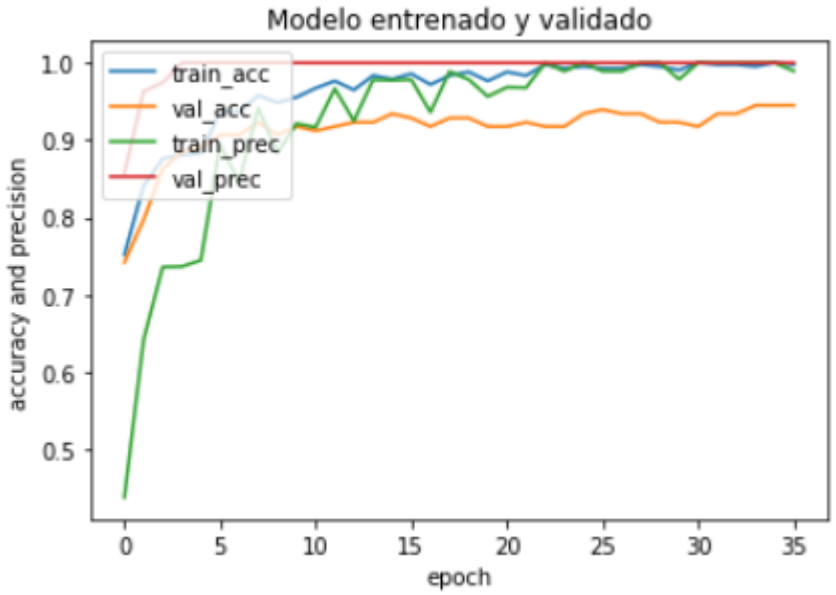


Figura 13: Resultados exactitud y precisión para entrenamiento y validación.
Fuente: Gráfica modelo propio en Google Colab.

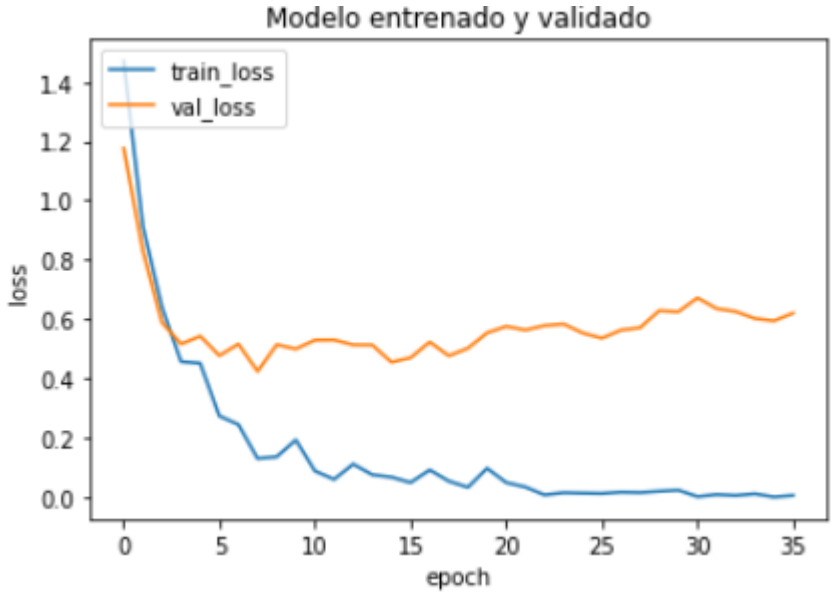


Figura 14: Resultados de pérdidas en entrenamiento y validación.

Fuente: Gráfica modelo propio en Google Colab.

Resultados con imágenes para pruebas:

19/19 - 2s 108ms/step - loss: 0.0652 - accuracy: 0.9672 - precision: 0.9524

test loss, test acc, test prec: [0.065212075710296631, 0.9672527444362640, 0.9523809552192688]

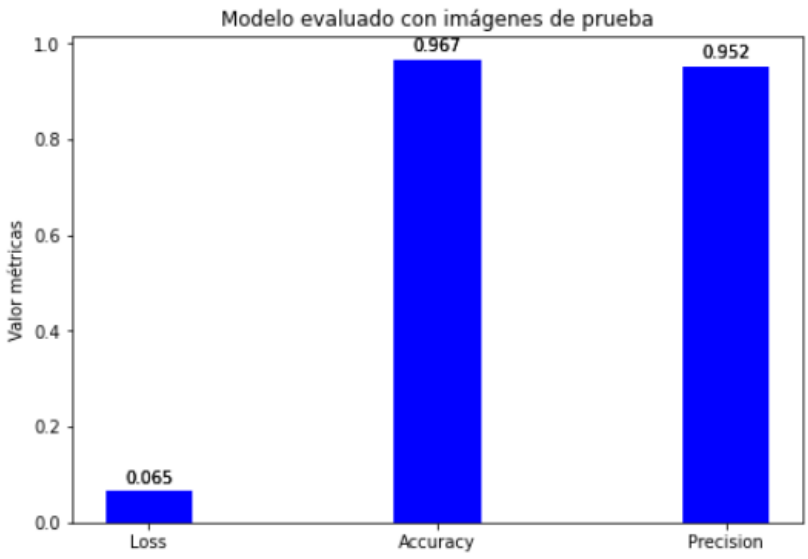


Figura 15: Resultados de pérdida, exactitud y precisión para pruebas.

Fuente: Gráfica modelo propio en Google Colab.

La Tabla 10 muestra las métricas de pérdida (loss), exactitud (accuracy) y precisión obtenidas en las pruebas con la base de datos 2, luego del entrenamiento y validación de la red VGG16 pre-entrenada, considerando diferentes parámetros.

Tabla 10: Pruebas y resultados de métricas con base de datos 2.

Cantidad de imágenes para entrenamiento = 475								
Cantidad de imágenes para validación = 158								
Cantidad de imágenes para pruebas = 158								
Prueba	**Epoch**	**Batch size**	**Steps per epoch**	**Validation steps**	**Capa densa 2**	**Pérdida (loss) (%)**	**Exactitud (accuracy) (%)**	**Precisión (%)**
1	36	10	48	16	32	26,9	94,3	100
2	36	10	48	16	64	16,3	96,2	98,3
3	36	10	48	16	128	20,8	96,8	98,3
4	36	10	48	16	256	11,8	97,5	98,3
5	36	10	48	16	512	22,5	97,5	100
6	18	25	19	7	32	11,3	94,3	98,2
7	18	25	19	7	64	22,0	91,1	88,7
8	18	25	19	7	128	40,4	92,4	98,1
9	18	25	19	7	256	16,4	96,2	96,7
10	18	25	19	7	512	41,6	93,7	100
11	18	10	48	16	32	18,4	94,3	98,2
12	18	10	48	16	64	18,4	96,2	98,3
13	18	10	48	16	128	17,3	95,6	98,2
14	18	10	48	16	256	24,7	95,6	98,2
15	**18**	**10**	**48**	**16**	**512**	**9,4**	**97,5**	**98,3**
16	10	25	19	7	32	36,2	89,2	97,9
17	10	25	19	7	64	29,1	93,7	98,1
18	10	25	19	7	128	29,0	91,8	86,6
19	10	25	19	7	256	11,9	95,6	93,7
20	10	25	19	7	512	15,1	95,6	98,3
21	10	10	48	16	32	27,2	95,6	100
22	10	10	48	16	64	23,5	96,2	98,3
23	10	10	48	16	128	15,5	95,6	100
24	10	10	48	16	256	17,7	96,2	100
25	10	10	48	16	512	10,9	98,1	98,4

Fuente: Elaboración propia.

A continuación, se muestran los resultados de la prueba 15 contenidos en Tabla 10 para la primera y última época en entrenamiento y validación, así como su evaluación con imágenes de pruebas, ver Figuras 16, 17 y 18.

Epoch 1/18

48/48 - 7s 99ms/step - loss: 1.3197 - accuracy: 0.8236 - precision: 0.6059 - val_loss: 0.4624 - val_accuracy: 0.9114 - val_precision: 0.9800

Epoch 18/18

48/48 - 4s 81ms/step - loss: 0.0020 - accuracy: 0.9976 - precision: 0.9870 - val_loss: 0.7602 - val_accuracy: 0.9367 - val_precision: 1.0000

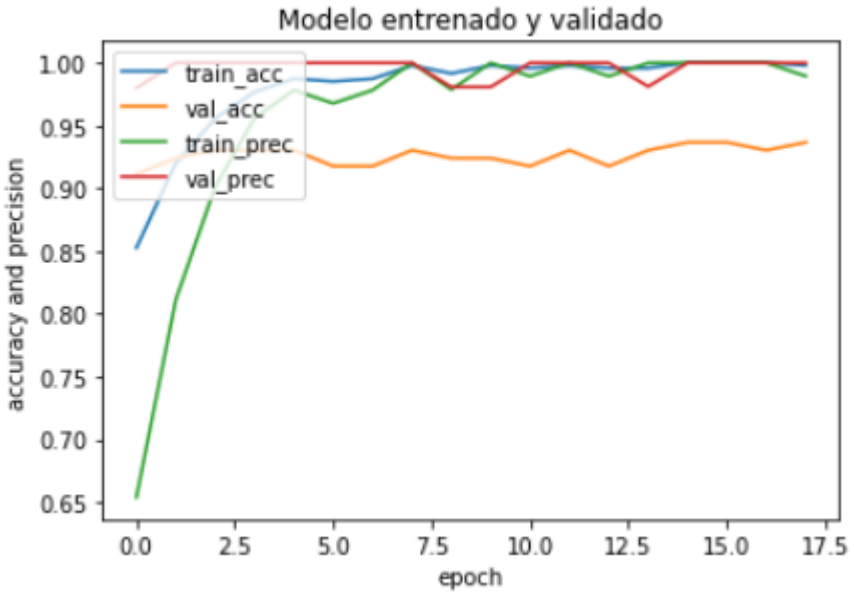


Figura 16: Resultados exactitud y precisión para entrenamiento y validación.
Fuente: Gráfica modelo propio en Google Colab.

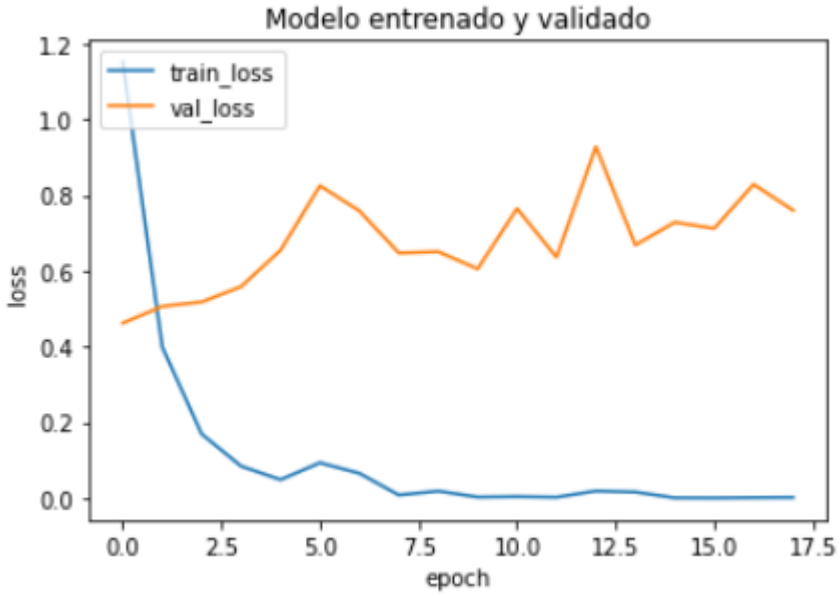


Figura 17: Resultados de pérdidas en entrenamiento y validación.

Fuente: Gráfica modelo propio en Google Colab.

Resultados con imágenes para pruebas:

16/16 - 1s 61ms/step - loss: 0.0940 - accuracy: 0.9747 - precision: 0.9833

test loss, test acc, test prec: [0.09395404160022736, 0.9746835231781006, 0.9833333492279053]

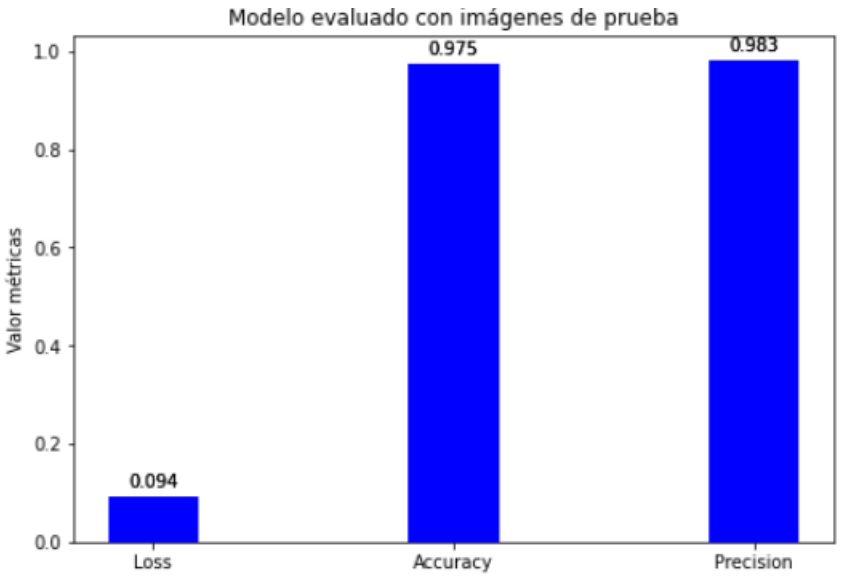


Figura 18: Resultados de pérdida, exactitud y precisión para pruebas.

Fuente: Gráfica modelo propio en Google Colab.

La Tabla 11 muestra las métricas de pérdida (loss), exactitud (accuracy) y precisión obtenidas en las pruebas con la base de datos 3, luego del entrenamiento y validación de la red VGG16 pre-entrenada, considerando diferentes parámetros.

Tabla 11: Pruebas y resultados de métricas con base de datos 3.

Cantidad de imágenes para entrenamiento = 387								
Cantidad de imágenes para validación = 162								
Cantidad de imágenes para pruebas = 162								
Prueba	**Epoch**	**Batch size**	**Steps per epoch**	**Validation steps**	**Capa densa 2**	**Pérdida (loss) (%)**	**Exactitud (accuracy) (%)**	**Precisión (%)**
1	**36**	**10**	**39**	**17**	**32**	**5,7**	**97,5**	**98,3**
2	36	10	39	17	64	15,7	96,9	96,7
3	36	10	39	17	128	12,0	96,3	96,7
4	36	10	39	17	256	9,3	95,7	93,7
5	36	10	39	17	512	9,1	97,5	98,3
6	18	25	16	7	32	21,5	93,8	96,4
7	18	25	16	7	64	23,2	96,9	98,3
8	18	25	16	7	128	25,8	93,2	91,8
9	18	25	16	7	256	25,3	93,8	98,2
10	18	25	16	7	512	27,7	93,2	94,7
11	18	10	39	17	32	18,4	94,4	98,2
12	18	10	39	17	64	14,9	95,1	98,2
13	18	10	39	17	128	16,8	96,3	100
14	18	10	39	17	256	32,8	94,4	100
15	18	10	39	17	512	22,7	93,8	96,4
16	10	25	16	7	32	37,1	91,4	96,2
17	10	25	16	7	64	25,1	93,2	100
18	10	25	16	7	128	31,2	93,8	93,3
19	10	25	16	7	256	41,1	85,2	72,1
20	10	25	16	7	512	35,1	95,7	100
21	10	10	39	17	32	18,5	92,0	85,5
22	10	10	39	17	64	19,8	93,8	98,2
23	10	10	39	17	128	24,8	95,1	98,2
24	10	10	39	17	256	38,4	93,2	96,4
25	10	10	39	17	512	34,4	93,2	100

Fuente: Elaboración propia.

A continuación, se muestran los resultados de la prueba 1 contenidos en Tabla 11 para la primera y última época en entrenamiento y validación, así como su evaluación con imágenes de pruebas, ver Figuras 19, 20 y 21.

Epoch 1/36

39/39 - 53s 170ms/step - loss: 2.4465 - accuracy: 0.5980 - precision: 0.3094 - val_loss: 0.5949 - val_accuracy: 0.8272 - val_precision: 1.0000

Epoch 36/36

39/39 - 3s 78ms/step - loss: 4.2529e-04 - accuracy: 1.0000 - precision: 0.9750 - val_loss: 0.5913 - val_accuracy: 0.9321 - val_precision: 1.0000

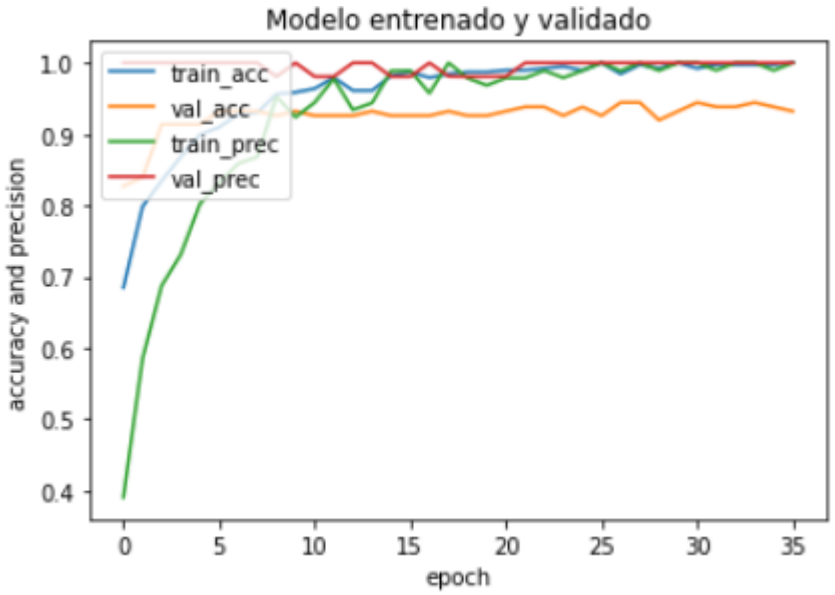


Figura 19: Resultados exactitud y precisión para entrenamiento y validación.
Fuente: Gráfica modelo propio en Google Colab.

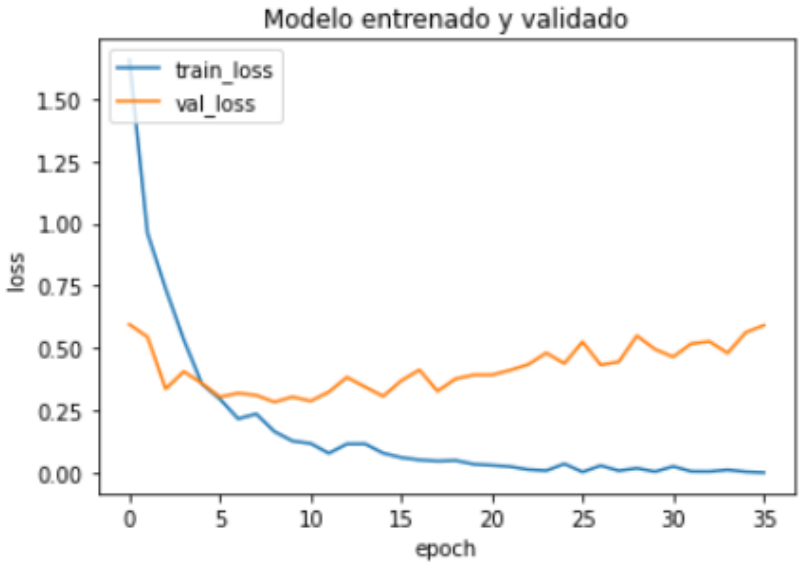


Figura 20: Resultados de pérdidas en entrenamiento y validación.

Fuente: Gráfica modelo propio en Google Colab.

Resultados con imágenes para pruebas:

17/17 - 1s 54ms/step - loss: 0.0566 - accuracy: 0.9753 - precision: 0.9833

test loss, test acc, test prec: [0.05662358179688454, 0.9753086566925049, 0.9833333492279053]

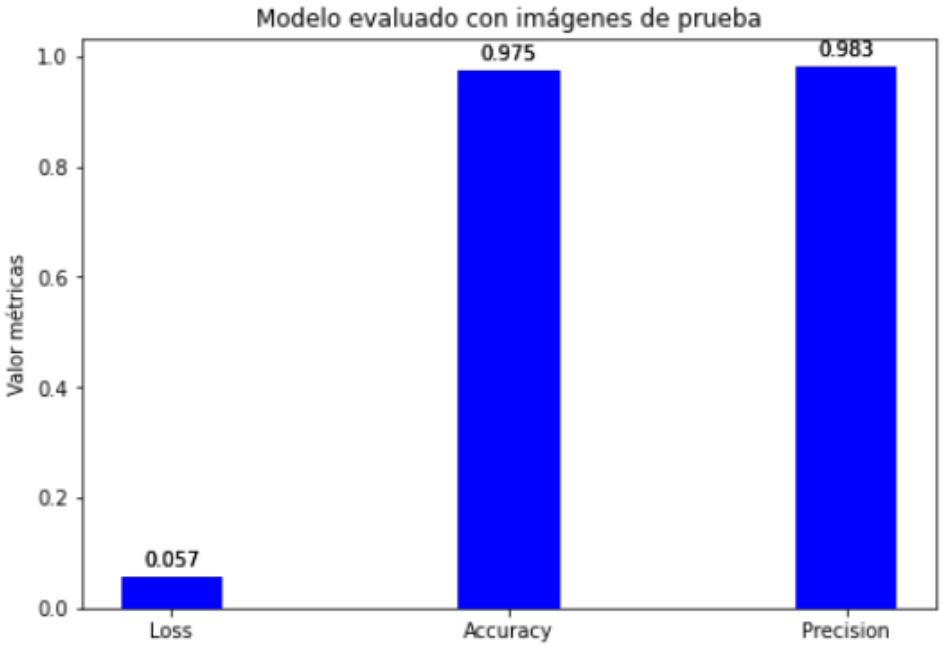


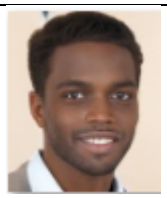
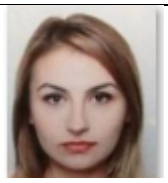
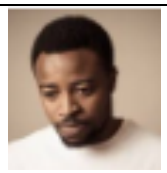
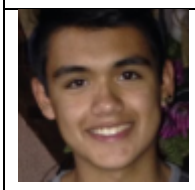
Figura 21: Resultados de pérdida, exactitud y precisión para pruebas.

Fuente: Gráfica modelo propio en Google Colab.

Predicción de clases

La propuesta considera un predictor de clases (estrés y no estrés) para verificar la respuesta del método, que consiste en someter al modelo a imágenes de rostros diferentes a las de las bases de datos en prueba, las cuales deben cumplir las características de estar en rango de edad, sin objetos en el rostro y con ojos abiertos, ver Tabla 12.

Tabla 12: Resultados de la predicción.

Imagen	Condición	Predicción
	Imagen con rango de edad, sin objetos en el rostro y con ojos abiertos.	No estrés
	Imagen con rango de edad, sin objetos en el rostro y con ojos abiertos.	Estrés
	Imagen con rango de edad, sin objetos en el rostro y con ojos abiertos.	No estrés
	Imagen con rango de edad, sin objetos en el rostro y con ojos abiertos.	Estrés
	Imagen con rango de edad, sin objetos en el rostro y con ojos abiertos.	No estrés
	Imagen con rango de edad, sin objetos en el rostro y con ojos abiertos.	No estrés
	Imagen con rango de edad, sin objetos en el rostro y con ojos abiertos.	Estrés

Fuente: Elaboración propia.

2.5. Discusión de los resultados

De los resultados obtenidos en las 25 pruebas realizadas al modelo y registrados en las Tablas 9, 10 y 11 con las bases de datos 1, 2 y 3, respectivamente, se considera solo una, la cual tiene como condición los valores de las métricas que van desde el valor más alto de exactitud en conjunto con el menor valor de las pérdidas.

El modelo encuentra su óptimo con los parámetros definidos en la prueba 1, ver Tabla 11, que tiene relación con la cantidad de épocas (36) y el número de neuronas de la capa 2 (32). En la Figura 19 se aprecian las curvas de aprendizaje del modelo para las métricas de exactitud y precisión con las imágenes para entrenamiento y la validación. Para el caso de la exactitud esta alcanza valores cercanos al 100% en el entrenamiento al completar los ciclos (épocas), lo cual se traduce en que tan bien el modelo está aprendiendo. En el caso de la validación los valores de exactitud se encuentran cercanos al 93% e indican como el modelo está generalizando al finalizar los ciclos. Los valores definitivos de las métricas se obtienen al evaluar el modelo con las imágenes de prueba, obteniendo una exactitud de 97,5%, ver Figura 21.

Para determinar el comportamiento del modelo en sus formas dinámicas (sub ajustado, bien ajustado y sobre ajustado), se utilizará la curva de aprendizaje que entrega la métrica de minimización, que en este caso son las pérdidas (loss). Se observa en la Figura 20 que las pérdidas disminuyen hasta un punto de estabilidad, pero existe una brecha de generalización entre las curvas de aprendizaje de pérdidas en entrenamiento y validación, que da cuenta probablemente de un conjunto pequeño de imágenes para validación. El valor definitivo de la pérdida se obtiene al evaluar el modelo con las imágenes de prueba, obteniendo una pérdida de 5,7%, ver Figura 21.

Es relevante indicar que, al verificar la respuesta del modelo esta se debe realizar con imágenes diferentes a las utilizadas en el entrenamiento, validación y pruebas, además de cumplir con las características establecidas de rango de edad, sin objetos en el rostro y con ojos abiertos, ya que su no cumplimiento produce errores en la predicción. En Tabla

13 se muestran los resultados comparativos de los métodos utilizados por otros autores y el propuesto, donde se puede apreciar que el método propuesto supera en un 12,1% de exactitud al método que cumple el requisito de no usar sensores adheridos al cuerpo y utilizar imágenes a nivel facial y a nivel de acción. El método propuesto es superado en un 2,3% de exactitud por el método que detecta señales fisiológicas del cuerpo, capturadas mediante instrumentación.

Tabla 13: Comparación de resultados entre métodos.

Método	**Autores**	**Instrumentación adherida al cuerpo**	**Resultados exactitud (accuracy)**
Detección de estrés mediante red neuronal convolucional unidimensional y red neuronal perceptrón multicapa, que utilizan señales fisiológicas del cuerpo (pulso del volumen de sangre, aceleración, actividad electrodérmica, ritmo cardiaco, temperatura de la piel, actividad eléctrica de los músculos esqueléticos, respiración).	Russell y Zhandong [5]	Si	99,8% CNN unidimensional 99,65 CNN perceptrón multicapa
Detección de estrés mediante red neuronal convolucional modificada con una arquitectura de memoria a corto y a largo plazo (LSTM) y la señal de un electrocardiograma (ECG).	Kang, Shin, Jung y Kim [6]	Si	98,3%
Detección de estrés de dos niveles basada en video, que integra un detector de nivel facial y un detector de nivel de acción para comprender las expresiones faciales y movimientos de acción.	Zhang, Feng, Li, Jin y Cao [7]	No	85,42%
Detección de estrés mediante red neuronal convolucional (VGG16 pre-entrenada) con aprendizaje supervisado y por transferencia, utilizando imágenes de rostros.	Muñoz, Cristian	No	97,5%

Fuente: Elaboración propia.

2.6. Conclusiones y trabajo futuro

El desafío de aplicar inteligencia artificial en la gestión de los recursos, ya sean estos económicos, humanos y tecnológicos, con el fin de tratar (eliminar o disminuir) los riesgos que los afectan, genera la necesidad de aprendizaje de los gestores y los obliga a explorar nuevos conocimientos en los dominios involucrados.

Esta investigación aplicada se desarrolla en dos ámbitos de los recursos, estos son humanos y tecnologías de la información, permite abordar los riesgos a que se ven afectadas las personas producto de los factores personales (estrés), y su reconocimiento oportuno mediante el aprendizaje profundo de las redes neuronales convolucionales, lo que permite identificar enfermedades laborales de salud mental y con ello evitar accidentes.

El desarrollo del método mediante un modelo de red neuronal (VGG16 pre-entrenada), el cual utiliza aprendizaje supervisado y por transferencia, permite una optimización de los recursos para lograr los objetivos propuestos y considerando la utilización de tres bases datos de imágenes reducidas en diferentes distribuciones. Es crucial que las imágenes se encuentren clasificadas con la condición a predecir y sus especificaciones de características (rango de edad, sin objetos en el rostro y con ojos abiertos), así como en una cantidad que logre un adecuado entrenamiento de la red con el fin de generalizar sus resultados. Lo anterior, permite que el modelo tenga un comportamiento ajustado en su forma dinámica.

El desempeño del método se probó mediante cambios en sus parámetros, dando cumplimiento a la métrica principal del objetivo (exactitud) y sus condiciones, las que se verifican mediante sus curvas de aprendizajes.

Como trabajo futuro se plantea implementar capas que permitan identificar en que partes de la imagen se está enfocando la CNN, además de desarrollar una aplicación en base al método para su uso en organizaciones productivas, donde el estrés es una condición de riesgo que puede provocar el fracaso de sus actividades. Con respecto a la base de

datos, se propone el construir una en colaboración con la Superintendencia de Seguridad Social de Chile (SUSESO) y las mutualidades del país, la cual incluiría imágenes de rostros de pacientes diagnosticados con estrés y no estrés, cuyos diagnósticos se basan en exámenes tradicionales realizados y aceptados por personal de salud y los organismos gestores y reguladores de enfermedades profesionales y de salud mental.

Finalmente es posible concluir que, el método propuesto permite reconocer la presencia de estrés y no estrés en una persona mediante imágenes de rostros y de manera oportuna.

REFERENCIAS

[1] Ministerio del Trabajo y Previsión Social, «Asociación Chilena de Seguridad,» 2021. [En línea]. Available: https://www.achs.cl/portal/trabajadores/Documents/Ley16.744_68.PDF. [Último acceso: 2021].

[2] Ministerio del Trabajo y Previsión Social, «Asociación Chilena de Seguridad,» 2021. [En línea]. Available: https://www.achs.cl/portal/trabajadores/Documents/ds-109.pdf. [Último acceso: 2021].

[3] P. Soto Altamirano, «SUSESO,» 2020. [En línea]. Available: https://www.suseso.cl/607/articles-632758_archivo_01.pdf. [Último acceso: 2021].

[4] Superintendencia de Seguridad Social, «SUSESO,» 2020. [En línea]. Available: https://www.suseso.cl/607/w3-article-617781.html. [Último acceso: 2021].

[5] L. Russell y L. Zhandong, «bmcmedinformdecismak,» 2020. [En línea]. Available: https://bmcmedinformdecismak.biomedcentral.com/track/pdf/10.1186/s12911-020-01299-4.pdf. [Último acceso: 2021].

[6] M. Kang, S. Shin, J. Jung y Y. T. Kim, «Hindawi,» 2021. [En línea]. Available: https://doi.org/10.1155/2021/9951905. [Último acceso: 2021].

[7] H. Zhang, L. Feng, N. Li, Z. Jin y L. Cao, «MDPI,» 2020. [En línea]. Available: https://doi.org/10.3390/s20195552. [Último acceso: 2021].

[8] R. Timarán Pereira, I. Hernández Arteaga, S. J. Caicedo Zambrano, A. Hidalgo Troya y J. Alvarado Pérez, «Ediciones UCC,» 2016. [En línea]. Available: http://dx.doi.org/10.16925/9789587600490. [Último acceso: 2021].

[9] G. Hinton, Y. LeCun y Y. Bengio, «Nature,» 2015. [En línea]. Available: https://s3.us-east-2.amazonaws.com/hkg-website-assets/static/pages/files/DeepLearning.pdf.

[10] C. Bishop, «Academia,» 2006. [En línea]. Available: http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf. [Último acceso: 2021].

[11] I. Goodfellow, Y. Bengio y A. Courville, «Deeplearningbook,» 2016. [En línea]. Available: https://www.deeplearningbook.org/contents/convnets.html. [Último acceso: 2021].

[12] D. Hawkins, «PSU,» 2003. [En línea]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.556.7571&rep=rep1&type=pdf. [Último acceso: 2021].

[13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever y R. Salakhutdinov, «Journal of Machine Learning Research,» 2014. [En línea]. Available: https://www.jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf?utm_campaign=buffer&utm_content=buffer79b43&utm_medium=social&utm_source=twitter.com. [Último acceso: 2021].

[14] E. Phaisangittisagul, «UKSIMS,» 2016. [En línea]. Available: https://uksim.info/isms2016/CD/data/0665a174.pdf. [Último acceso: 2021].

[15] K. Shah, «Towards Tech Intelligence,» 2019. [En línea]. Available: https://medium.com/towards-tech-intelligence/a-quick-overview-to-the-transfer-learning-and-its-significance-in-real-world-applications-790fb57debad. [Último acceso: 2021].

[16] P. Ureña, «Revista MHSalud,» 2006. [En línea]. Available: https://www.revistas.una.ac.cr/index.php/mhsalud/article/view/317/10905. [Último acceso: 2021].

[17] K. Simonyan y A. Zisserman, «Arxiv,» 2015. [En línea]. Available: https://arxiv.org/pdf/1409.1556.pdf%20http://arxiv.org/abs/1409.1556.pdf. [Último acceso: 2021].

[18] M. ul Hassan, «Neurohive,» 2018. [En línea]. Available: https://neurohive.io/en/popular-networks/vgg16/. [Último acceso: 2021].

[19] M. Rivera, «Cimat,» 2019. [En línea]. Available: http://personal.cimat.mx:8181/~mrivera/cursos/aprendizaje_profundo/preentrenadas/preentrenadas.html. [Último acceso: 2021].

[20] M. Sokolova y G. Lapalme, «Sciencedirect,» 2009. [En línea]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0306457309000259. [Último acceso: 2021].

Anexo 1: Instrumentos para evaluar estrés

A continuación, se mencionan algunos instrumentos para evaluar estrés:

- Escala Sintomática de Estrés (ESE). Explora síntomas de estrés con referencias subjetivas de la frecuencia con que se sienten los síntomas emocionales y psicofisiológicos relacionados con el estrés, cuenta con 18 ítems, dos categorías propuestas por Seppo Aro (1983).
- Inventario de Personalidad Resistente. Escala diseñada para evaluar la personalidad resistente, sigue el modelo propuesto por Kobasa (1979), operativizada en tres dimensiones: compromiso, reto y control, que se miden a través de 15 proposiciones en una escala Likert entre 0 y 3.
- Escala de modos de afrontamiento (EMA). De los autores Folkman y Lazarus (1984), permite a través de 67 ítems evaluar los estilos de afrontamiento.
- Valoración Resultados Estrategia de Confrontación de la Escala Modos de Afrontamiento. Del autor Dionisio Zaldívar Pérez (1990), cuenta con 8 subescalas.
- Test de Vulnerabilidad al Estrés. De los autores L. H. Miller y A. D. Smith, cuenta con 20 items para valorar el estrés tomando como punto de partida los estilos de conductas.
- Lista de Indicadores de Vulnerabilidad al Estrés. Adaptación cubana del autor Dionisio Zaldívar Pérez (1990), cuenta con 20 indicadores.
- Inventario para la Evaluación del Estrés Laboral *Burnout o Maslach Inventory Burnout* (MBI). De los autores C. Maslach y S. E. Jackson (1986), evalúa el estrés laboral a partir de tres escalas con un total de 22 items.
- Cuestionario Breve de *Bornout* (CBB). Es una adaptación española del MBI del autor Bernardo Moreno (2005) y colaboradores, cuenta con 21 items y una pregunta abierta.
- Escala General de Satisfacción. Adaptación realizada por los autores Pérez Bilbao y Fidalgo Vega (1979), explora la respuesta subjetiva o nivel de satisfacción de los

trabajadores, lo que permite indagar sobre cuáles son las causas que se ocultan tras el estrés laboral.

- Escala de Estrés Percibido (EEP). Mide la respuesta psicológica general frente a los estresores, de los autores Cohen, Kamarck y Mermelstein (1983), existen diferentes versiones, la original compuesta por 14 incisos (EEP-14), entre otras que son resultado de un proceso de refinamiento; en este último grupo, las dos más conocidas son una versión de 10 (EEP-10) y otra de 4 reactivos (EEP-4), con traducciones en diferentes idiomas, incluido el español.
- Cuestionario Sucesos de Vida. De los autores Lucio y Durán (2003), tiene como fin evaluar los sucesos de vida estresantes en adolescentes de 13 a 18 años de edad. Es un autoinforme de 129 reactivos y una pregunta abierta; permite evaluar siete áreas. Las respuestas deben corresponder a eventos ocurridos en un período no mayor a un año.