Universidad Autónoma de Querétaro

Facultad de Ingeniería

Maestría en Ciencias en
Inteligencia Artificial

Clasificación de estados de atención visual con métricas de seguimiento ocular mediante técnicas de aprendizaje profundo

TESIS

Que como parte de los requisitos para obtener el grado de
Maestra en Ciencias en Inteligencia Artificial

Presenta:

Alea Fernanda Bello Díaz

Dirigido por:

Dr. Marco Antonio Aceves Fernández

Co-dirigido por:

Dr. Sebastián Salazar Colores

Dr. Marco Antonio Aceves Fernández
Presidente
Dr. Sebastián Salazar Colores
Secretario
Dr. Jesús Carlos Pedraza Ortega
Vocal
Dr. Saúl Tovar Arriaga
Suplente
Dra. Danjela Ibrahimi
Suplente

Centro Universitario
Querétaro, QRO
México.
Julio 2023

A mi familia...

Agradecimientos

Me gustaría agradecer al Consejo Nacional de Ciencia y Tecnología (CONACYT) por brindar el apoyo económico que permitió el desarrollo de esta investigación. Adicional me gustaría agradecer a la Universidad Autónoma de Querétaro y a la dirección de Investigación y Posgrado de la Facultad de Ingeniería por brindar las herramientas y elementos necesarios a lo largo de los dos años del programa de estudios. Quisiera extender un agradecimiento a mis asesores que brindaron apoyo, consejo e ideas durante la realización de esta investigación.

Agradezco también a mi familia y amigos por su gran apoyo durante este paso más. A mis hermanos, Juan Pablo y Santiago quisiera agradecerles por siempre estar para mi y celebrar mis logros como suyos, siempre los llevo conmigo. A mi abuela, un ser humano maravilloso, por darme todas las herramientas, facilidades y su apoyo incondicional en cada paso profesional de mi vida.

A mi novio, gracias por todas tus palabras, desveladas, ideas y códigos durante estos dos años. Gracias por ser mi incondicional siempre, por siempre tener un consejo sabio y por nunca dejarme sola.

Abstract

Estimating attentional states is an essential task in different areas of knowledge. So far, the classification of attentional states was limited by the classes that the state of the art used to describe subjects' cognitive states. This contribution proposes a new way of processing data from eye-tracking sessions: converting one-dimensional information to a two-dimensional visualization. This approach provides the opportunity to visualize the physical properties of different eye movements that, during conventional processing may go unnoticed. One of the main problems of this research is the lack of public databases to work with, so it is proposed to create a database based on identifying the physical properties of each state of attention in the dataset. This identification allows the database to be labeled for the five states of care proposed by the clinical model of Solberg and Mateer. This information is subjected to an optical flow transformation for domain modification. Through domain shifting and deep learning techniques, specifically transfer learning dynamics, it is possible to successfully identify four attention levels by analyzing eye-tracking sessions.

Resumen

La estimación de estados de atención es una tarea importante en diferentes áreas del conocimiento. Hasta el momento la clasificación de estados de atención se veía limitada por las clases que el estado del arte ha utilizado para describir el estado cognitivo de los sujetos. En esta contribución se plantea una nueva forma de procesar los datos de sesiones de seguimiento ocular: convertir la información una dimensión a una visualización bidimensional. Esta propuesta brinda la oportunidad de visualizar propiedades físicas de los diferentes movimientos oculares que durante un procesamiento convencional pueden pasar desapercibidas. Uno de los problemas principales de esta investigación es la falta de bases de datos públicas con las que trabajar, por lo que se propone la creación de una base de datos basada en la identificación de propiedades físicas de cada estado de atención en el conjunto de datos. Esta identificación permite etiquetar la base de datos para los cinco estados de atención propuestos por el modelo clínico de Solberg y Mateer. Esta información es sometida a una transformación de flujo óptico con el propósito de modificar el dominio. A través de una combinación entre el cambio de dominio y técnicas de aprendizaje profundo, específicamente la dinámica de transferencia de aprendizaje, es posible identificar exitosamente cuatro estados de atención a través del análisis de sesiones de seguimiento ocular.

Índice general

Agradecimientos I

Abstract III

Resumen V

Índice VII

Índice de Figuras IX

Índice de Tablas XI

1. Introducción **1**
- 1.1. Motivación . . . 1
- 1.2. Planteamiento del Problema . . . 2
- 1.3. Justificación . . . 4
- 1.4. Objetivos . . . 5
  - 1.4.1. Objetivos Específicos . . . 5
- 1.5. Estructura de la Tesis . . . 5

2. Revisión de Literatura **7**
- 2.1. Estados de atención . . . 7
- 2.2. Movimientos oculares . . . 8
- 2.3. Estimación de estados de atención con EEG . . . 9
- 2.4. Estimación de estados de atención con movimientos oculares . . . 11

3. Metodología **13**
- 3.1. Inteligencia Artificial . . . 14
  - 3.1.1. Aprendizaje Automático . . . 15
  - 3.1.2. Aprendizaje Profundo . . . 15
  - 3.1.3. Aplicaciones de la atención . . . 18
- 3.2. Flujo Óptico . . . 19
- 3.3. Base de datos . . . 22

3.3.1. GazeCom . . . 23
3.3.2. Lund2013 . . . 24
3.3.3. 360EM . . . 25
3.4. Métricas de Seguimiento Ocular . . . 27
3.4.1. Fijaciones . . . 27
3.4.2. Movimientos sacádicos . . . 31
3.4.3. Seguimientos Suaves . . . 34
3.5. Etiquetado . . . 36
3.6. Preprocesamiento . . . 38
3.6.1. Etapa I: Limpieza e interpolación . . . 38
3.6.2. Etapa II: Transformación Inicial . . . 39
3.6.3. Etapa III: Cambio de dominio: Flujo Óptico . . . 40
3.7. Clasificación . . . 43
3.7.1. Arquitectura . . . 43
3.7.2. Entrenamiento del modelo . . . 43
3.8. Recursos de Hardware . . . 44
3.9. Métricas de Desempeño . . . 46

4. Resultados y Discusión **49**
4.1. Resultados . . . 49
4.1.1. Clasificación de Movimientos Oculares . . . 49
4.1.2. Clasificación de Niveles de Atención . . . 50
4.1.3. Estimación de Estados de Atención . . . 52
4.2. Discusión . . . 54
4.2.1. Clasificación de Movimientos Oculares . . . 54
4.2.2. Clasificación de Niveles de Atención . . . 57
4.2.3. Estimación de Estados de Atención . . . 60
4.3. Impacto . . . 63
4.3.1. Impacto Social . . . 63
4.4. Publicaciones . . . 65
4.5. Trabajo Futuro . . . 66

5. Conclusiones **67**

6. Anexos **69**
6.1. Artículo Presentado . . . 69
6.2. Constancias Manejo Lengua Extranjera . . . 71

Bibliografía **81**

Índice de figuras

2.1. Interpretación del ambiente a través de la atención visual [creación propia]. . 8
2.2. Técnica comúnmente utilizada para registrar movimientos oculares[creación propia]. 9

3.1. Metodología propuesta. 14
3.2. Ejemplo de secuencia del recorrido de una cuadrícula con un kernel [creación propia]. 16
3.3. Arquitectura ejemplo de una RNN sin capa de salida[creación propia]. 18
3.4. Comparativa y explicación a detalle de dos modelos de color: RGB y HSV. . 22
3.5. Ejemplos del escenario de pruebas utilizado durante la grabación de la sesión de eye-tracking [1]. 24
3.6. Escenarios empleados durante la sesión de eye-tracking de la base de datos Lund2013 [2]. 25
3.7. Ejemplos de escenarios empleados durante la sesión de eye-tracking de la base de datos 360EM [3]. 26
3.8. Etapa de limpieza e interpolación del archivo inicial de seguimiento ocular. . 39
3.9. Extracción de los diferentes eventos del movimiento ocular para la transformación inicial a una secuencia de vídeo. 40
3.10. El proceso de obtención del flujo óptico denso, (a) corresponde a la representación HSV y (b) a la representación en flechas. 42
3.11. Arquitectura del modelo EfficientNet-B0, inspirado en [4] 44
3.12. Entrenamiento del modelo en las diferentes etapas [creación propia]. 45

4.1. Primera etapa de entrenamiento del modelo: clasificación de eventos de movimientos oculares [creación propia]. 50
4.2. Segunda etapa de entrenamiento del modelo: clasificación de niveles de atención [creación propia]. 51
4.3. Tercera y última etapa de entrenamiento del modelo: estimación de estados de atención [creación propia]. 53
4.4. Resultados de las métricas de rendimiento en los diez ensayos ejecutados en la clasificación del conjunto de datos GazeCom. 56

4.5. Comportamiento de los cinco modelos para tres métricas: Accuracy, Precisión y Recall. 56
4.6. Comportamiento del modelo durante el entrenamiento para dos folds del primer experimento. 58
4.7. Comportamiento del modelo durante el entrenamiento en dos folds del primer experimento. 58
4.8. Análisis de los resultados alcanzados para la clasificación de niveles cognitivos. 60
4.9. Análisis de las métricas de la estimación de sostenida y dividida. 62
4.10. Análisis de las métricas de la estimación de atención enfocada y selectiva. . . 63

Índice de Tablas

2.1. Estado del arte. 12

3.1. Bases de datos que conforman la base de datos maestra. Donde *FX* se refieren a Fijaciones, *SC* Movimientos sacádicos, *PSO* Oscilaciones Posacádicas y *SP* Persecuciones suaves. 23
3.2. Propiedades físicas de los movimientos oculares. La primera fila de cada movimiento indica la duración, mientras que la segunda es la velocidad en g/s. . 36
3.3. Propiedades físicas de los movimientos oculares y naturaleza de estímulos en relación a los estados de atención. La primera fila de cada movimiento indica la duración, mientras que la segunda es la velocidad en g/s. 37
3.4. Relación entre el valor del Coeficiente Kappa y el nivel de acuerdo entre las muestras reales y predicciones. 48

4.1. Comparación del rendimiento de los métodos más avanzados. Los marcados con * corresponden a contribuciones con el base de datos GazeCom, y los marcados con + corresponden a contribuciones con el base de datos Lund2013. Las métricas en negrita corresponden a las puntuaciones más altas. 50
4.2. Métricas alcanzadas por el modelo durante cinco pruebas realizadas y el promedio de cada métrica. 52
4.3. Métricas alcanzadas por el modelo en la estimación del nivel de atención bajo: atención enfocada y atención selectiva, durante cinco pruebas realizadas. . . 53
4.4. Métricas alcanzadas por el modelo en la estimación del nivel de atención alto: atención sostenida y atención dividida, durante cinco pruebas realizadas. . . 54

Siglas

AA Aprendizaje Automático

AP Aprendizaje Profundo

CNN Redes Neuronales Convolucionales

EEG Electroencefalogramas

fMRI Imagen por Resonancia Magnética Funcional

FX Fijaciones

IA Inteligencia Artificial

LSTM Memoria a corto y largo plazo

OF Flujo Óptico

SC Movimientos Sacádicos

SP Seguimientos Suaves

SVM Máquina de Soporte de Vectores

TDAH Transtorno por Déficit de Atención e Hiperactividad

CAPÍTULO 1

Introducción

La estimación de estados de atención visual se ha convertido en un tema de gran importancia en la investigación en interacción humano-computadora. La técnica de seguimiento ocular proporciona información valiosa sobre la atención visual del usuario, pero su análisis y clasificación pueden ser desafiantes debido a la complejidad de los datos recopilados. Aquí es donde entran en juego las técnicas de Aprendizaje Profundo (AP). Los métodos de aprendizaje profundo son capaces de procesar grandes cantidades de datos y aprender patrones complejos, lo que los hace ideales para la clasificación de estados de atención visual.

Los sensores de seguimiento ocular recopilan información precisa sobre la ubicación y duración de las fijaciones oculares en un monitor de computadora. Al combinar esta información con técnicas de AP, se puede analizar la dinámica de los movimientos oculares y clasificar los estados de atención que caracterizan la actividad neurológica durante tareas cognitivas. En esta investigación, se analizará la importancia de la estimación de estados de atención visual y se explorará cómo el AP puede ser utilizado para mejorar la precisión y la fiabilidad de la estimación. Se discutirán diferentes enfoques y técnicas utilizadas en la literatura para abordar este problema, incluyendo técnicas basadas en AP y enfoques híbridos que combinan con otras técnicas. Además, se investigarán las implicaciones prácticas de la estimación de estados de atención visual en aplicaciones específicas relacionadas a campos específicos como la psicología y neurología. Finalmente, se discutirán las limitaciones actuales y las posibles áreas de investigación futura en este campo en constante evolución.

1.1. Motivación

Como los movimientos oculares están directamente relacionados con los mecanismos neuronales involucrados en la atención visual, el seguimiento ocular es una herramienta valiosa para estudiar el comportamiento de la atención [5]. Además, el análisis de los patrones de movimiento ocular también puede proporcionar información sobre trastornos neurológicos y otros procesos complejos relacionados con la atención[6]. Por lo tanto, es importante desarrollar sistemas robustos que puedan determinar con precisión el estado de atención del sujeto a partir de la información proporcionada por el seguimiento ocular.

El uso de información de seguimiento ocular no invasivo y aprendizaje profundo para estimar estados de atención es motivado por la necesidad de comprender y medir la concentración y la dirección de la mirada de una persona en tiempo real. La combinación de los datos de seguimiento ocular con técnicas de AP permite crear modelos precisos y eficientes para predecir la atención de una persona, lo que puede ser útil en una variedad de aplicaciones, como la investigación en neurociencia. Además, el uso de estas estrategias para procesar grandes cantidades de datos de seguimiento ocular permite una mayor precisión y velocidad en la estimación de estados de atención. Este enfoque no invasivo permite evitar la subjetividad de los diagnósticos tradicionales en la evaluación de la función cognitiva, especialmente relacionada con la atención y el aprendizaje. Además, la identificación precisa de los estados de atención relacionados con trastornos neurológicos brinda mayor confianza en el diagnóstico a los especialistas.

1.2. Planteamiento del Problema

La atención visual es un aspecto fundamental en la investigación en neurociencia y psicología. Sin embargo, los dos medios principales utilizados para estudiar la atención, como la resonancia magnética funcional y la potenciales evocados, son invasivos y limitados en su capacidad de proporcionar información precisa sobre el estado de atención del sujeto en situaciones reales. Cuando se observan diferentes síntomas relacionados con trastornos de atención, usualmente en el aula, se notifica a padres o tutores para que el sujeto sea sometido a una entrevista con un profesional. El llamado diagnostico tradicional consiste en la recopilación de información sobre los antecedentes médicos y sociales, y una evaluación en donde se analizan aspectos como la expresión emocional, el habla y funciones cognitivas, en donde destacan el nivel de alerta, espacio y tiempos, memoria, razonamiento y concentración o atención.

Es común debido a la naturaleza del análisis del sujeto, que el especialista emita un diagnóstico basado en la subjetividad y el padecimiento sufrido sea distinto al diagnosticado. La atención es un factor importante asociado con el efecto de aprendizaje [5]. A través de la evaluación de la atención, es posible especular que el sujeto tiene enfermedades relacionadas con la atención. Recientemente se han utilizado distintas fuentes de información para corroborar diagnósticos a través del análisis de atención en la evaluación mental, se utiliza Electroencefalogramas (EEG) para registrar y estudiar el comportamiento cerebral durante la ejecución de tareas, Imagen por Resonancia Magnética Funcional (fMRI) para diferenciar anatómicamente al sujeto de grupos de control y como análisis adicional el registro de movimientos oculares, pues existe un estrecho vínculo entre la dirección de la mirada humana y el foco de atención [7].

Actualmente, existen varias estrategias para la estimación de estados de atención, como la observación directa, la retroalimentación subjetiva y el seguimiento ocular. Sin embargo, estas estrategias presentan algunas desventajas importantes. La observación directa puede ser subjetiva y estar influenciada por la percepción individual del observador. La retroalimentación subjetiva depende de la capacidad del sujeto para autoevaluar su estado de atención,

lo que puede ser limitante en caso de trastornos neurológicos. El seguimiento ocular invasivo requiere el uso de electrodos en el ojo, lo que puede ser incómodo para el sujeto y limitar la aplicabilidad en entornos reales.

Otra estrategia para la estimación de estados de atención es el uso de EEG, que registra la actividad eléctrica del cerebro y es menos invasivo que el seguimiento ocular. El EEG se puede utilizar para identificar patrones de actividad cerebral asociados con diferentes estados de atención y para analizar la dinámica de la actividad cerebral durante tareas cognitivas. Sin embargo, la resolución temporal y espacial del EEG es limitada y puede no ser suficiente para capturar detalles finos en la dinámica de la actividad cerebral. Además, el EEG requiere la aplicación de electrodos en la cabeza, lo que puede ser incómodo y limitar su aplicación en entornos reales. Por lo tanto, aunque el EEG es una herramienta valiosa para la estimación de estados de atención, todavía presenta desafíos y limitaciones importantes que deben abordarse en futuros estudios.

El mayor problema de la clasificación con fuentes de información mencionadas anteriormente recae en las etiquetas que se utilizan. Clasificar estados de atención visual como *atento* o *noatento* puede tener desventajas, como la simplificación de la complejidad del comportamiento humano. La atención no es un concepto binario y puede variar en intensidad y en dirección. Además, la clasificación puede ser influenciada por factores externos como la tarea en sí misma y la disposición emocional de la persona, lo que puede llevar a resultados poco precisos. Otro problema es que las tecnologías utilizadas para medir la atención, como un $eye-tracker$, pueden ser invasivas y limitar la capacidad de la persona para realizar tareas naturalmente. En resumen, clasificar la atención como *atento* o *noatento* puede proporcionar información útil, pero es importante ser consciente de sus limitaciones y considerar otros factores que puedan afectar la precisión de la clasificación.

En lo que respecta al problema computacional, existe un problema relacionado a modelos que en el pasado han intentado clasificar estados de atención con información de una sesión de seguimiento ocular. Específicamente utilizan rRedes Neuronales Convolucionales (CNN) y series de tiempo extraídas del eye-tracker sin preprocesar. Estos enfoques tienen áreas de oportunidad significativas. En primer lugar, los datos de eye-tracking pueden ser muy ruidosos y contener artefactos que afecten la precisión del modelo. Además, los modelos CNN son propensos a sobresaturarse cuando se les presentan grandes cantidades de datos sin preprocesamiento, lo que puede resultar en una disminución en la precisión. Además, el procesamiento de series de tiempo requiere una gran cantidad de recursos computacionales y tiempo, lo que puede ser un obstáculo para su aplicación en situaciones reales.

1.3. Justificación

La existencia de trastornos neurológicos no es algo precisamente nuevo, de los mencionados anteriormente el más común es Transtorno por Déficit de Atención e Hiperactividad (TDAH), cuyo diagnostico tradicional se basa en observación y aplicación de cuestionarios. Esta forma de análisis tradicional guía hacia un diagnóstico erróneo, se estima que el 20 % de los niños son mal diagnosticados con TDAH sólo por la diferencia de edad con sus compañeros de clase. Estadísticas mundiales arrojan un incremento del 8.5 al 9.5 % en el número de niños diagnosticados con TDAH del 2011 al 2017 [8]. Este trastorno tiene altas probabilidades de prevalencia en la vida adulta, algunas de las secuelas presentadas en adultos diagnosticados son la poca habilidad para relacionarse socialmente, deficiencia en el desempeño académico y profesional, problemas para manejar ansiedad y depresión [9].

Swanson y Volkow [10] han considerado que los estragos de la pandemia por coronavirus 2019 (COVID-19) no sólo se limitan a las muertes por el virus, el confinamiento obligatorio y la reducción de la interacción con la sociedad tendrá un efecto visible en el comportamiento cerebral de niños, jóvenes y adultos. Es preocupante que al término del confinamiento y regreso a las aulas el desempeño neurológico de estudiantes no sea el esperado, pues se piensa que un subtipo etiológico novedoso de TDAH podría detectarse en dicha población. Los expertos especulan que los efectos residuales de la enfermedad COVID-19 pueden afectar selectivamente las regiones del cerebro subyacentes a los déficits de atención y motivación asociados con el TDAH.

En un reciente estudio, Zhao et al. [11] analiza el comportamiento de estudiantes universitarios en el regreso a las aulas. Se observan que los principales problemas de salud mental en jóvenes adultos es ansiedad, depresión y un porcentaje considerable de ellos han sido diagnosticados con TDAH. Padecimientos como altos niveles de ansiedad y estrés en una madre gestante joven incrementan la probabilidad de que el feto sufra de trastornos del desarrollo neurológico en edad adulta [12], como el TDAH, el trastorno del espectro autista, los trastornos del espectro de la esquizofrenia, comportamiento antisocial y síntomas depresivos.

La innovación de sistemas impulsados por Inteligencia Artificial (IA) utilizando EEG, cámaras térmicas con seguimiento ocular e fMRI con seguimiento ocular fueron mencionadas con anterioridad. Pese a esto la clasificación de atención que realizan en muchos casos no es suficiente, pues es una cuestión binaria: atento y no atento. En términos del TDAH, por ejemplo, la observación del comportamiento atencional se basa en estados de atención un poco más complejos como atención sostenida y alternada, pues quien sufre del trastorno tiene dificultades para prestar atención a los detalles, para mantener la atención durante un período prolongado de tiempo y para alternar la atención a objetos relevantes [13].

1.4. Objetivos

Clasificar estados de atención utilizando la segmentación del área de interés visual y seguimiento ocular a través de técnicas de aprendizaje profundo.

1.4.1. Objetivos Específicos

- Determinar los movimientos oculares a evaluar, en base a su relevancia con estados de atención visual.
- Establecer las áreas de interés visual requeridas para el análisis de la información obtenida por el seguimiento ocular.
- Definir los estados de atención que brinden más información sobre el comportamiento de atención visual del sujeto para su clasificación.
- Analizar los datos obtenidos para extraer las características de los movimientos oculares definidos.
- Diseñar y proponer un modelo con técnicas de aprendizaje profundo para la clasificación de los estados de atención.
- Realizar las pruebas necesarias para determinar el desempeño del modelo, y en su caso ajustarlo para obtener resultados satisfactorios.
- Evaluar los resultados obtenidos y comparar con el estado del arte.

1.5. Estructura de la Tesis

La tesis se organiza del siguiente modo:

- Capitulo 2 presenta una revisión por los trabajos relacionados. En este capítulo se realiza una exhaustiva revisión de los trabajos previos y estudios relevantes relacionados con el tema de investigación. Se analizan y resumen los hallazgos más importantes de la literatura existente, proporcionando un contexto sólido para el estudio actual.
- Capitulo 3 muestra a detalle la metodología planteada. En este capítulo se describe detalladamente la metodología propuesta para llevar a cabo la investigación. Se explica el enfoque adoptado, los procedimientos utilizados, los instrumentos de recolección de datos y cualquier otra técnica o estrategia empleada en el estudio. Se busca proporcionar una comprensión clara de cómo se ha llevado a cabo la investigación.
- Capitulo 4 detalla los resultados alcanzados. En este capítulo se describe detalladamente la metodología propuesta para llevar a cabo la investigación. Se explica el enfoque adoptado, los procedimientos utilizados, los instrumentos de recolección de datos y

cualquier otra técnica o estrategia empleada en el estudio. Se busca proporcionar una comprensión clara de cómo se ha llevado a cabo la investigación.

- Capitulo 5 se mencionan las conclusiones de la investigación. En este capítulo se presentan las conclusiones derivadas de la investigación realizada. Se resumen los principales resultados, se destacan las contribuciones y se discuten las implicaciones prácticas y teóricas del estudio. También se mencionan posibles direcciones futuras para la investigación, brindando recomendaciones adicionales o áreas que podrían ser exploradas en trabajos posteriores.

Esta estructura permite a los lectores tener una visión clara y organizada de la tesis, abarcando desde la revisión de la literatura existente hasta los resultados y conclusiones finales de la investigación.

CAPÍTULO 2

Revisión de Literatura

La estimación de estados de atención es una tarea crucial en diversas aplicaciones, incluyendo la medición de la eficacia de la publicidad, la detección de fatiga en conductores y la evaluación de la atención de pacientes con trastornos neurológicos. Con el avance de la inteligencia artificial (IA), se han desarrollado métodos automáticos para la estimación de estados de atención con una mayor precisión y eficiencia que los métodos tradicionales. La estimación de estados de atención es un tema de gran importancia en el campo de la neurociencia y la psicología, ya que la atención es un proceso cognitivo crucial para el desempeño humano. En los últimos años, se ha demostrado que tanto la electroneuroencefalografía (EEG) como que los movimientos oculares contienen información valiosa relacionada a los diferentes estados de atención. La EEG permite registrar la actividad eléctrica cerebral, mientras que los movimientos oculares reflejan el estado de alerta y concentración del sujeto. La revisión de literatura sobre esta área de investigación busca examinar los avances y desafíos en la estimación de estados de atención a través de EEG y movimientos oculares, así como explorar su aplicación en diversas tareas y situaciones.

2.1. Estados de atención

La atención puede definirse como un mecanismo central de control del procesamiento de información, que actúa de acuerdo con los objetivos del organismo activando e inhibiendo procesos, y que puede orientarse hacia los sentidos, las estructuras de conocimiento en memoria y los sistemas de respuesta. En la Figura 2.1 se muestra cómo es que los seres humanos extraemos información del ambiente para su posterior análisis a través de la mirada.

La atención visual ha sido estudiada desde hace décadas para distintos fines [14, 15, 16], este mecanismo propio del ser humano brinda amplia información sobre el desarrollo neuronal y permite analizar distintos aspectos psicológicos. No existe un acuerdo universal en la literatura sobre procesamiento de información con respecto a los mecanismos de atención. La mayoría de los modelos de atención basados en el enfoque del procesamiento de información humana fueron introducidos por primera vez por Broadbent [17]. El problema con todos los modelos de atención introducidos anteriormente que ninguno aborda adecuadamente el

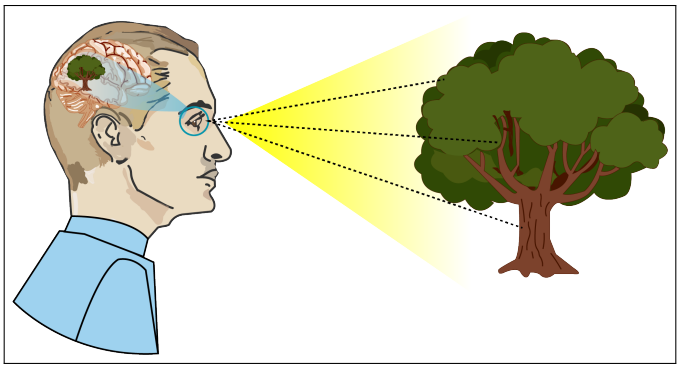


Figura 2.1: Interpretación del ambiente a través de la atención visual [creación propia].

fenómeno clínico de los déficits de atención o su remediación. Los pocos programas de tratamiento para la atención que existen tienden a estar orientados a tareas sin una base teórica sólida. Los estados de atención que se enlistan a continuación fueron descritos por Sohlberg y Mateer [18], estos brindan información certera sobre el comportamiento del sujeto, no se limita únicamente a la resolución de tareas y donde el sujeto analizado está atento o no atento.

- **Atención enfocada:** capacidad de responder discretamente a estímulos visuales, auditivos o táctiles.
- **Atención sostenida:** capacidad de mantener una respuesta conductual constante durante la actividad continua o repetitiva.
- **Atención selectiva:** capacidad de mantener un conjunto cognitivo que requiere activación e inhibición de respuestas dependientes de la discriminación de estímulos.
- **Atención alternada:** capacidad de flexibilidad mental que permite moverse entre tareas que tienen diferentes requisitos cognitivos.
- **Atención dividida:** capacidad de responder simultáneamente a múltiples tareas.

Estos estados de atención visual pueden ser modulados por diferentes factores, como la tarea que se está realizando, el nivel de fatiga o la presencia de distracciones. La capacidad de identificar y medir estos diferentes estados de atención visual a través de EEG y movimientos oculares es crucial para comprender mejor el procesamiento cognitivo humano.

2.2. Movimientos oculares

El comportamiento del movimiento ocular es el resultado de complejos procesos cognitivos, las métricas de la mirada pueden revelar objetivos e información cuantificable sobre la calidad, previsibilidad y consistencia de dichos procesos [19], por lo tanto, son una fuente de

información natural para sistemas proactivos que analizan el comportamiento del usuario. En comparación con otros tipos de datos, los movimientos oculares contienen características menos complejas, cuya dimensionalidad, reducida en su forma original, resulta ideal para la identificación de trastornos neurológicos [9]. La atención visual y el movimiento ocular nos permiten interactuar con ambientes complejos seleccionando información relevante para ser procesada en el cerebro. El comportamiento de la atención junto con los movimientos oculares contiene una huella biométrica de la función o disfunción del cerebro del individuo. Estas huellas tienen potencial para identificar desordenes neurológicos, pues tienen impregnada información característica de desórdenes como el trastorno de déficit de atención por hiperactividad (TDAH) [20]. En la Figura 2.2 podemos ver el método común para la obtención de información relacionada con la mirada.

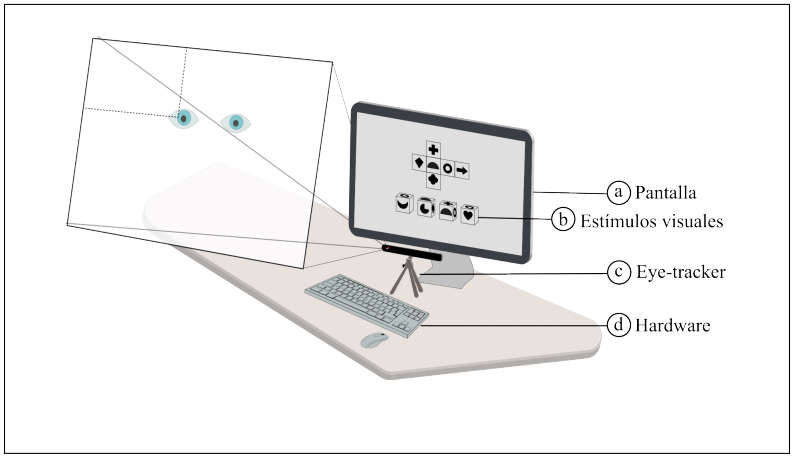


Figura 2.2: Técnica comúnmente utilizada para registrar movimientos oculares[creación propia].

2.3. Estimación de estados de atención con EEG

El proceso cognitivo de la atención es fuente de distintas interrogantes a las que se ha intentado dar respuesta por diferentes caminos. Específicamente determinar el estado de atención que presta el sujeto a determinada tarea. La relevancia de los estados de atención recae en que no sólo pueden ser utilizados como fuente de información sobre desempeños académicos, por ejemplo, en la aplicación de alguna prueba. Previamente mencionamos la importancia atención visual en diferentes patologías, pues los patrones visuales propios de cada trastorno pueden aportar al diagnóstico diferencial de cada uno.

Dentro de las principales técnicas de estimación de atención visual nos encontramos con aquellas propuestas que hacen uso de actividad cerebral obtenidas por señales EEG. Belle et al. [21] proponen desarrollar un sistema de monitorización que utilice señales ECG para analizar y predecir la presencia o falta de atención cognitiva en los individuos durante la ejecución de una tarea. Se utilizar diferentes modelos de Aprendizaje Automático (AA) para

la clasificación. Los resultados muestran que la precisión global del modelo de clasificación basado en bosques aleatorios fue superior al modelo clasificación basado regresión, con una precisión de clasificación de casi el 86 % para el conjunto de características del EEG. Se concluye entonces que el análisis de las señales de EEG sirve principalmente para establecer un punto de referencia con el que comparar el análisis de las características fisiológicas del brazalete. El sistema propuesto se centra principalmente en la señal del electrocardiograma y en ella se aplican varios métodos de descomposición.

Dado que las emociones, el estado mental y la atención de una persona se rigen por diversas partes frontales del cerebro, la observación de las señales EEG de esta zona es un método viable para determinar si los alumnos están atentos. Liu et al. [22] presenta una propuesta para observar e identificar si los alumnos están atentos mediante una sencilla detección y clasificación de señales EEG. Los resultados del estudio indican que el estado de atención es un fenómeno continuo, y que la observación de datos de un periodo limitado es ligeramente superior a la precisión de clasificación de una sola entrada de datos. Además, los resultados en muestran que Máquina de Soporte de Vectores (SVM) puede proporcionar una mayor precisión cuando se emplean todas las características de las señales EEG, específicamente alcanzando una exactitud del 71 % para la clasificación de dos estados de atención. Los autores discuten sobre las señales de atención del EEG son más fáciles de identificar en comparación con las de desatento; esto se debe a que las señales de falta de atención del EEG contienen más información.

S. M. Yang et al. [23] proponen un análisis adicional a las señales EEG, argumentando que las señales del deben ser realzadas porque, en general, se miden con señales eléctricas débiles del cerebro. Después de utilizar una optimización con algoritmos genéticos con selección óptima de modelos y características, este estudio desarrolla un nuevo modelo basado en señales EEG para identificar los niveles de alta y baja atención de los estudiantes en un entorno autónomo de e-learning. De acuerdo con sus resultados, el modelo propuesto puede identificar con precisión los periodos de baja atención de la videoconferencia que los alumnos generaron en cierta medida basándose en las medidas de rendimiento de la precisión, la tasa de recall y la F1 score. A través de una clasificación con SVM se obtuvieron valores de precisión del 91.60 % para el estado atento y de 87.44 % para el estado alto y bajo, respectivamente. Nuevamente se presenta la observación que para estados reducidos de atención existe menos información en señales EEG.

Aliakbaryhosseinabadi et al. [24] concluyen que, con la precisión del modelo análisis discriminante lineal que alcanzó un 71 % de exactitud, es posible definir un criterio global para investigar niveles de atención durante la ejecución de tareas motoras con el monitoreo de señales EEG. Toa et al. [25] presentan un sistema útil para quienes estén interesados en desarrollar monitoreo de la atención en diversos ambientes, en el entorno escolar, la investigación del comportamiento y durante una evaluación de ejecución en la industria. Se propone inicialmente combinar Memoria a corto y largo plazo (LSTM) y una red neuronal totalmente conectada para aprender características de alto nivel de la señal de EEG sin procesar y realizar la clasificación. En segundo lugar, se propone un modelo de AP con modelado *Vec2Vec*

[26] que puede aprender en un enfoque de extremo a extremo mediante una red neuronal. El modelo propuesto se denomina Red Neuronal de Convolución con Memoria de Atención que utiliza Vec2Vec. Concluyen que CAMNN puede analizar eficazmente las señales EEG de los participantes a través del modelo de AP y clasificar su nivel de atención con una exactitud del 92 %.

2.4. Estimación de estados de atención con movimientos oculares

Los movimientos oculares han sido foco de investigaciones desde hace varias décadas, podrá creerse que se encuentran en un área emergente de un campo nuevo de investigación, pero lo cierto es que desde hace 20 años se había notado el potencial que tiene la información brindada por la vista [27]. La clasificación de estados de atención utilizando movimientos oculares se ha convertido en un tema de interés creciente en la investigación neuropsicológica y de neurociencia. La atención es una habilidad fundamental que permite a las personas procesar y responder a la información relevante en su entorno. La capacidad de medir y clasificar los estados de atención puede tener aplicaciones prácticas en una amplia gama de áreas, incluyendo la evaluación de trastornos de atención, la medición de la eficacia de las intervenciones de atención y la investigación en neurociencia cognitiva.

Los movimientos oculares se han utilizado como un indicador de la atención y se ha demostrado que están correlacionados con la atención y la concentración. Los estudios han utilizado diversas técnicas para medir y clasificar los movimientos oculares, incluyendo la velocidad de los movimientos oculares, la dirección y la frecuencia de los movimientos. Se han desarrollado varios modelos de clasificación de estados de atención utilizando movimientos oculares, incluyendo modelos basados en la teoría de señalización, modelos basados en el análisis de señales y modelos basados en la teoría de la dinámica de sistemas. Algunos de estos modelos han sido evaluados en estudios empíricos y se ha demostrado que son efectivos en la identificación de diferentes estados de atención.

El interés por desarrollar sistemas capaces de clasificar estados de atención de una forma simple y que no involucren al individuo a procesos invasivos para la obtención de información, junto con el creciente interés del seguimiento ocular de los últimos años, inspira a diversos autores a presentar propuestas de estimación de atención con técnicas de IA y movimientos oculares tales como fijaciones o movimientos sacádicos.

Mohammadi-Aragh et al. [28] investigan la precisión del uso un sistema de clasificación de ondas de Haar para distinguir a los estudiantes patrones de atención en una conferencia. Los resultados y discusión demuestran la utilidad de aplicar SVM para la clasificación de estados de atención con precisión del 82,8 %. Chen et al. [5] hacen uso de gafas inteligentes que registran información de la mirada e implementaron un sistema con SVM y algoritmos

genéticos. Los experimentos realizados revelan que el sistema de estimación de atención propuesto puede lograr la precisión del 93,1 %. Zaletelj y Košir [29] desde otro enfoque utilizan datos en 2 y 3 dimensiones obtenidos por el sensor *Kinect One* para construir un conjunto de características que caracterizan propiedades faciales y corporales de un estudiante, incluido el punto de mirada y la postura corporal. A través de diferentes técnicas de AA la exactitud alcanzada fue del 75.3 %. Abdelrahman et al. [30] realizan un estudio con una cámara térmica y dispositivos de seguimiento ocular para estimar cinco estados de atención que brindan más información sobre el proceso cognitivo. Con tres técnicas de AA, concluyen que el mejor desempeño fue alcanzado con regresión logística con características extraídas de ambas fuentes de información con 86.90 % de exactitud. La Tabla 2.1 presenta las aportaciones relevantes del estado del arte presentado en líneas anteriores.

ESTADO DEL ARTE				
EEG				
Contribución	Técnica	Sujetos	Clasificación	Exactitud
[21]	Random Forest	21	Atención y no atención	85.70 %
[22]	SVM	24	Atención, no atención	71 %
[23]	SVM	4	Atención alta, atención baja	85.70 %
[24]	Análisis discriminante lineal	12	Nivel atención controlada, nivel de atención compleja	90 %
[25]	AP	30	Atento, desatento	92 %
Seguimiento ocular				
[28]	SVM lineal	222	Atención No atención	82.80 %
[29]	Árboles de decisión K-Vecinos más Cercanos	3	Atención, no atención	71 %
[5]	SVM y algoritmos genéticos	10	Atención No atención	93.10 %
[30]	Regresión logística	22	Atención enfocada, atención mantenida, atención alternada, atención selectiva compleja, atención dividida	90 %

Tabla 2.1: Estado del arte.

CAPÍTULO 3

Metodología

En el campo de la psicología y la neurociencia cognitiva, la medición de la atención visual es una herramienta importante para comprender cómo procesamos y comprendemos el mundo que nos rodea. Una forma común de medir la atención visual es mediante la estimación de estados de atención visual, que se refiere a la identificación de momentos específicos en los que un sujeto está prestando atención a un estímulo visual en particular. Existen diversas estrategias que han sido implementadas previamente para la estimación o clasificación de estados de atención. Como se mencionó en la sección anterior, las dos fuentes principales de información son los movimientos oculares y EEG. Esta contribución se enfoca en utilizar una fuente de información que no sea invasiva al sujeto durante la adquisición de información, es por ello que las señales de seguimiento ocular resultan ideales para el análisis de atención. En ocasiones, como se mencionó anteriormente, al utilizar un método invasivo en la adquisición, los sujetos pueden ser sometidos a estímulos adicionales por error y existiría un sesgo considerable en la información.

El método planteado en esta investigación consta específicamente de una primera estimación de cinco estados de atención visual, propuestos en el modelo clínico de atención de Sohlberg [18]. Este modelo clínico es considerado como uno de los más completos, pues como se menciono anteriormente, las clasificaciones realizadas por trabajos previos constan de etiquetas o clasificaciones básicas que no contienen información relevante relacionada al estado cognitivo del sujeto. Concretamente existen dos niveles de atención globales: nivel alto y nivel bajo.

Esta primera estimación de atención se realizará a través de un análisis del comportamiento ocular del sujeto durante una prueba. Específicamente se contabilizan los movimientos oculares ejecutados por la mirada: fijaciones, movimientos sacádicos y persecuciones suaves, la duración y velocidad de cada evento realizado. Sin embargo, es importante mencionar que estas medidas solo proporcionan una estimación aproximada de la atención visual y no deben considerarse como la única medida para evaluar la atención visual, esta aproximación brindará información a un especialista para lograr una correcta clasificación de la atención visual.

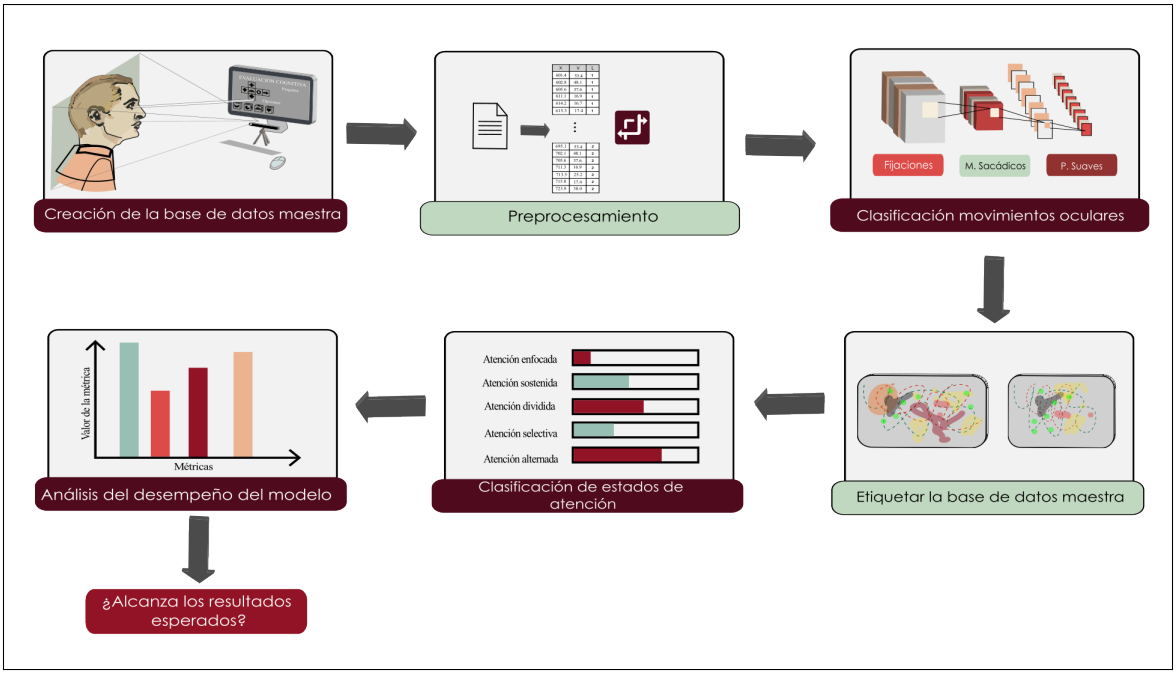


Figura 3.1: Metodología propuesta.

La estimación de atención se realiza a través de un modelo de AP entrenado en una base de datos maestra, compuesta por la base de datos más grande de la literatura relacionada a los movimientos oculares, GazeCom [31], Lund2013 [2] y 360EM [32]. Debido a la naturaleza del modelo, se requiere de una cantidad de datos considerable para el entrenamiento del modelo, pues se puede esperar un mejor desempeño en la tarea con una mayor cantidad de ejemplos. Una vez entrenado el modelo es capaz de estimar el estado de atención visual presentado durante una prueba. El modelo y una presentación del comportamiento ocular se presenta en una interfaz de usuario, embebido el proceso completo en un sistema: limpieza de muestras, preprocesamiento, clasificación y una transformación a un dominio de flujo óptico que representa visualmente el comportamiento de la mirada en la pantalla. Concretamente en la Figura 3.1 encuentran las etapas correspondientes a la metodología de esta investigación.

3.1. Inteligencia Artificial

La Inteligencia Artificial (IA) es un campo de la informática que se enfoca en la creación de sistemas y algoritmos que pueden realizar tareas que normalmente requieren la intervención humana, como el reconocimiento de voz, la visión por computadora, el procesamiento del lenguaje natural y la toma de decisiones. La IA ha evolucionado rápidamente en las últimas décadas y ha encontrado aplicaciones en diversos campos, como la medicina, la robótica, el comercio y la educación. Además, el AA y el AP han permitido el desarrollo de sistemas que pueden mejorar y adaptarse continuamente en función de su experiencia y datos de entrada.

La IA ha encontrado una gran cantidad de aplicaciones en la biomedicina, desde el descubrimiento de nuevos fármacos hasta la predicción de enfermedades y la identificación de biomarcadores para un diagnóstico temprano. Por ejemplo, se están utilizando algoritmos de AA para analizar grandes cantidades de datos de pacientes y de investigaciones médicas para identificar patrones y factores de riesgo para enfermedades. Además, se está utilizando esta tecnología para diseñar nuevas terapias y tratamientos personalizados para enfermedades como el cáncer, la diabetes y la enfermedad de Alzheimer. En el futuro, puede desempeñar un papel cada vez más importante en la mejora de la atención médica y la detección temprana de enfermedades.

Específicamente en psicología y la neurociencia, desde el análisis de datos hasta la evaluación y tratamiento de trastornos mentales. Los algoritmos de AA se están utilizando para analizar grandes cantidades de datos de imágenes para identificar patrones y factores de riesgo para enfermedades neurológicas y trastornos mentales. Además, se están desarrollando sistemas de inteligencia artificial para mejorar la precisión de los diagnósticos y el tratamiento personalizado de los trastornos mentales, como la depresión, la ansiedad y el trastorno del espectro autista. En la psicología, la inteligencia artificial se está utilizando para crear modelos de procesamiento cognitivo, como el aprendizaje y la memoria, y para desarrollar herramientas de evaluación y tratamiento para pacientes con trastornos cognitivos y del desarrollo. En resumen, tiene el potencial de transformar la forma en que se estudian y tratan los trastornos mentales y neurológicos en la psicología y la neurociencia.

3.1.1. Aprendizaje Automático

El aprendizaje automático, o machine learning en inglés, es una rama de la inteligencia artificial que se enfoca en desarrollar algoritmos que permiten a las computadoras aprender y mejorar su rendimiento en una tarea específica a partir de la experiencia previa y la retroalimentación recibida. Esto significa que los algoritmos de aprendizaje automático son capaces de detectar patrones en grandes conjuntos de datos y utilizar esta información para hacer predicciones o tomar decisiones sin necesidad de una programación explícita. El aprendizaje automático tiene una amplia variedad de aplicaciones en campos como la medicina, la economía, la seguridad, el transporte, entre otros [33].

3.1.2. Aprendizaje Profundo

El aprendizaje profundo (AP), una rama del aprendizaje automático basado en redes neuronales profundas, se ha convertido en una herramienta poderosa para la investigación en biomedicina. Esta técnica ha demostrado su utilidad en diversas áreas, como la identificación de patrones en imágenes médicas, la clasificación de datos genómicos y la predicción de resultados de tratamientos. Por ejemplo, el AP se ha utilizado para analizar imágenes de resonancia magnética y tomografía computarizada para detectar tumores o anomalías

estructurales en el cerebro, así como para clasificar patologías a partir de imágenes de microscopía. Además, el AP ha permitido la identificación de biomarcadores y la predicción de riesgos en enfermedades crónicas, como la diabetes o el cáncer, a partir de análisis genómicos y clínicos.

Redes Neuronales Convolucionales

Las redes neuronales convolucionales (CNN, por sus siglas en inglés) son una arquitectura específica de redes neuronales profundas que han demostrado un gran éxito en el procesamiento y clasificación de imágenes. Su funcionamiento se basa en la aplicación de operaciones de convolución y pooling en cada capa de la red para extraer características relevantes de la imagen de entrada.

La operación de convolución implica la aplicación de un conjunto de filtros o *kernels* sobre la imagen de entrada. Cada filtro se desplaza por la imagen realizando una multiplicación de los valores de los píxeles de la imagen que están bajo el filtro, y luego se suma el resultado de todas las multiplicaciones para generar un solo valor en la salida de la capa convolucional. Este proceso se repite con diferentes filtros para extraer diferentes características de la imagen, como bordes, texturas y formas. En su forma más general, la convolución es una operación sobre dos funciones de un argumento de valor real [34]. La salida $S(i, j)$ a la que se le aplica la operación convolución, en términos de las posiciones i y j dentro de una cuadrícula de dos dimensiones, se define de la siguiente manera:

$$S(i, j) = (K * I)(i, j) \tag{3.1}$$

Donde:

K: kernel
I: entrada

En la Figura 3.2 se muestra un ejemplo de cómo el *kernel* va realizando la operación con partes de una cuadrícula.

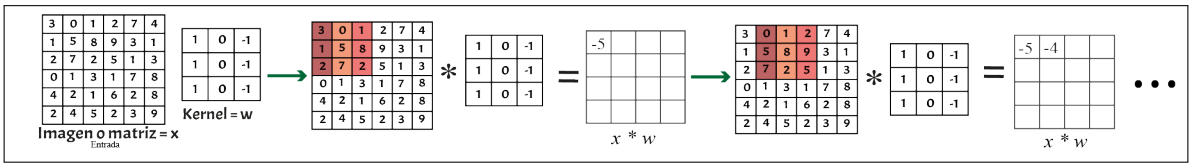


Figura 3.2: Ejemplo de secuencia del recorrido de una cuadrícula con un kernel [creación propia].

La operación de pooling se utiliza para reducir el tamaño de la representación de la imagen en la red, lo que permite una computación más eficiente. En esta operación, se divide la imagen en regiones y se toma el valor máximo (en el caso del max-pooling) o el valor promedio (en el caso del average-pooling) de cada región para generar una nueva imagen

reducida en tamaño.

Las CNN suelen estar compuestas por varias capas convolucionales y de pooling, seguidas de capas totalmente conectadas que realizan la clasificación final. Durante el entrenamiento, se ajustan los pesos de las capas para minimizar la función de pérdida que mide la diferencia entre las predicciones de la red y las etiquetas reales de las imágenes de entrenamiento.

Redes Neuronales Recurrentes

Las Redes Neuronales Recurrentes (RNN) son un tipo de arquitectura de redes neuronales que permite el procesamiento de datos secuenciales, como el lenguaje natural o señales de series de tiempo. A diferencia de las redes neuronales convencionales, las RNN tienen conexiones retroalimentadas, lo que les permite mantener una memoria interna que les permite aprender y recordar patrones de entrada a lo largo del tiempo. as redes neuronales recurrentes (RNN) son una familia de redes neuronales para procesar datos secuenciales. Una red neuronal recurrente es una red neuronal especializada en procesar una secuencia de valores $x^1, ..., x^n$, estas redes se pueden construir de muchas formas diferentes [34].

$$h^{(t)} = f(h^{(t-1)}, x^{(t)}; \theta) \tag{3.2}$$

Donde:

h: estado
x: entrada
t: tiempo
θ: valor inicial

Una de las tareas más comunes al utilizar RNN es predecir el futuro a partir del pasado, la red normalmente aprende a usar $h^{(t)}$ como una especie de resumen con pérdidas de los aspectos relevantes para la tarea de la secuencia pasada de entradas hasta t [34].La unidad básica de una RNN es la célula de memoria, que toma como entrada la entrada actual y la memoria anterior y produce una salida y una nueva memoria para la siguiente iteración. La memoria se *retroalimenta* para la siguiente iteración, lo que permite a la RNN tener en cuenta el contexto temporal y procesar la información de manera secuencial. La arquitectura tradicional de una RNN se muestra en la Figura 3.3.

Las RNN tienen diversas aplicaciones, como el procesamiento del lenguaje natural, la generación de texto, la traducción automática, la clasificación de imágenes secuenciales y la predicción de series de tiempo. Una variante popular de las RNN son las LSTMs (redes neuronales de memoria a largo plazo), que están diseñadas para resolver el problema de la desaparición del gradiente en las RNN, lo que les permite mantener la memoria a largo plazo y aprender patrones de entrada de larga duración.

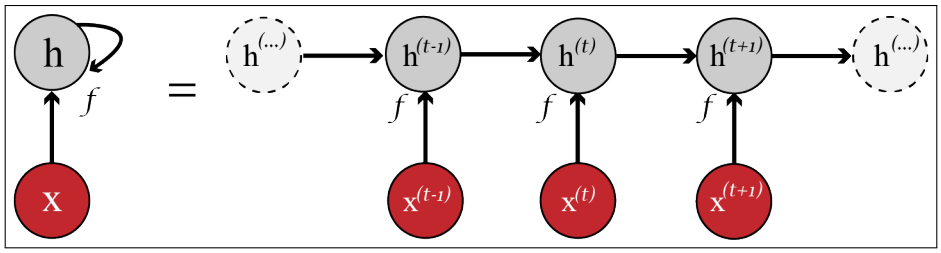


Figura 3.3: Arquitectura ejemplo de una RNN sin capa de salida[creación propia].

Transformers

Un *Transformer* es un modelo de aprendizaje profundo utilizado en el procesamiento del lenguaje natural (NLP) que se basa en la atención. Vaswani et al. [35] presentan el *Transformer*, el primer modelo de transducción de secuencias basado enteramente en la atención, sustituyendo las capas recurrentes más comúnmente utilizadas en arquitecturas codificador-decodificador por la autoatención multi-cabezal. Este modelo puede entrenarse significativamente más rápido que las arquitecturas basadas en capas recurrentes o convolucionales para tareas de traducción. Como se menciono previamente, el funcionamiento se basa en el concepto de atención, que permite al modelo enfocarse en partes relevantes de la entrada al momento de realizar la predicción. El proceso de atención se realiza mediante un mecanismo que permite a la red ponderar la importancia de cada elemento de entrada en función de la tarea en cuestión.

Esta arquitectura consta de dos partes principales: el codificador y el decodificador. El codificador toma una secuencia de entrada y la transforma en una serie de vectores de características, mientras que el decodificador toma esta representación de entrada y genera una secuencia de salida. En el corazón del *Transformer* se encuentra el mecanismo de atención multi-cabezal, que permite que el modelo preste atención a diferentes partes de la entrada simultáneamente. Cada cabeza de atención aprende a enfocarse en una parte diferente de la entrada y a calcular su importancia para la tarea en cuestión. Los resultados de las distintas cabezas de atención se combinan mediante una capa lineal para producir una representación final de la entrada. Una vez que la entrada ha sido procesada por el codificador, se utiliza el decodificador para generar la salida. El decodificador también utiliza el mecanismo de atención multi-cabezal, pero en este caso se utiliza para enfocarse en la entrada codificada y generar la salida paso a paso.

3.1.3. Aplicaciones de la atención

El Transformer utiliza la atención multicabezal de tres formas distintas:

- En las capas de *atención codificador-decodificador*, las consultas proceden de la capa decodificadora anterior, y las claves y valores de memoria proceden de la salida del codificador. Esto permite que cada posición del decodificador atienda a todas las

posiciones de la secuencia de entrada. Esto imita los mecanismos típicos de atención codificador-decodificador en modelos secuencia-a-secuencia como [38, 2, 9].

- El codificador contiene capas de *autoatención*. En una capa de autoatención todas las claves, valores y consultas proceden del mismo lugar, en este caso, la salida de la capa anterior del codificador. codificador. Cada posición del codificador puede atender a todas las posiciones de la capa anterior del del codificador.
- Del mismo modo, las capas de autoatención del descodificador permiten que cada posición del descodificador atienda a todas las posiciones del decodificador hasta esa posición inclusive. Debemos evitar el flujo de en el decodificador para preservar la propiedad autorregresiva.

3.2. Flujo Óptico

Flujo óptico es un método utilizado en visión por computadora para analizar el movimiento de los objetos en una secuencia de imágenes. Se basa en la suposición de que los píxeles de una imagen se mueven de manera continua en el tiempo, y que el movimiento entre dos imágenes adyacentes puede ser representado por un vector de desplazamiento para cada píxel. Existen diferentes métodos para calcular el flujo óptico, pero en general se pueden dividir en dos categorías: métodos densos y métodos dispersos. Los métodos densos calculan el flujo óptico para cada píxel de la imagen, mientras que los métodos dispersos solo calculan el flujo óptico en puntos específicos seleccionados previamente.

La estimación de un campo de movimiento denso, correspondiente al desplazamiento de cada píxel, se denomina flujo óptico. La noción de flujo óptico se refiere a los desplazamientos de los patrones de intensidad. Esta definición tiene su origen en una descripción fisiológica de la percepción visual del mundo a través de la formación de imágenes en la retina. Existen dos tipos: El flujo óptico disperso y el flujo óptico denso. El flujo óptico disperso calcula el vector de movimiento para un conjunto específico de objetos, lo que significa que no habrá información de movimiento sobre los píxeles que no estén contenidos en él. El flujo óptico denso computa el vector de flujo óptico para cada píxel del fotograma, lo que conduce a un posible resultado más preciso [36].

Aunque el campo de flujo óptico se aproxima a la proyección del movimiento real de la escena, proporciona información valiosa en diferentes aplicaciones. Se utiliza ampliamente para tareas de vigilancia visual, incluyendo el seguimiento [37], la segmentación [38], y la detección de anomalías [39]. El concepto de flujo óptico fue propuesto por Gibson, Poggio y Reiehardt [40] presentaron un enfoque para calcular el movimiento de cada píxel en una imagen. Finalmente, el primer modelo práctico de flujo óptico fue establecido por el trabajo clásico de Horn y Schunck [41]. El problema del flujo óptico puede expresarse de la siguiente manera, si $I(x, y, t)$ es la intensidad de una imagen en función del espacio (x, y) durante un tiempo t y los píxeles se desplazan $(\delta x, \delta y)$ después de un tiempo δt, resultará una nueva

imagen que se describirá como el supuesto de intensidad constante, correspondiente a la Ecuación 3.3.

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta y) \tag{3.3}$$

Aplicando una aproximación de Taylor al lado derecho y dividiendo por δt la Ecuación 3.3, se obtiene la Ecuación fundamental del flujo óptico que se describe por la Ecuación 3.4.

$$\frac{dI}{dx}\, u + \frac{dI}{dy}\, v + \frac{dI}{dt} = 0 \tag{3.4}$$

Donde $u = dx/dt$ y $v = dy/dt$, dI/dx es el gradiente de la imagen a lo largo del eje horizontal, dI/dy es el gradiente de la imagen a lo largo del eje vertical, y dI/dt es un tiempo largo. Al tener sólo una ecuación para encontrar dos valores desconocidos, se utilizan diferentes técnicas para encontrarlos.

Existen varias técnicas de flujo óptico, cada una de las cuales utiliza diferentes métodos para resolver el problema descrito en la Ecuación 3.4. Estas técnicas pueden clasificarse en los siguientes grupos: técnicas diferenciales, basadas en regiones, basadas en características, basadas en frecuencias y basadas en CNN. En la presente contribución, se utiliza el algoritmo de estimación de flujo óptico de Farneback [42] para extraer el flujo óptico denso. Esta técnica de flujo óptico diferencial ha demostrado previamente su eficacia en diferentes aproximaciones con señales médicas [43].

Algoritmo de Farneback

El algoritmo Farneback incrusta un modelo de movimiento de traslación entre vecindarios de dos fotogramas consecutivos mediante expansión polinómica. La idea de la expansión polinómica es aproximar algún vecindario de cada píxel con un polinomio cuadrático y estimar las intensidades de los píxeles contenidos en él. La magnitud y dirección del flujo óptico se calculan a partir de una matriz de vectores de flujo (u, v); la visualización de la dirección del flujo se representa por el tono y la distancia o magnitud del flujo por el valor de la representación de color HSV(Hue, Saturation, Value) [42]. Este es un modelo de color ampliamente utilizado en el campo de la informática gráfica y el procesamiento de imágenes. A diferencia del modelo RGB (Red, Green, Blue), que se basa en la combinación aditiva de colores primarios para generar otros colores, el modelo HSV se centra en los atributos perceptuales del color.

Fanerback es un enfoque clásico utilizado en el análisis del flujo óptico. Esta técnica se basa en la correlación de píxeles en diferentes fotogramas de una secuencia de imágenes para detectar el movimiento de objetos en la escena. La técnica de Fanerback implica el cálculo de un campo de Flujo Óptico (OF) en cada punto de la imagen. En primer lugar, se calcula la correlación de píxeles entre dos fotogramas consecutivos. Luego, se busca el desplazamiento que maximiza la correlación. Este desplazamiento se utiliza para calcular el OF para cada punto de la imagen. Los vectores de OF se pueden visualizar como flechas que indican la dirección y magnitud del movimiento en cada punto de la imagen.

El algoritmo de Farneback es una técnica de flujo óptico de segundo orden que se utiliza para estimar la velocidad y dirección del movimiento de los objetos en una secuencia de imágenes. La implementación de este algoritmo en OpenCV se basa en la estimación de la matriz de derivadas espaciales y temporales de las imágenes de entrada. Primero, las imágenes se convierten a escala de grises y se suavizan para reducir el ruido. Luego, se calculan las matrices de derivadas parciales de segundo orden para cada píxel de la imagen. Después, se aplica un método de pirámide Gaussiana para reducir el tamaño de las imágenes y el cálculo de las matrices de derivadas. Esto permite aumentar la precisión del algoritmo al considerar diferentes escalas espaciales y temporales del movimiento. A continuación, se calcula el flujo óptico entre las dos imágenes mediante la estimación de la matriz de correlación de los vectores de flujo de los píxeles vecinos. Por último, se utiliza una técnica de interpolación para suavizar el flujo óptico y eliminar el ruido.

Modelo de color HSV

La representación del flujo óptico con el modelo de color HSV permite una mejor visualización de la magnitud y la dirección del movimiento en una secuencia de imágenes. la representación del flujo óptico con el modelo de color HSV en aplicaciones de detección de movimiento ofrece una mayor distinción de colores, independencia de la luminosidad y sensibilidad a los cambios de color. Esto permite una detección de movimiento más precisa y confiable en escenas complejas, como en la vigilancia y el análisis de vídeo. El canal Hue representa la dirección del movimiento, mientras que los canales Saturation y Value representan la magnitud. El canal Hue utiliza un esquema de colores cíclicos, donde el color rojo se asigna a un movimiento hacia la derecha, el verde se asigna a un movimiento hacia abajo, el azul se asigna a un movimiento hacia la izquierda y el morado se asigna a un movimiento hacia arriba. La saturación y el valor se utilizan para indicar la fuerza del movimiento.

Los componentes principales del modelo HSV son:

- Matiz (Hue): Representa el tono del color y se refiere a la ubicación del color en el espectro. Los valores de matiz abarcan un rango de 0 a 360 grados, donde 0 y 360 corresponden al rojo puro, 120 al verde y 240 al azul.

- Saturación (Saturation): Indica la pureza o intensidad del color. Valores más altos de saturación representan colores más vivos e intensos, mientras que valores más bajos se acercan a tonos de gris. La saturación se mide en porcentaje, generalmente en un rango de 0 % a 100 %.

- Valor (Value): Representa el brillo o la claridad del color. Valores más altos de brillo se traducen en colores más claros y brillantes, mientras que valores más bajos se acercan al negro. El valor también se mide en porcentaje, generalmente en un rango de 0 % a 100 %.

El modelo HSV permite una representación más intuitiva del color, ya que separa los aspectos perceptuales del color en componentes independientes. Esto facilita el ajuste y

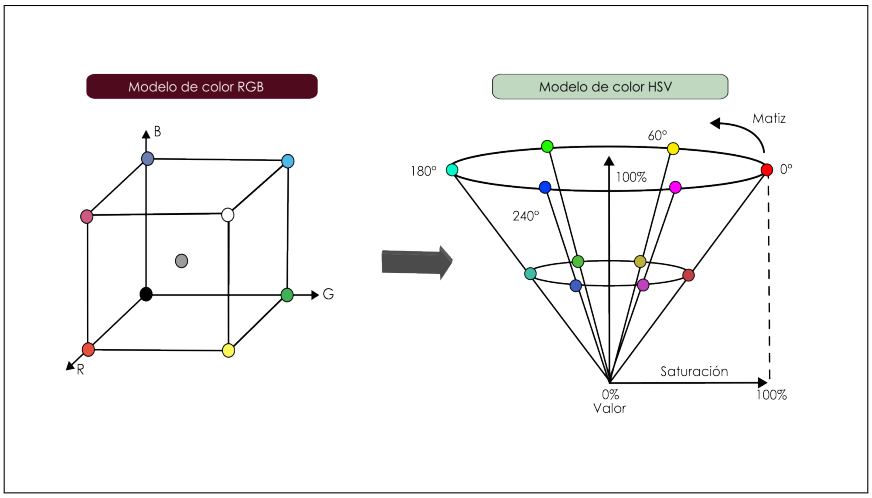


Figura 3.4: Comparativa y explicación a detalle de dos modelos de color: RGB y HSV.

manipulación de colores, así como la selección de tonos específicos en aplicaciones gráficas y de procesamiento de imágenes.

La representación del flujo óptico con el modelo de color HSV es particularmente útil en aplicaciones donde la detección de movimiento es crucial, como en la vigilancia y el análisis de vídeo. Además, esta representación también se utiliza en la detección de objetos en movimiento y en el seguimiento de objetos en un vídeo. Sin embargo, es importante tener en cuenta que esta representación no siempre es precisa, especialmente en escenas con cambios bruscos de iluminación o donde hay oclusiones. En tales casos, se pueden requerir técnicas más avanzadas de procesamiento de imágenes para una representación precisa del flujo óptico. El uso de la representación de flujo óptico en la investigación no tiene los principales problemas de la iluminación ya que la información de las sesiones de eye-tracking transformadas a este dominio no sufren de cambios bruscos de entorno.

3.3. Base de datos

Como se ha mencionado anteriormente, las sesiones de eye-tracking contienen una gran cantidad de información relacionada a los movimientos oculares realizados. Estos contienen información sobre el comportamiento ocular y por consecuencia, del estado cognitivo del sujeto. El principal reto de esta investigación son las bases de datos, ya que en la literatura no existen bases de datos numerosas relacionadas al problema planteado. A continuación se muestra la información de las bases de datos públicas disponibles etiquetadas con los movimientos oculares principales: Fijaciones (FX), Movimientos Sacádicos (SC) y Seguimientos Suaves (SP). La base de datos maestra, será la unión de las bases de datos contenidas en la Tabla 3.1. A través del cambio de dominio a las diferentes bases de datos, que fueron capturadas con diferentes dispositivos, y por consecuencia con diferentes frecuencias de muestreo, se pretende agregar robustez al modelo.

Dataset	Duración	Pruebas	Frecuencia	Equipo	Distribución
Lund2013 [2]	14.9 min	imágenes puntos en movimiento clips de video	500 Hz	SMI Hi-Speed 1250	46.49 % FX 5.88 % SC 3.34 % PSO 41.60 % SP
GazeCom [31]	14.1 h	imágenes clips de video	250 Hz	EyeLink II	73.96 % FX 10.67 % SC 9.83 % SP
360EM [3]	32.9 min	clips de video	120 Hz	FOVE	75.15 % FX 10.44 % SC 9.76 % SP

Tabla 3.1: Bases de datos que conforman la base de datos maestra. Donde *FX* se refieren a Fijaciones, *SC* Movimientos sacádicos, *PSO* Oscilaciones Posacádicas y *SP* Persecuciones suaves.

3.3.1. GazeCom

El conjunto contiene grabaciones de 18 vídeos con 47 sujetos en cada uno. Los sujetos estaban sentados a 45 cm de una pantalla de 40 cm por 30 cm con imágenes a una resolución de 1280 por 720 píxeles. El estímulo abarcaba 48 por 27 grados de ángulo visual, correspondiendo 1 grado a unos 26,7 píxeles [1]. Se realizó una calibración binocular, pero sólo se registraron datos monoculares, lo que dio lugar a un error de validación medio de 0,62 grados en todos los sujetos. En los metadatos de la grabación se almacenan las especificaciones del experimento. El conjunto de datos contiene las posiciones brutas de la mirada, las etiquetas los resultados de los distintos algoritmos de movimiento ocular y las características extraídas. La verdad básica contiene el tiempo en microsegundos, la posición x (horizontal) e y (vertical) como posición de la mirada en la pantalla como la posición de la mirada en la pantalla, la confianza del rastreador ocular y la puntuación de cada etiquetador y una puntuación final combinada. El conjunto de datos contiene 5 clases: fijaciones, sacadas, persecuciones suaves y ruido, etiquetadas de 0 a 4 en el orden respectivo.

Figura 3.5: Ejemplos del escenario de pruebas utilizado durante la grabación de la sesión de eye-tracking [1].

Escenario experimental

Dorr et al. [1] grabaron a 76 sujetos con un rastreador ocular *SR Research EyeLink II* a 250 Hz. Los sujetos se distribuyeron en tres experimentos. El primer experimento 54 sujetos tuvieron que ver 18 películas de escenas reales de Lübeck y sus alrededores. El segundo experimento contó con 11 sujetos que acudieron dos días seguidos para medidas repetidas. Vieron cuatro *trailers* de películas de Hollywood seis películas seleccionadas de las 18 películas reales, como se muestra en la Figura 3.5. En el tercer experimento también participaron 11 sujetos, se les mostraron nueve películas en *stop motion* hechas a partir de las películas del mundo real del primer experimento. Después se les mostraron imágenes estáticas de las nueve películas del primer experimento. Agtzidis et al. [44] etiquetaron todo el conjunto de datos etiquetando primero automáticamente todo el conjunto de datos y luego haciendo que codificadores manuales lo revisaran y lo corrigieran. Solo se etiquetaron los datos del experimento uno, lo que dio como resultado 4,3 millones de muestras repartidas entre 38629 fijaciones, 39217 sacadas y 4631 persecuciones suaves. Este es el conjunto etiquetado más grande a disposición del público y supone unas 4,8 horas de grabación. Podría considerarse un conjunto de datos de referencia, ya que se han probado múltiples algoritmos de clasificación de movimientos oculares de última generación con los datos y su rendimiento también está disponible públicamente.

3.3.2. Lund2013

Larsson et al. [2] grabaron a 31 sujetos con un rastreador ocular *Hi-Speed 1250 de Senso-Motoric Instruments* a 500 Hz. A los sujetos se les presentaron imágenes estáticas, lecturas, videoclips, estímulos de puntos en movimiento y texto de desplazamiento vertical. Sólo un subconjunto de imágenes, estímulos de puntos móviles y videoclips fue etiquetado manualmente por dos evaluadores. Marcus Nyström (MN) y Richard Andersson (RA). El dataset contiene etiquetas para diferentes movimientos oculares: fijaciones, movimientos sacádicos, oscilaciones posacádicas (PSO) y persecuciones suaves. Y etiquetas adicionales como par-

padeos y muestras indefinidas. Tiene una distribución de eventos del 46.49 % corresponde a fijaciones, del 5.8 % movimientos sacádicos, del 3.34 % a oscilaciones posacádicas y del 41.60 % a persecuciones suaves, con un total de 1015 eventos.

Escenario experimental

Los datos se registraron utilizando un sistema Hi-Speed 1250 montado en torre, de Senso-Motoric Instruments GmbH (Teltow, Alemania), que minimiza los movimientos de la cabeza mediante un apoyo para la barbilla y la frente. Todos los datos se registraron durante una sesión en una de las salas de experimentos del Laboratorio de Humanidades de la Universidad de Lund. El desplazamiento medio de la mirada de los participantes según un procedimiento de validación de 4 puntos fue de 0,40°. La precisión se estimó midiendo la desviación cuadrática media (RMSD) de las coordenadas x, y de las muestras identificadas por ambos expertos humanos como fijaciones. Las instrucciones consistían en ver libremente las imágenes mirar los objetos en movimiento de los vídeos y seguir los puntos móviles. Algunas de las imagenes y videos empleados en este escenario experimental se muestran en la Figura 3.6.

Figura 3.6: Escenarios empleados durante la sesión de eye-tracking de la base de datos Lund2013 [2].

3.3.3. 360EM

Agtzidis et al. [3] crearon una base de datos contiene 16 vídeos de 360 grados de diferentes escenas, como un bosque, un concierto y un mercado, y fue capturada utilizando una cámara de 360 grados y un sistema de seguimiento ocular. Este conjunto se compone de los datos de seguimiento ocular de 24 participantes, que observaron los vídeos y sus movimientos oculares

se registraron durante el proceso. La base de datos contiene información sobre la posición de los ojos, el diámetro de la pupila, el parpadeo y la dirección de la mirada, y también incluye información sobre el comportamiento de la cabeza y el cuerpo. Además de la base de datos, los autores también presentan un algoritmo de clasificación de movimientos oculares que se utilizó para analizar los datos de seguimiento ocular. El algoritmo se basa en el análisis de los movimientos oculares en tres categorías principales: fijaciones, movimientos sacádicos y persecuciones suaves. Contiene un total de 24 archivos de análisis del movimiento ocular, uno para cada sujeto que participó en los experimentos.

Escenario experimental

La colección de vídeos reunida incluye 14 clips naturalistas de YouTube y un vídeo generado sintéticamente. Los clips seleccionados en seleccionados representan un conjunto diverso de diferentes categorías de contenido y contexto, por ejemplo, cámara estática, caminando, en bicicleta o conduciendo, así como propiedades tales como que el contenido represente una escena de interior o una escena al aire libre, el entorno abarrotado o vacío urbano o mayoritariamente natural, etc. como se muestra en la Figura 3.7 La duración de los vídeos completos era muy variable, por lo que se decidió utilizar un máximo de un minuto por estímulo.

Figura 3.7: Ejemplos de escenarios empleados durante la sesión de eye-tracking de la base de datos 360EM [3].

3.4. Métricas de Seguimiento Ocular

Dada la naturaleza del modelo de AP, ya se ha mencionado que la cantidad de datos es un factor determinante en el desempeño del modelo. La base de datos maestra, contiene tres bases de datos publicas etiquetadas unicamente con los movimientos oculares. Una de las aportaciones más significativas de esta investigación consiste en plasmar una relación directa entre los tres principales movimientos oculares y el modelo clínico de atención de Sohlberg y Mateer [18]. Este modelo describe un sistema jerárquico de la atención, cuyos componentes aumentan cada vez más en complejidad, es decir que los últimos niveles de atención, requieren un esfuerzo atencional mayor que los precedentes. Este modelo se describió buscando una manera de evaluar la atención.

La atención y los movimientos oculares están estrechamente relacionados, ya que los movimientos de los ojos son una forma clave en que nuestro cerebro dirige nuestra atención hacia los estímulos visuales. La atención puede ser enfocada en un objeto específico mediante el movimiento de los ojos hacia ese objeto, y los movimientos oculares también pueden ser utilizados para explorar un escenario visual y detectar objetos relevantes. Además, el patrón de movimientos oculares puede indicar el nivel de atención de una persona hacia un estímulo visual, lo que permite a los investigadores medir la atención de manera objetiva.

Por otro lado, la atención también influye en los movimientos oculares. Por ejemplo, cuando estamos enfocados en una tarea específica, nuestros ojos se moverán de manera diferente en comparación con cuando estamos distraídos o aburridos. Además, las interrupciones en la atención pueden llevar a cambios en los patrones de movimientos oculares, lo que indica una distracción temporal. Por lo tanto, la relación entre la atención y los movimientos oculares es bidireccional y ambos aspectos son importantes para comprender cómo procesamos y comprendemos el mundo visual que nos rodea. En general, los movimientos oculares pueden ser un indicador útil del estado de atención de una persona y de cómo están procesando los estímulos visuales en su entorno [45].

3.4.1. Fijaciones

Las fijaciones son un fenómeno importante en el estudio de la percepción visual y la atención, ya que nos permiten examinar la relación entre la fijación de la mirada y la cognición. Las fijaciones son momentos en los que el ojo permanece relativamente inmóvil en un punto específico del campo visual durante un corto período de tiempo. Estas fijaciones pueden ser utilizadas para entender cómo los seres humanos procesan y organizan la información visual, y cómo la atención influye en la percepción.

Las propiedades de las fijaciones son de gran interés ya que permiten medir el tiempo que los seres humanos dedican a examinar objetos específicos. Las fijaciones pueden proporcionar

información sobre la ubicación de la atención, la calidad de la percepción y la cantidad de información que se está extrayendo del estímulo visual. Además, las fijaciones también pueden ser utilizadas para estudiar la memoria visual, ya que las fijaciones pueden ser utilizadas para determinar si un objeto específico ha sido examinado antes o no. En general, las propiedades de las fijaciones son una herramienta valiosa para el estudio de la percepción visual y la atención, ya que pueden proporcionar información detallada sobre cómo procesamos la información visual y cómo nuestra atención se enfoca en diferentes aspectos del entorno visual [46].

Atención Enfocada

Las fijaciones son un proceso complejo que involucra no solo la dirección del movimiento ocular, sino también la velocidad, la duración y la amplitud. Estas propiedades físicas pueden variar según la tarea visual, la complejidad del estímulo, la fatiga visual y otros factores, lo que las hace una herramienta valiosa para el estudio de la percepción visual y la atención.

El número de fijaciones y el tiempo que los ojos pasan fijos en un objeto o estímulo pueden proporcionar información útil sobre la atención enfocada, pero no es suficiente para estimarla de forma precisa y completa. Si bien la duración de las fijaciones pueden indicar que los ojos están dirigidos hacia un estímulo en particular, no necesariamente indican que la atención cognitiva se está enfocando en ese estímulo. Para determinar de manera más precisa la atención enfocada, es necesario considerar otros factores, como la calidad de la fijación (por ejemplo, si los ojos están enfocados adecuadamente en el estímulo), el tiempo que los ojos pasan en desplazamiento entre estímulos, el tamaño y la complejidad de los estímulos, así como la velocidad de procesamiento cognitivo. La duración promedio de fijaciones durante la atención enfocada puede variar dependiendo de la tarea y de las características del estímulo que se está observando. Sin embargo, se ha encontrado que en promedio, durante la atención enfocada, la duración de las fijaciones puede oscilar entre 200 y 500 milisegundos [47].

La velocidad de las fijaciones se refiere a la velocidad a la que el ojo se mueve de un punto a otro en el campo visual. Una alta velocidad de fijaciones puede indicar una atención superficial y una baja velocidad de fijaciones puede indicar una atención profunda y enfocada. Además, la duración de las fijaciones también puede ser utilizada para medir la atención enfocada, ya que una fijación prolongada en un punto específico del campo visual indica que el procesamiento cognitivo está en curso. Se ha encontrado que la velocidad promedio de una fijación durante la atención enfocada es de alrededor de 30 a 60 grados por segundo [48].

Atención Selectiva

La atención selectiva se refiere a la capacidad de enfocar la atención en un estímulo específico mientras se ignoran otros estímulos irrelevantes. En este caso, la velocidad de las

fijaciones puede indicar la eficacia de la atención selectiva, ya que una mayor velocidad de las fijaciones en el estímulo relevante puede indicar una atención más enfocada. Además, la duración de las fijaciones en el estímulo relevante también puede indicar la capacidad de mantener la atención en un objeto específico durante un período prolongado de tiempo.

La relación entre las propiedades físicas de velocidad y duración de las fijaciones en la atención selectiva y las pruebas estáticas es importante, ya que las pruebas estáticas suelen utilizarse para evaluar la capacidad de los individuos para enfocar la atención en un estímulo específico mientras se ignoran otros estímulos irrelevantes. En estas pruebas, los individuos deben buscar un estímulo específico en un campo visual lleno de estímulos irrelevantes. La velocidad y la duración de las fijaciones pueden ser utilizadas para evaluar la eficacia de la atención selectiva en estas pruebas, ya que las fijaciones en el estímulo relevante pueden indicar una atención más enfocada y prolongada, mientras que las fijaciones en los estímulos irrelevantes pueden indicar una atención dispersa. Durante la atención selectiva, su duración promedio puede ser más larga que durante la atención dividida o alternada, ya que se enfoca la atención en un estímulo específico. La duración promedio de las fijaciones durante la atención selectiva puede variar entre 200 y 500 milisegundos, con una velocidad promedio de alrededor de 50 a 100 grados por segundo.

Atención Sostenida

La atención sostenida se refiere a la capacidad de mantener la atención en una tarea específica durante un período prolongado de tiempo. En esta situación, la velocidad de las fijaciones puede indicar la eficacia de la atención en la tarea, ya que una alta velocidad de fijaciones puede indicar una atención superficial y una baja velocidad de fijaciones puede indicar una atención profunda y sostenida. Además, la duración de las fijaciones puede indicar la capacidad de mantener la atención en un objeto específico durante un período prolongado de tiempo [49].

Las propiedades físicas de velocidad y duración de las fijaciones son herramientas importantes para el estudio de la atención sostenida en diferentes tareas y situaciones. La evaluación de la atención sostenida utilizando la velocidad y duración de las fijaciones puede ser utilizada para identificar patrones de atención en diferentes grupos de individuos, como aquellos con trastornos de atención. En general, las propiedades físicas de velocidad y duración de las fijaciones son una herramienta valiosa para el estudio de la atención sostenida, ya que pueden proporcionar información detallada sobre cómo los seres humanos mantienen y enfocan su atención en tareas prolongadas y repetitivas.

Durante la atención sostenida, la duración promedio de fijaciones puede ser más larga que durante la atención dividida o alternada, ya que se enfoca la atención en un estímulo específico durante un período prolongado de tiempo. La duración promedio de las fijaciones durante la atención sostenida puede variar entre 300 y 800 milisegundos, con una velocidad promedio de alrededor de 50 a 100 grados por segundo [50].

Atención Alternada

La atención alternada implica cambiar la atención entre dos o más tareas o estímulos de forma rápida y efectiva. En este caso, la velocidad de las fijaciones puede indicar la eficacia de la atención alternada, ya que una mayor velocidad en cambiar la atención entre los estímulos puede indicar una mayor capacidad para realizar la tarea de forma eficiente. Además, la duración de las fijaciones puede ser utilizada para medir la capacidad de cambiar la atención de forma rápida y eficiente, ya que una duración corta en las fijaciones puede indicar una mayor capacidad para cambiar rápidamente la atención entre los estímulos.

La relación entre las propiedades físicas de velocidad y duración de las fijaciones en la atención alternada y las pruebas dinámicas es importante, ya que pueden ser utilizadas para evaluar la capacidad de los individuos para cambiar la atención entre diferentes tareas o estímulos de forma rápida y efectiva. En estas pruebas, los individuos deben cambiar rápidamente la atención entre diferentes estímulos en un campo visual lleno de estímulos irrelevantes en constante movimiento. Es importante tomar en cuenta tanto la duración como la velocidad de las fijaciones en la estimación de la atención alternada, ya que este tipo de atención implica alternar entre múltiples estímulos o tareas en un corto período de tiempo. Por lo tanto, se espera que las fijaciones sean más cortas y rápidas en comparación con otras formas de atención, como la atención enfocada o la atención sostenida. Además, la capacidad para cambiar entre estímulos o tareas de manera eficiente se relaciona con una mejor función cognitiva y una mayor flexibilidad mental.

En cuanto a las características promedio de las fijaciones durante la atención alternada, la duración suele oscilar entre 200 y 300 milisegundos, y la velocidad media se encuentra en el rango de 250-300 grados por segundo. Sin embargo, estas cifras pueden variar según el tipo de estímulos y las demandas de la tarea en cuestión. En general, una mayor velocidad y una menor duración de las fijaciones pueden indicar una mejor capacidad para cambiar de una tarea a otra en la atención alternada [51].

Atención Dividida

La atención divida implica la capacidad de procesar dos o más tareas al mismo tiempo. En este caso, la velocidad de las fijaciones puede ser utilizada para evaluar la capacidad de los individuos para procesar dos o más estímulos al mismo tiempo, mientras que la duración de las fijaciones puede indicar la capacidad de los individuos para mantener la atención en las diferentes tareas durante un período prolongado de tiempo [52].

La relación entre las propiedades físicas de velocidad y duración de las fijaciones en la atención dividida y las pruebas dinámicas es crucial, ya que las pruebas dinámicas pueden evaluar la capacidad de los individuos para procesar múltiples tareas simultáneamente en entornos complejos y cambiantes. Las fijaciones pueden ser utilizadas para identificar los puntos de interés relevantes y cómo se distribuye la atención en las diferentes tareas. Además,

la velocidad y la duración de las fijaciones pueden indicar la eficacia de la atención dividida en estas pruebas, ya que una mayor velocidad en cambiar la atención entre los diferentes estímulos y una duración prolongada de las fijaciones en los estímulos relevantes puede indicar una mayor eficacia en la atención dividida.

La duración y velocidad promedio de las fijaciones durante la atención dividida pueden variar dependiendo de la complejidad de la tarea y la capacidad cognitiva del individuo. Sin embargo, en general, se ha encontrado que la duración promedio de las fijaciones en la atención dividida es de alrededor de 200-250 ms y la velocidad promedio de las fijaciones es de aproximadamente 200-250 g/s. Es importante tener en cuenta estos valores en la estimación de la atención dividida, ya que una disminución en la duración y velocidad de las fijaciones puede indicar una menor capacidad para procesar la información de manera simultánea en múltiples tareas [48].

3.4.2. Movimientos sacádicos

Los movimientos sacádicos son rápidos movimientos oculares que se producen cuando los ojos se mueven de un punto de interés a otro en el campo visual. Estos movimientos son importantes en la estimación de la atención visual, ya que reflejan la capacidad del sistema visual para cambiar rápidamente el enfoque y la atención en diferentes estímulos. Los movimientos sacádicos son esenciales para la exploración visual y la selección de información relevante en el ambiente

Los movimientos sacádicos son especialmente importantes para el procesamiento rápido de la información visual. La capacidad de mover rápidamente los ojos entre diferentes puntos de interés en el campo visual permite una mayor eficiencia en el procesamiento de la información y una mayor capacidad para captar información relevante en entornos visuales complejos. Además, los movimientos sacádicos son importantes en la atención selectiva, ya que pueden ser utilizados para identificar rápidamente los estímulos relevantes y separarlos de los estímulos irrelevantes en el campo visual [52].

Atención Enfocada

En la atención enfocada, los movimientos sacádicos también juegan un papel importante. La capacidad de mover los ojos rápidamente de un objeto a otro permite enfocar la atención en la tarea visual y permite un procesamiento visual eficiente. Los movimientos sacádicos también ayudan a las personas a cambiar la atención de una tarea a otra de manera rápida y eficiente. Durante la atención enfocada, la duración y velocidad promedio de un movimiento sacádico pueden variar dependiendo de la tarea y del estímulo que se está observando. Sin embargo, en general, la duración promedio suele ser de alrededor de 20 a 40 milisegundos, mientras que la velocidad promedio puede ser de 300 a 600 grados por segundo [53].

Además, los movimientos sacádicos también pueden proporcionar información sobre la capacidad de la atención enfocada. Las personas con déficit de atención pueden mostrar movimientos sacádicos atípicos o irregulares, lo que indica una dificultad para enfocar la atención de manera efectiva. También se ha encontrado que las personas con trastornos neurológicos, como el síndrome de Tourette o la esquizofrenia, pueden mostrar patrones atípicos de movimientos sacádicos, lo que indica una alteración en la función de la atención [54].

Atención Selectiva

La capacidad de mover rápidamente los ojos de un objeto a otro permite a las personas elegir los objetos relevantes y filtrar los irrelevantes en su campo visual. Los movimientos sacádicos permiten una selección visual rápida y eficiente, lo que es esencial para realizar tareas que requieren atención selectiva. Durante la atención selectiva, estos movimientos suelen ser más cortos y más frecuentes que durante la atención enfocada, ya que se necesita cambiar rápidamente la atención entre diferentes objetos o estímulos. La duración promedio de los movimientos sacádicos durante la atención selectiva puede ser de alrededor de 20 a 50 milisegundos, con una velocidad promedio de 300 a 600 grados por segundo.

Además, los movimientos sacádicos también pueden proporcionar información sobre la capacidad de la atención selectiva. Las personas que tienen dificultades para realizar tareas de atención selectiva pueden mostrar patrones atípicos de movimientos sacádicos, como una mayor frecuencia de movimientos hacia objetos irrelevantes, lo que indica una dificultad para filtrar la información no relevante [55].

Atención Sostenida

En la atención sostenida, la capacidad de mantener la atención en una tarea durante un período prolongado de tiempo es crucial. Los movimientos sacádicos son importantes en la atención sostenida ya que permiten al ojo explorar el campo visual para detectar cambios y mantener la atención en la tarea visual. La investigación ha demostrado que la velocidad y la amplitud de las sacadas disminuyen con el tiempo durante una tarea prolongada, lo que indica una disminución en la eficiencia de la atención visual. Durante la atención sostenida, suelen ser menos frecuentes que durante la atención dividida o alternada, ya que se necesita cambiar menos la atención entre diferentes objetos o estímulos. La duración promedio de los movimientos sacádicos durante la atención sostenida puede ser de alrededor de 20 a 50 milisegundos, con una velocidad promedio de 300 a 600 grados por segundo [56].

Los movimientos sacádicos también pueden proporcionar información sobre la fatiga visual y la carga cognitiva durante la atención sostenida [57]. Estudios han demostrado que las personas que experimentan fatiga visual durante una tarea prolongada muestran una disminución en la amplitud y velocidad de las sacadas, lo que sugiere una disminución en la eficiencia de la atención visual [58]. Por lo tanto, medir y analizar los movimientos sacádicos

durante la atención sostenida puede ser útil para comprender los procesos cognitivos subyacentes y la fatiga visual en tareas prolongadas, lo que puede ayudar a mejorar la eficiencia de la atención y la productividad [57].

Atención Alternada

En análisis de estos movimientos es importante en la estimación de atención alternada ya que permiten a las personas mover la atención rápidamente entre dos o más tareas o estímulos [59]. Los movimientos sacádicos rápidos y precisos son esenciales para la realización de tareas de atención alternada, ya que permiten a las personas mover la atención de manera rápida y efectiva entre diferentes tareas o estímulos [60]. Durante la atención alternada, pueden ser más frecuentes y más cortos que durante la atención enfocada, con una duración promedio de alrededor de 10 a 20 milisegundos y una velocidad promedio de 300 a 600 grados por segundo. Sin embargo, también pueden ser más grandes y más rápidos cuando se cambia la atención de un estímulo a otro [61].

Además, los movimientos sacádicos también pueden proporcionar información sobre la capacidad de la atención alternada. Por ejemplo, las personas con trastornos de atención pueden mostrar patrones de movimientos sacádicos atípicos, como una mayor frecuencia de movimientos entre dos tareas o estímulos, lo que indica una dificultad para mantener la atención en una tarea por un período prolongado [55].

Atención Dividida

La atención dividida es un tipo de atención que se enfoca en la capacidad de procesar múltiples estímulos simultáneamente [62]. Los movimientos sacádicos, en particular la velocidad y duración de estos movimientos, son fundamentales en la medición de la atención dividida y su capacidad para procesar y responder a múltiples estímulos [63]. La velocidad y duración de los movimientos sacádicos pueden ser utilizados para evaluar la capacidad del sistema visual para procesar múltiples estímulos, y cómo los estímulos compiten por la atención limitada del sujeto [64].

La medición de la velocidad y duración de los movimientos sacádicos también puede ser utilizada para evaluar la eficacia de la atención dividida en diferentes tareas. Por ejemplo, en tareas de búsqueda visual, la velocidad y duración de los movimientos sacádicos pueden ser utilizados para medir la capacidad del sistema visual para procesar y seleccionar información relevante entre distracciones irrelevantes. En tareas de seguimiento de objetos, la velocidad y duración de los movimientos sacádicos pueden ser utilizados para medir la capacidad del sistema visual para mantener la atención dividida entre múltiples objetos en movimiento. En general, durante la atención dividida, las fijaciones pueden ser más cortas y más rápidas que durante la atención enfocada, con una duración promedio de alrededor de 150 a 300 milisegundos y una velocidad promedio de 60 a 120 grados por segundo [47].

La implementación presentada en esta investigación pretende que no se requiera de un dispositivo (eye-tracker) especifico, ya que en el mercado existe una gran variedad. Al entrenar el modelo de aprendizaje profundo con datos obtenidos con diferentes equipos a distintas frecuencias de muestreo se pretende lograr generalizar en los patrones de los diferentes movimientos oculares, no en los datos que se les son entregados como normalmente sucede en algunos modelos de AP.

3.4.3. Seguimientos Suaves

Los seguimientos suaves son un tipo de movimiento ocular que se produce cuando la persona sigue un objeto en movimiento de forma suave y continua. Estos movimientos oculares son importantes para la percepción visual y para el seguimiento de objetos en movimiento. Durante los seguimientos suaves, el sistema visual y motor trabajan juntos para mantener la fijación en un objeto en movimiento y seguirlo con precisión [62].

En cuanto a su relación con la atención visual, los seguimientos suaves pueden proporcionar información útil sobre la capacidad de la atención selectiva. La atención selectiva es la capacidad de enfocar la atención en un objeto específico y excluir distracciones. los seguimientos suaves son una medida de la capacidad de la persona para seguir un objeto en movimiento mientras mantiene la atención selectiva [65].

Estos movimientos pueden ser utilizados para evaluar la capacidad de la atención visual y para identificar posibles trastornos de atención. Además, la capacidad de seguir objetos en movimiento con precisión puede ser importante para tareas visuales complejas, como la conducción o el deporte, por lo que el estudio de los seguimientos suaves también puede tener aplicaciones prácticas en estos campos [65].

Atención Enfocada

En la atención enfocada, las seguimientos suaves pueden ser utilizadas para seguir un objeto en movimiento en una tarea específica, como seguir un objeto en movimiento en un videojuego [66]. La capacidad de realizar persecuciones suaves precisas y estables es un indicador de una buena atención enfocada, ya que requiere una alta precisión y un seguimiento constante del objeto en movimiento [67]. La velocidad promedio suele ser de alrededor de 50-100 grados por segundo, con una duración promedio de alrededor de 100-200 milisegundos[68, 66]. Es importante tener en cuenta que estos valores pueden variar dependiendo de la tarea específica y de las características individuales del sujeto.

Además, los seguimientos suaves también se utilizan para la detección temprana de trastornos neurológicos, como el déficit de atención y la hiperactividad (TDAH) [69], la esclerosis múltiple [70] y la enfermedad de Parkinson [71]. En estos casos, las alteraciones en los seguimientos suaves pueden ser un indicador temprano de cambios en la función neurológica relacionados con la atención visual. Por lo tanto, el estudio de los seguimientos suaves en la

atención enfocada puede tener implicaciones clínicas importantes en la detección temprana de trastornos neurológicos relacionados con la atención visual [62].

Atención Selectiva

Los seguimientos suaves también son relevantes en la atención selectiva, ya que pueden reflejar la capacidad del individuo para seguir un objeto específico en presencia de distracciones o estímulos irrelevantes. Por ejemplo, un individuo que puede seguir un objeto en movimiento a través de un entorno lleno de distracciones y estímulos irrelevantes tendría un mejor desempeño en una tarea de atención selectiva que aquellos que tienen dificultades para mantener el enfoque [72]. Durante la atención selectiva, suelen ser menos frecuentes que durante la atención dividida o alternada, ya que la atención se enfoca en un estímulo específico [65]. La duración promedio de los seguimientos suaves durante la atención selectiva puede ser de alrededor de 300 a 600 milisegundos, con una velocidad promedio de 30 a 60 grados por segundo [73]. Además, los seguimientos suaves también pueden ser utilizadas para medir la flexibilidad cognitiva y la capacidad de cambio entre tareas en una tarea de atención selectiva. Un individuo que puede cambiar rápidamente su enfoque entre diferentes estímulos en una tarea de atención selectiva tendría un mejor desempeño en la tarea y se consideraría más flexible cognitivamente que aquellos que tienen dificultades para cambiar de enfoque de manera efectiva [74].

Atención Sostenida

Se ha encontrado que la duración y la frecuencia de los seguimientos suaves están relacionadas con la capacidad de atención sostenida. Por ejemplo, en un estudio se encontró que los participantes que mostraron una mayor frecuencia de persecuciones suaves también tuvieron un mejor rendimiento en una tarea de atención sostenida, lo que sugiere que los seguimientos suaves pueden ser un indicador útil de la capacidad de atención sostenida. Además, también se ha encontrado que la falta de persecuciones suaves puede ser un indicador de fatiga visual y mental durante una tarea de atención sostenida. Durante la atención sostenida, pueden ser más frecuentes que durante la atención selectiva, ya que el ojo puede seguir el movimiento de un objeto o estímulo que se mueve en la pantalla. La duración promedio de los seguimientos suaves durante la atención sostenida puede ser de alrededor de 300 a 600 milisegundos, con una velocidad promedio de 30 a 60 grados por segundo [56].

Atención Alternada

los seguimientos suaves son un tipo de movimiento ocular que se produce cuando la persona sigue un objeto en movimiento de forma suave y continua. En la atención alternada, los seguimientos suaves también son importantes ya que permiten a la persona seguir un objeto o tarea en movimiento de forma efectiva mientras cambia su atención a otra tarea o estímulo [62].

Los seguimientos suaves pueden proporcionar información útil sobre la capacidad de la atención alternada [66]. Por ejemplo, las personas con trastornos de atención pueden tener dificultades para seguir objetos en movimiento de forma suave y continua, lo que puede indicar una capacidad reducida para mantener la atención en una tarea mientras se realiza una tarea diferente [75, 76]. Además, los estudios han demostrado que las personas con trastornos de atención pueden mostrar patrones de persecuciones suaves atípicos [75], como una menor velocidad o precisión en el seguimiento de objetos en movimiento. Por lo tanto, la medición y análisis de los seguimientos suaves durante la atención alternada puede ser útil para identificar trastornos de atención y mejorar la eficiencia en la atención alternada [77, 78]

Atención Dividida

En los seguimientos suaves en la atención dividida, se ha observado que la velocidad promedio es mayor que en la atención sostenida o enfocada. Además, la duración promedio de los seguimientos suaves también es menor. Esto puede indicar una mayor dificultad en mantener la atención en un objeto en particular mientras se procesa información de otras fuentes simultáneamente. En general, la atención dividida se asocia con una mayor frecuencia de movimientos oculares y una mayor velocidad de procesamiento visual [77].

La Tabla 3.2 muestra las características de los tres movimientos oculares relacionadas con los cinco tipos de atención visual mencionadas en párrafos anteriores.

Atención	Fijaciones	Mov. Sacádicos	Persecuciones Suaves
Selectiva	200 - 500 ms 50 - 100 g/s	20 - 50 ms 300 - 600 g/s	300 - 600 ms 30 - 60 g/s
Enfocada	200 - 500 ms 30 - 60 g/s:	20 - 40 ms 300 - 500 g/s	200 - 500 ms 30 - 40 g/s
Sostenida	300 - 800 ms 50 - 100 g/s	20 - 50 ms 300 - 600 g/s	300 - 600 ms 30 - 60 g/s
Alternada	200 - 300 ms 250 - 300 g/s	15 - 25 ms 500 - 800 g/s	100 - 200 ms 20 - 40 g/s
Dividida	200 - 250 ms 200 - 250 g/s	10 - 15 ms 400 - 600 g/s	200 - 400 ms 20 - 30 g/s

Tabla 3.2: Propiedades físicas de los movimientos oculares. La primera fila de cada movimiento indica la duración, mientras que la segunda es la velocidad en g/s.

3.5. Etiquetado

Se plantea entonces un etiquetado basado dos aspectos: naturaleza de estímulos visuales y las propiedades de los movimientos oculares ejecutados. Los estímulos visuales utilizados

en pruebas de evaluación cognitiva en donde se registran movimientos oculares: estáticos y dinámicos. Las pruebas cognitivas con estímulos estáticos y dinámicos requieren de una atención visual diferente debido a las diferencias en la forma en que se presentan los estímulos visuales y en la naturaleza del procesamiento visual que se requiere para cada tipo de prueba. En pruebas estáticas, el estímulo visual es presentado de manera fija en una pantalla, y el participante debe examinarlo para detectar detalles y características específicas. Por lo tanto, se requiere una atención visual más enfocada y sostenida para poder examinar el estímulo visual con detenimiento [45]. Además, el procesamiento visual en pruebas estáticas y dinámicas involucra diferentes áreas del cerebro y diferentes procesos cognitivos. En las pruebas estáticas, el procesamiento visual se enfoca en la percepción de detalles y la identificación de patrones, mientras que en las pruebas dinámicas, se requiere una mayor atención al movimiento y al cambio de la información visual [45].

Es por lo anterior que se propone realizar una primera estimación basada en las propiedades físicas de duración y velocidad de los movimientos oculares: fijaciones, movimientos sacádicos y persecuciones suaves. Para cada tipo de estimulo visual se analizan distintos niveles de enfoque. En la Tabla 3.3 se relacionan estos aspectos y se presenta formalmente la estimación de atención como un primer acercamiento a la clasificación formal.

Nivel	Estimulo	Atención	FX	SC	SP
Bajo	Estático	Selectiva	200 - 500 ms 50 - 100 g/s	20 - 50 ms 300 - 600 g/s	300 - 600 ms 30 - 60 g/s
		Enfocada	200 - 500 ms 30 - 60 g/s:	20 - 40 ms 300 - 500 g/s	200 - 500 ms 30 - 40 g/s
Alto	Dinámico	Sostenida	300 - 800 ms 50 - 100 g/s	20 - 50 ms 300 - 600 g/s	300 - 600 ms 30 - 60 g/s
		Alternada	200 - 300 ms 250 - 300 g/s	15 - 25 ms 500 - 800 g/s	100 - 200 ms 20 - 40 g/s
		Dividida	200 - 250 ms 200 - 250 g/s	10 - 15 ms 400 - 600 g/s	200 - 400 ms 20 - 30 g/s

Tabla 3.3: Propiedades físicas de los movimientos oculares y naturaleza de estímulos en relación a los estados de atención. La primera fila de cada movimiento indica la duración, mientras que la segunda es la velocidad en g/s.

La base de datos maestra se etiqueta a través de un *script* que requiere primero del tipo de prueba: estática o dinámica. Para sujeto se extraen los eventos de los diferentes movimientos oculares y se realiza un análisis de la duración en milisegundos y velocidad en grados por segundo. A través de las métricas mostradas en la Tabla 3.3 se realiza la clasificación de cada sujeto.

Para obtener la velocidad en grados por segundo se calcula primero la correspondencia por píxel a través de la Ecuación 3.5. Esta información es proporcionada por los diseñadores de cada base de datos. El script del etiquetado toma en consideración este factor para calcular

la velocidad. Donde Δx es la diferencia entre las dos muestras de mirada horizontal y Δy es la diferencia entre dos muestras de mirada vertical, $\Delta tiempo$ es la diferencia entre las dos muestras utilizadas, v es en $grados/s$.

$$v = \frac{\sqrt{\Delta x^2 + \Delta y^2}}{\Delta tiempo} \tag{3.5}$$

Después de obtener todas las propiedades físicas de cada par de muestras, la duración y velocidad son promediadas para realizar la evaluación de acuerdo a la Tabla 3.3. Se evalúa la correspondencia de cada sujeto con esta información y se suma el porcentaje de la relación que existe entre el sujeto con el estado de atención y se etiqueta a aquel donde el porcentaje es mayor.

3.6. Preprocesamiento

El preprocesamiento de una base de datos en el aprendizaje profundo es crucial para garantizar la calidad y la eficacia del modelo. La base de datos puede contener datos incompletos, ruido, datos erróneos o datos no normalizados que pueden afectar negativamente al rendimiento del modelo. El preprocesamiento de la base de datos implica la limpieza de los datos, la normalización, la eliminación de duplicados y la transformación de los datos en un formato adecuado para el modelo. Además, el preprocesamiento de la base de datos también puede incluir la selección de características y la reducción de la dimensionalidad para reducir la complejidad de los datos y hacer que el modelo sea más eficiente. En esta contribución el preprocesamiento está compuesto por 3 etapas: Etapa I, Etapa II y Etapa III, a continuación se describe cada una de las etapas a detalle.

3.6.1. Etapa I: Limpieza e interpolación

El registro de una sesión de eye-tracking puede almacenarse en diferentes formatos, por lo que, inicialmente, es necesario transformar el archivo en un archivo separado por comas. Además, dado que la información procede de sujetos humanos, pueden producirse errores durante el registro de una sesión de eye-tracking. Por ejemplo, supongamos que hay otros estímulos visuales o distracciones en la sesión. En ese caso, los sujetos pueden apartar la mirada de la pantalla y el *eye-tracker* no podrá registrar información valiosa. Por este motivo, se realiza una limpieza inicial del archivo para eliminar los registros falsos resultantes de un error de grabación. A continuación, se eliminan del archivo de sesión todos los registros que faltan. Este procedimiento se realiza debido a la naturaleza de la transformación a flujo óptico. La traslación en las distintas posiciones en las que se situó la mirada en la pantalla no debe sufrir cambios drásticos. Por último, se interpolan veinte puntos entre cada par de coordenadas para equilibrar el número de muestras mediante una interpolación *Akima* [79]. El método de interpolación *Akima* utiliza un *subspline* continuamente diferenciable construido a partir de polinomios cúbicos a trozos. La curva resultante pasa por los puntos de datos

dados y tendrá un aspecto suave y natural. El proceso descrito anteriormente corresponde a la Figura 3.8 y al Algoritmo 1.

Algorithm 1 Preprocesamiento.

Input: Archivo de una sesión de Eye-tracking *data*.
Output: Archivo limpio e interpolado *interpolated_data*.

```
df = read_file(data)
missing_samples = df_null_values(df)
df = drop_null_values(missing_samples)
interpolated_data = create_dataframe()
for all i in range(0, len (df) - 1) do
    x = df_loc[i:i+2, 0]
    y = df_iloc[i:i+2, 1]
    akima = Akima1DInterpolator(x, y)
    x_interpolated = linspace(x[0], x[1], 20)
    y_interpolate = make_akima(x,20)
interpolated_data.append(x_interpolated, y_interpolated)
```

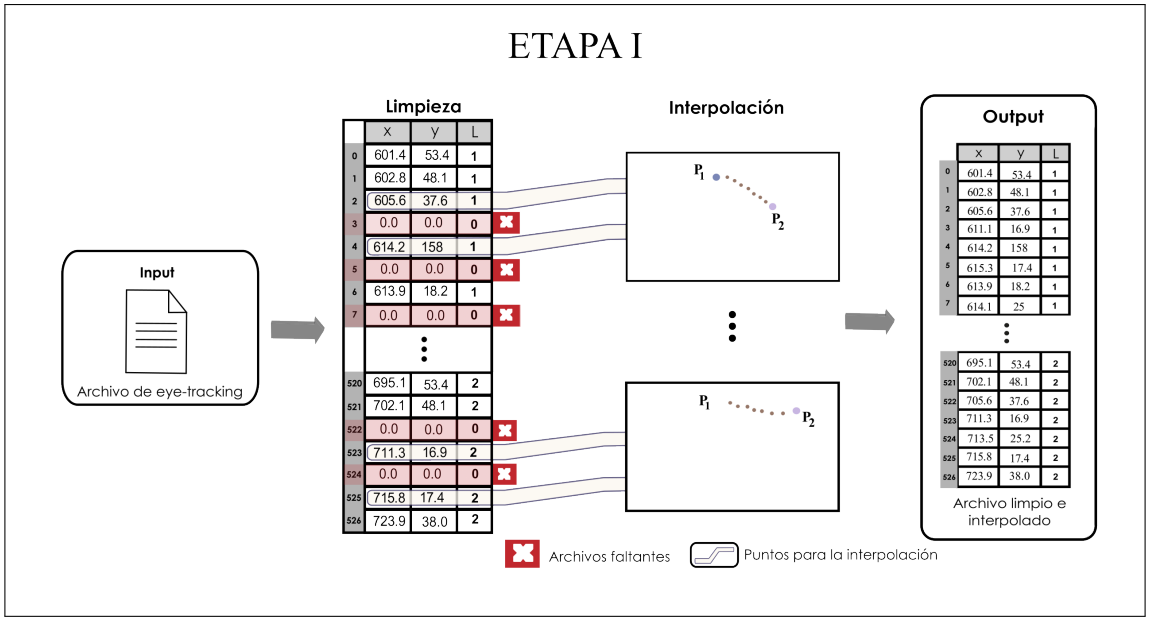


Figura 3.8: Etapa de limpieza e interpolación del archivo inicial de seguimiento ocular.

3.6.2. Etapa II: Transformación Inicial

Una vez que el conjunto de datos está limpio e interpolado se pasa a la Etapa II: transformación inicial. El fichero limpiado por el paso anterior contiene la información de las coordenadas de la mirada (x,y) y la etiqueta asignada por un experto identificada como

Label, donde 1 corresponde a la fijación y 2 a un movimiento sacádico. En primer lugar, se diseñó un algoritmo para automatizar la extracción de eventos del archivo. El algoritmo analiza la secuencia de registros: si un conjunto de muestras de una *Label* específica es continuo, el conjunto corresponde a un evento del movimiento ocular en cuestión.

Ya que el algoritmo ha identificado el evento, la información que contiene se somete a una transformación plana bidimensional. Para cada evento, se genera un marco, consistente en un círculo en las coordenadas x e y de la mirada en un lienzo del tamaño de la pantalla en la que se ha realizado la sesión de eye-tracking. El siguiente paso es concatenar todos los fotogramas generados, dando como resultado una secuencia de vídeo v_i. Con esta transformación inicial, es posible visualizar el movimiento realizado por la mirada sobre la pantalla en un vídeo; el proceso descrito en estas líneas corresponde a la Figura 3.9.

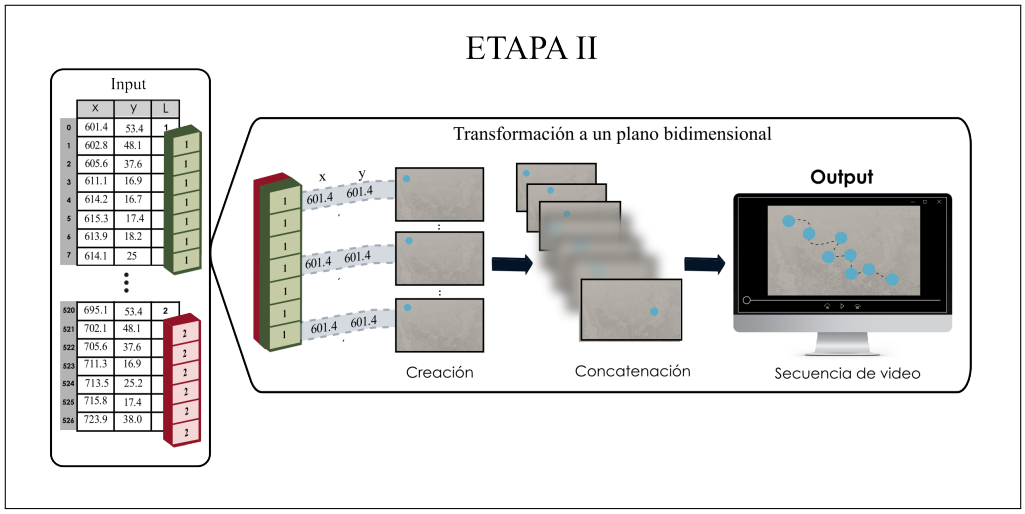


Figura 3.9: Extracción de los diferentes eventos del movimiento ocular para la transformación inicial a una secuencia de vídeo.

3.6.3. Etapa III: Cambio de dominio: Flujo Óptico

Se utiliza un dominio diferente al trabajado por propuestas anteriores, las imágenes. Esta propuesta no se ha abordado anteriormente para el análisis de movimientos oculares, y en consecuencia, para el procesamiento de la información relacionada con niveles de atencióm. Este cambio de dominio se realiza a través de la función Farneback acelerada por una tarjeta grafica dedicada contenida en la librería OpenCV. La función utiliza un modelo de seis parámetros para describir el movimiento entre las dos imágenes. Este modelo calcula el flujo óptico en cada punto de la imagen mediante una optimización iterativa para ajustar los parámetros del modelo a los datos. La función toma como entrada dos imágenes consecutivas de tamaño *(L, M, 3)*. Los tamaños de las imágenes de entrada dependen del conjunto de datos.

Tras calcular el flujo óptico, devuelve una matriz de dos canales que contiene el flujo óptico en los ejes x e y en cada punto de la imagen. Este proceso se realiza para cada par de fotogramas del vídeo generado por la transformación inicial.

Una vez obtenido el flujo óptico de todos los fotogramas, el siguiente paso es calcular la magnitud y dirección del flujo óptico a partir de los valores de flujo óptico en los ejes x e y contenidos en las matrices correspondientes. Para ello, es necesario convertir la representación de coordenadas cartesianas en una representación de coordenadas polares. A continuación, se normaliza la magnitud del flujo óptico para que sea visible en la imagen, y el último paso corresponde a la representación del flujo óptico. Esta representación puede ser mediante una escala de colores que integra dirección y magnitud o una representación vectorial de la dirección del flujo óptico a través de vectores en una imagen. Ambas representaciones se obtuvieron en esta contribución, Figura 3.10 (1) corresponde a una representación de color y (2) la dirección a través de vectores.

La representación del flujo óptico en formato HSV *(hue, saturation, value)* es una técnica para visualizar el flujo óptico en una imagen utilizando un espacio de color diferente del tradicional RGB (rojo, verde, azul). En este formato, el tono se utiliza para representar la dirección del flujo óptico, la saturación se utiliza para representar la magnitud del flujo óptico y el valor se utiliza para representar el nivel de confianza en la medición del flujo óptico. Por ejemplo, un matiz rojo puede representar un flujo óptico hacia la derecha, mientras que un matiz azul puede representar un flujo óptico hacia arriba. La saturación representa la magnitud del flujo óptico, con una saturación máxima para la mayor magnitud de flujo y una saturación 0 para la ausencia de flujo. Por último, el brillo (valor) representa la confianza en la medición del flujo óptico, con un brillo máximo para la confianza más alta y 0 para la confianza más baja. La ventaja de esta representación es que permite visualizar fácilmente el flujo óptico en una imagen porque el tono permite ver intuitivamente la dirección del flujo óptico y la saturación permite ver la magnitud del flujo óptico. Además, el brillo permite ver el nivel de confianza en la medición, la descripción del modelo de color HSV de muestra en la Figura 3.4, donde además se puede apreciar la diferencia con un modelo de color RGB, en el cual no sería posible observar todas las propiedades de la transformación obtenida por técnicas de flujo óptico.

El último paso es ajustar la representación del flujo óptico a un tamaño arbitrario de *(224, 224, 3)*. Lo descrito en las líneas anteriores se refiere a la Etapa III correspondiente al Algoritmo 2, y el resultado se puede encontrar en la Figura 3.10.

Aunque la transformación de flujo óptico ha sido ampliamente utilizada desde su propuesta, no ha sido explotada en aplicaciones relacionadas con el eye-tracking. El cambio de dominio en los datos de eye-tracking podría proporcionar información valiosa sobre nuevas características de las trayectorias de los movimientos oculares no detectadas con los métodos presentados anteriormente.

Algorithm 2 Cambio al dominio de flujo óptico.

Input: Video sequence v_i.
Output: Optical Flow HSV imagen bgr.

while true **do**
    prvs = change_BGR_to_GRAY(v_0)
    hsv = create_hsv_vector_like(v_0)
    $hsv[..., 1] = 255$
    while true **do**
        $frame$ = read_each_frame(v_{i+1})
        next = change_BGR_to_GRAY($frame$)
        flow_x,flow_y = calc_Farneback()
        mag, ang = cord_cart_to_Polar(flow_x,flow_y)
        max_mag = estimate_max_mag_gradient(mag)
        hsv= draw_hsv(flow_x,flow_y, ang, max_mag)
        bgr = change_HSV_to_BGR(hsv)
        $i = i + 1$

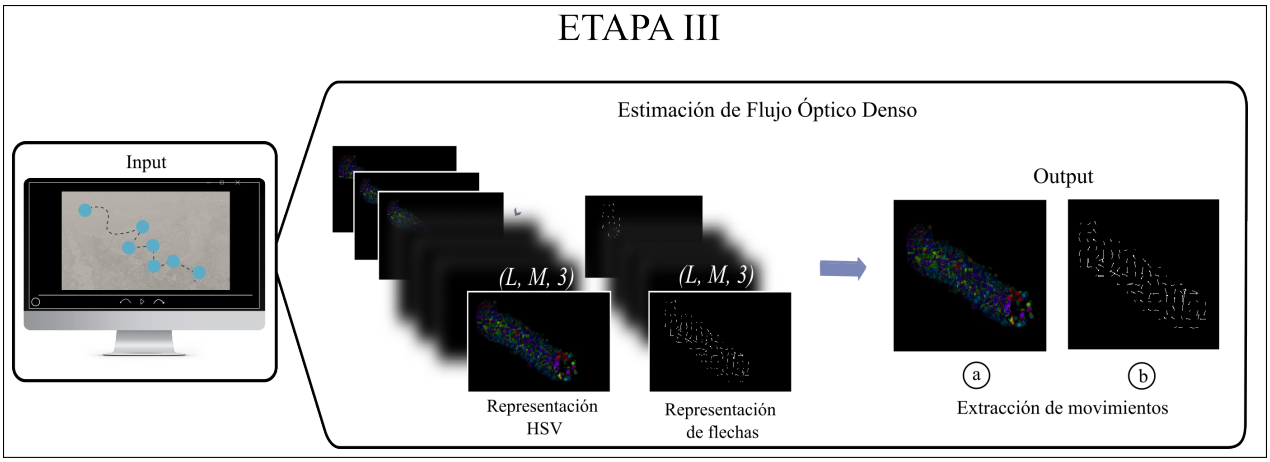


Figura 3.10: El proceso de obtención del flujo óptico denso, (a) corresponde a la representación HSV y (b) a la representación en flechas.

3.7. Clasificación

Los modelos basados en CNN tienen un punto débil: el proceso de entrenamiento con pesos aleatorios inicializados puede tardar mucho en converger y necesita la disponibilidad de un conjunto de datos a gran escala. El entrenamiento completo o desde cero de una CNN es complicado, requiere ajustes constantes de los parámetros para garantizar un aprendizaje equivalente de todas las capas y es más propenso a sobreentranamiento. El sobreentranamiento se produce cuando hay un buen ajuste de los datos de entrenamiento, pero el modelo no puede generalizarse adecuadamente a los nuevos datos. Hay varias formas de prevenir este comportamiento durante el entrenamiento de un modelo de DL, la primera opción es preparar más datos de entrenamiento; por desgracia, en las aplicaciones del mundo real a menudo no es posible preparar más debido al tiempo y al costo [80].

La clasificación de la base de datos maestra se realiza a través del procesamiento de un modelo de AP. A través de una estrategia de Transfer Learning se utiliza un modelo que ha sido entrenado en la base de datos GazeCom para la clasificación de movimientos oculares con el mismo preprocesamiento y cambio de dominio descrito en la sección previa. Se realiza esta estrategia para acelerar la convergencia del modelo en los datos. Se ha demostrado anteriormente que utilizando este tipo de estrategias es posible alcanzar resultados prometedores en pocas épocas de entrenamiento. Adicionalmente esto permite administrar los recursos computacionales disponibles.

3.7.1. Arquitectura

El modelo de clasificación es un EfficientNet-B0 [4] que ha sido previamente entrenado en el popular conjunto de datos ImageNet [81], compuesto por mil clases. La arquitectura del modelo fue elegida tras una serie de experimentos realizados con otros modelos preentrenados. Los detalles se pueden encontrar en la sección Experimentos: Entrenamiento del modelo.

EfficientNet-B0 es una arquitectura CNN y un método de escalado que escala uniformemente todas las dimensiones de profundidad/anchura/resolución utilizando un coeficiente compuesto. La base EfficientNet-B0 se basa en los bloques residuales de cuello de botella invertido de MobileNetV2 [82] y bloques de compresión y excitación. La arquitectura resultante utiliza la convolución de cuello de botella invertido móvil (MBConv), compuesta específicamente por bloques repetidos MBConv1, MBConv3 y MBConv6 que son diferentes tipos de bloques MBConv [4]. La Figura 3.11 muestra la arquitectura descrita anteriormente.

3.7.2. Entrenamiento del modelo

El entrenamiento del modelo presentado en la sección anterior conlleva un proceso complejo. Uno de los principales aspectos del AP es la cantidad de datos, el principio de la cantidad de ejemplos que se suministran al modelo durante su etapa de aprendizaje resulta vital en el desempeño del modelo. La base de datos maestra, descrita en la Sección 3.3,

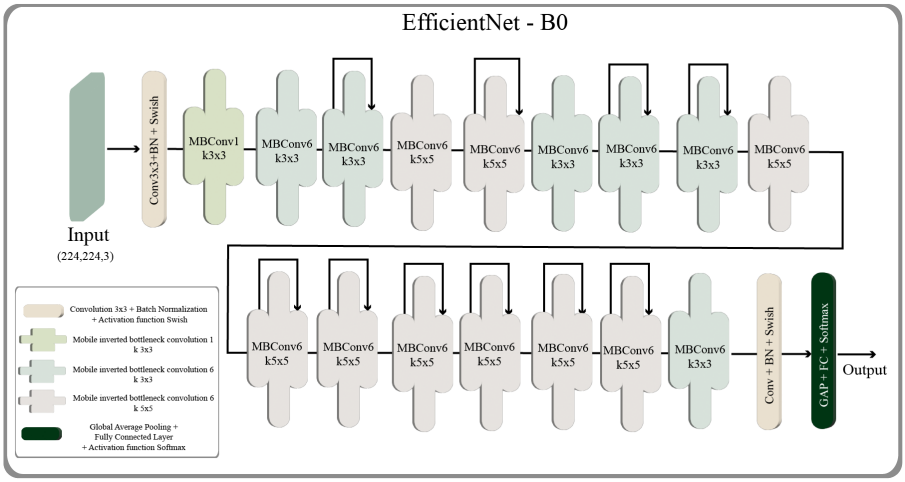


Figura 3.11: Arquitectura del modelo EfficientNet-B0, inspirado en [4]

contiene un total de 1,000 ejemplos que no resultan ser suficientes en un proceso de entrenamiento de un modelo profundo como lo es EfficientNet. Es por ello que se propone un enfoque de Transferencia de Aprendizaje (Transfer Learning) a través del entrenamiento del modelo para una tarea similar: la clasificación de eventos de movimientos oculares en el dominio de flujo óptico.

El enfoque de aprendizaje por transferencia de esta contribución consiste en utilizar el aprendizaje adquirido por las capas convolucionales durante el entrenamiento anterior y ajustar la parte densa de la arquitectura, comúnmente denominada clasificador, permitiendo que el modelo se ajuste a la naturaleza del conjunto de datos. Esta estrategia permite que el modelo se generalice en pocas épocas con un alto rendimiento de clasificación. La Figura 3.12 muestra las tres etapas de entrenamiento realizadas, en la primera etapa se realiza la clasificación de eventos de movimientos oculares. Esta es una tarea compleja en el estado puro de los datos, pues consiste el análisis de una gran cantidad de muestras a mano. Este entrenamiento se realiza con la base de datos GazeCom [1], que contiene cerca de 80,000 ejemplos. El preprocesamiento y cambio de dominio descritos en la sección previa han sido aplicados para el conjunto de datos, de manera que el modelo profundice en las características de las propiedades del dominio de flujo óptico, el extractor de características va transfiriéndose a cada etapa de entrenamiento. La siguiente etapa corresponde a la transferencia del aprendizaje obtenido en la clasificación de niveles de atención. Finalmente, la última etapa consiste en la estimación de los cinco estados de atención mencionados en secciones previas.

3.8. Recursos de Hardware

Los experimentos se implementaron en dos estaciones de computo con las siguientes características: un procesador Ryzen 5 5600g con 56 GB de RAM y una tarjeta Nvidia RTX

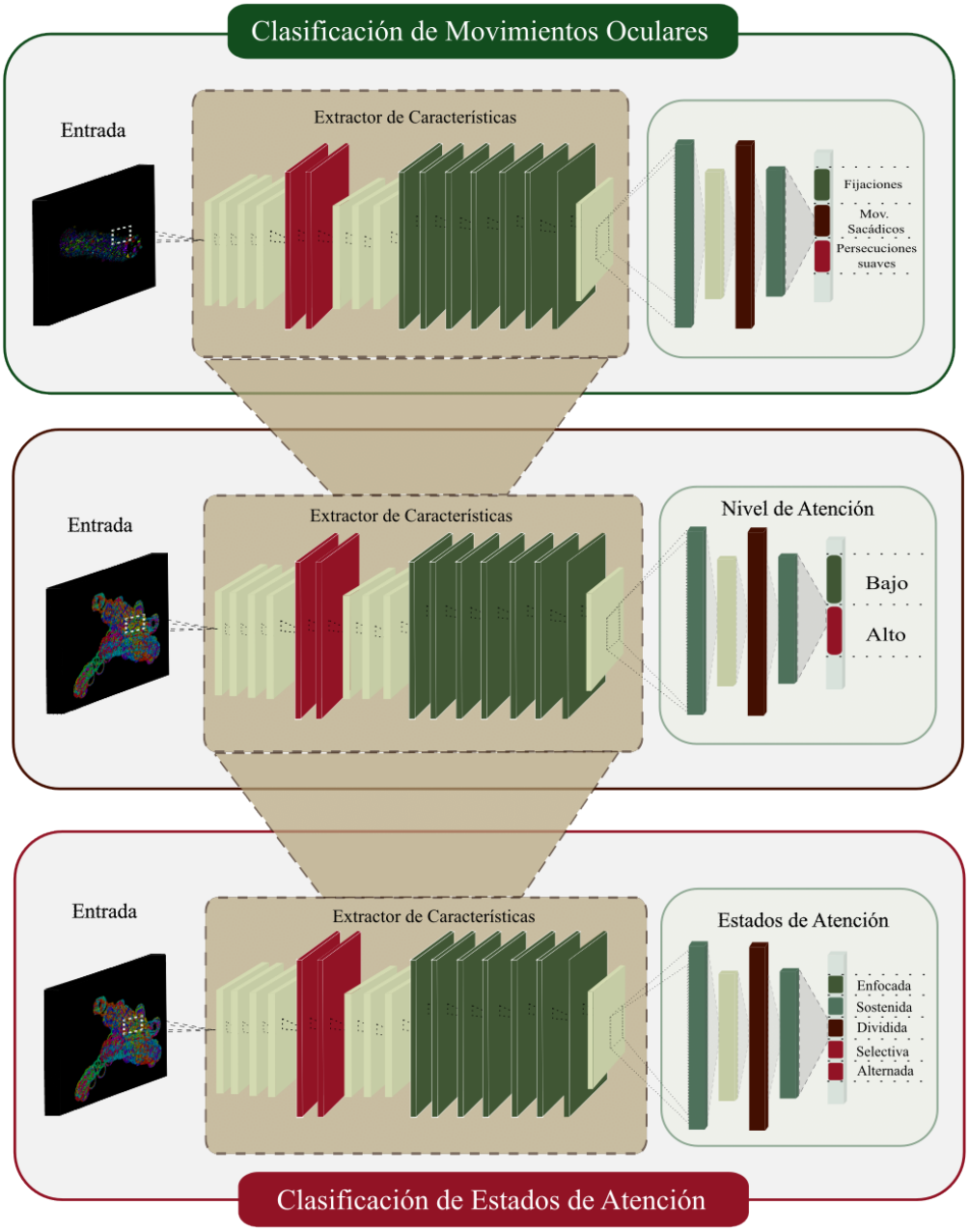


Figura 3.12: Entrenamiento del modelo en las diferentes etapas [creación propia].

3090. El sistema operativo utilizado es Linux Ubuntu 20.4, dotado del framework Python 3.10, Keras 2.3.0 y Tensorflow 2.8.1. La transformación del dominio se realiza a través de la función del framework OpenCV del enfoque Farneback. Mediante el procesamiento de la CPU, la transformación requirió una media de diez minutos para obtener el dominio de flujo óptico. Se realizó una nueva compilación de OpenCV para incluir el procesamiento mediante aceleración por GPU, por lo que fue posible acelerar la transformación. En concreto, utilizando la aceleración por GPU y el hardware descrito anteriormente, es posible obtener la transformación al dominio de flujo óptico una media de 16 veces más rápido.

3.9. Métricas de Desempeño

Una de las principales áreas de oportunidad son las métricas de rendimiento utilizadas en enfoques anteriores. Algunas propuestas sólo presentan el coeficiente kappa de Cohen para medir el rendimiento de la precisión. Esta métrica intenta corregir el sesgo de evaluación al considerar la clasificación correcta mediante una suposición aleatoria. Cuanto más difieran las distribuciones de clase objetivo predicha y real, menor será el valor kappa de Cohen máximo alcanzable. El valor kappa de Cohen máximo representa el caso límite en el que el número de falsos negativos o falsos positivos en la matriz de confusión es cero. Alcanza su máximo cuando el modelo se aplica a datos equilibrados, por lo que esta métrica expresa poco sobre la precisión de predicción esperada. El mismo modelo puede obtener valores más bajos de kappa de Cohen para datos no equilibrados que para datos equilibrados [83] [84]. Por lo tanto, es incorrecto afirmar que los modelos presentados anteriormente clasifican correctamente los movimientos oculares basándose únicamente en el comportamiento del coeficiente kappa de Cohen.

Es necesario utilizar métricas de rendimiento para determinar con precisión la eficiencia de clasificación del conjunto de datos. Aunque algunos enfoques utilizan F1-Score o la precisión, las métricas se presentan individualmente. Hasta ahora no se ha presentado un análisis completo de la precisión de los modelos clasificadores. En esta contribución, se presentan todas las métricas de rendimiento utilizadas en el campo de la investigación: exactitud, precisión, recuerdo, F1-Score, coeficiente kappa de Cohen e intersección sobre la unión (IoU).

Todas las métricas se obtienen mediante un análisis de la matriz de confusión, una herramienta para visualizar el rendimiento de un algoritmo de aprendizaje supervisado. Contiene cuatro posibles escenarios para evaluar el rendimiento del modelo: verdadero positivo (TP), verdadero negativo (TN), falso negativo (FN) y falso positivo (FP). Las ecuaciones 3.6 a 3.10 corresponden a las métricas basadas en los resultados de la matriz de confusión. El uso de diversas métricas de rendimiento permite abarcar todas las características que componen el rendimiento de un modelo de AP.

Accuracy La exactitud es una métrica que generalmente describe el rendimiento del modelo en todas las clases. Es útil cuando todas las clases tienen la misma importancia. Se calcula como la relación entre el número de predicciones correctas y el número total de predicciones. El comportamiento de esta métrica se muestra en la Ecuación 3.6.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.6}$$

Recall Recall se define como la relación entre el número de muestras Positivas correctamente clasificadas como Positivas y el número total de muestras Positivas. La recuperación mide la capacidad del modelo para detectar muestras positivas. Cuanto mayor sea, mayor será el número de muestras positivas detectadas. El comportamiento de esta métrica se muestra en la Ecuación 3.7.

$$Recall = \frac{TP}{TP + FN} \tag{3.7}$$

Precisión La precisión se define como la relación entre el número de muestras Positivas clasificadas correctamente y el número total de muestras clasificadas como Positivas (ya sea correcta o incorrectamente). La precisión mide la exactitud del modelo a la hora de clasificar una muestra como positiva. El comportamiento de esta métrica se muestra en la Ecuación 3.8.

$$Precision = \frac{TP}{TP + FP} \tag{3.8}$$

F1-Score La precisión y el recall son los dos componentes básicos de la F1-Score. El objetivo de esta métrica es combinar la precisión y recall en una única métrica. Al mismo tiempo, la F1-Score ha sido diseñada para funcionar bien con datos desequilibrados. Esta métrica también se define como la media armónica de la precisión y Recall. El comportamiento de esta métrica se muestra en la Ecuación 3.9.

$$F1 - Score - score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{3.9}$$

Intersección sobre la unión La intersección sobre la unión (IoU) se utiliza para evaluar el rendimiento de la detección de objetos comparando el cuadro delimitador real con el cuadro delimitador previsto. Se incluye como métrica de desempeño en la presente contribución ya que autores la han utilizado previamente en tareas de clasificación de movimientos oculares. El comportamiento de esta métrica se muestra en la Ecuación 3.10.

$$IoU = \frac{TP}{TP + FP + FN} \tag{3.10}$$

Coeficiente Kappa de Cohen Esta métrica mide el acuerdo entre dos evaluadores que clasifican N elementos en categorías C mutuamente excluyentes. Es una medida estadística que ajusta el efecto del azar en la proporción de la concordancia observada. Se incluye como métrica para la evaluación del desempeño del modelo ya que se utiliza en el estado del arte, de manera que la comparativa de esta contribución no estaría completa. La Ecuación 3.11 resume el comportamiento de esta métrica, donde *Pr(a)* es el acuerdo observado relativo entre los observadores, y Pr(e) es la probabilidad hipotética de acuerdo por azar, utilizando los datos observados para calcular las probabilidades de que cada observador clasifique aleatoriamente cada categoría. Si los evaluadores están completamente de acuerdo, entonces $\kappa = 1$. Si no hay acuerdo entre los calificadores distinto al que cabría esperar por azar (según lo definido por *Pr(e)*),$\kappa = 0$. La Tabla 3.4 muestra la relación entre el valor del coeficiente kappa (κ) y el grado de acuerdo entre las predicciones de un modelo y etiquetas reales.

$$\kappa = \frac{Pr(a) - Pr(b)}{1 - Pr(e)} \tag{3.11}$$

Kappa κ	Estimación del grado de acuerdo
0	No de acuerdo
0.0 - 0.2	Insignificante
0.2 - 0.4	Bajo
0.4 - 0.6	Moderado
0.6 - 0.8	Bueno
0.8 - 1.0	Muy bueno

Tabla 3.4: Relación entre el valor del Coeficiente Kappa y el nivel de acuerdo entre las muestras reales y predicciones.

CAPÍTULO 4

Resultados y Discusión

En esta sección, se presentan los hallazgos y conclusiones que surgieron como resultado del procesamiento y análisis de los datos. Los resultados que se presentan son fundamentales para cumplir con los objetivos planteados y la hipótesis formulada en esta investigación. Además, la interpretación y discusión de los resultados permiten obtener una comprensión más profunda del tema de estudio y proporcionan nuevas perspectivas que pueden ser exploradas en futuros estudios. Por lo tanto, esta sección es crucial para demostrar la contribución de la investigación al conocimiento existente y la importancia de los resultados obtenidos. Esta sección presenta los resultados alcanzados durante la experimentación de la metodología planteada. Se presenta por separado el resultado de las tres etapas de entrenamiento del modelo.

4.1. Resultados

4.1.1. Clasificación de Movimientos Oculares

El primer entrenamiento del modelo, descrito en la Figura 4.1 consistió en la clasificación de eventos de movimientos oculares con el base de datos GazeCom, con 79,983 ejemplos. Este proceso sirve como *calentamiento* o *warmup* al modelo, ya que para el problema principal existen pocos ejemplos, es necesario adoptar estrategias que proporcione un mejor rendimiento en una muestra tan pequeña de ejemplos. Esta estrategia ayuda a la red neuronal a ajustar su tasa de aprendizaje de manera gradual y eficiente durante el entrenamiento, lo que puede mejorar el rendimiento y la estabilidad del modelo para el problema principal.

Para demostrar la capacidad del método propuesto, se presentan comparaciones con tres métodos de vanguardia utilizando GazeCom como conjunto de datos principal. El primer método, [31], utilizó un modelo CNN 1D combinado con BLSTM para lograr el aprendizaje secuencia a secuencia de extremo a extremo. El segundo [85], presentó gazeNet, el modelo más popular para la clasificación de movimientos oculares. Por último, [86] utilizó una red convolucional temporal (TCNs). Estas comparaciones permiten determinar el impacto del

cambio de dominio propuesto en las métricas de rendimiento.

Como se muestra en la Tabla 4.1, el método propuesto alcanza los valores más altos para todas las métricas presentadas por las contribuciones anteriores. El rendimiento comparado con los enfoques anteriores demuestra que el cambio de dominio aplicado al conjunto de datos GazeCom consigue mejores resultados en la clasificación de eventos de movimientos oculares.

Figura 4.1: Primera etapa de entrenamiento del modelo: clasificación de eventos de movimientos oculares [creación propia].

Tabla 4.1: Comparación del rendimiento de los métodos más avanzados. Los marcados con * corresponden a contribuciones con el base de datos GazeCom, y los marcados con + corresponden a contribuciones con el base de datos Lund2013. Las métricas en negrita corresponden a las puntuaciones más altas.

	Fijaciones						Movimientos Sacádicos					
Contribution	Acc	Prec	Recall	F1	IoU	Kappa	Acc	Prec	Recall	F1	IoU	Kappa
[31]*	-	-	-	93.9 %	-	-	-	-	-	89.3 %	85.85	-
[85]*	-	-	-	-	-	0.915	-	-	-	-	-	0.845
[86]*	-	92.9 %	96.1 %	94.5 %	-	-	-	89.9 %	88.8 %	89.4 %	-	-
[87]*	-	-	-	93.60 %	-	-	-	-	-	**97.16 %**	-	-
Propuesta*	**96.61 %**	**95.93 %**	**97.31 %**	**96.63 %**	**93.45 %**	**0.931**	**96.61 %**	**97.31 %**	**95.85 %**	96.59 %	**93.38 %**	**0.9331**
[85]+	-	-	-	-	-	0.959	-	-	-	-	-	0.947
[88]+	-	-	-	-	-	0.936	-	-	-	-	-	0.940
Propuesta+	**99.29 %**	**99.17 %**	**99.35 %**	**99.26 %**	**98.53 %**	**0.985**	**99.28 %**	**99.29 %**	**99.26 %**	**99.26 %**	**98.56 %**	**0.985**

4.1.2. Clasificación de Niveles de Atención

La segunda etapa de entrenamiento del modelo consta de la clasificación entre niveles de atención, de acuerdo a la Tabla 3.3, existen dos niveles de atención generales: alto y bajo, de los cuales se desglosan los cinco estados de atención. La base de datos contiene 981 ejemplos para ambas clases. La arquitectura del modelo es la misma mostrada en la Figura 3.11, utilizando los pesos de la etapa de entrenamiento previa, se realiza la clasificación de los dos niveles de atención. En la Figura 4.2 se muestra cómo es que el extractor de características basado en el modelo EfficientNet-B0 se mantiene para esta nueva clasificación, al igual que

en la etapa previa, la parte convolucional se congela para conservar el aprendizaje alcanzado.

La base de datos de esta etapa de entrenamiento descrita en la Sección 3.3 conformada por tres diferentes conjuntos, no ha sido utilizada anteriormente para la clasificación de atención. Como se explicó en previamente, se realiza el etiquetado de la base de datos de acuerdo con las métricas de los diferentes movimientos oculares, basados en propiedades físicas analizadas por la literatura mencionada en la Sección 3.4. Lo anterior provoca que no sea posible proporcionar una comparativa en los mismos términos para las diferentes contribuciones mencionadas en la Sección 2.

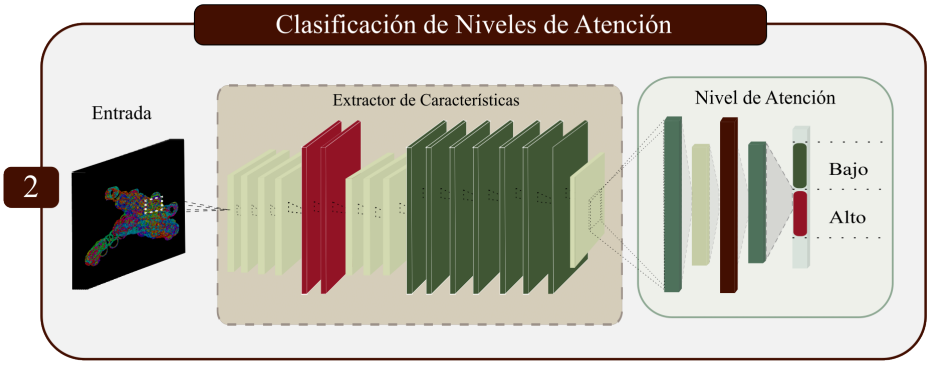


Figura 4.2: Segunda etapa de entrenamiento del modelo: clasificación de niveles de atención [creación propia].

Las métricas de desempeño para evaluar el modelo son las mismas que en el entrenamiento previo. Es importante demostrar el rendimiento de esta etapa de entrenamiento ya que nuevamente el aprendizaje alcanzado en este proceso de clasificación será transferido una ultima vez para clasificar los cinco estados de atención. En el campo del aprendizaje profundo, es fundamental utilizar varias métricas de desempeño para evaluar la calidad de los modelos. El uso de una sola métrica puede ser insuficiente para capturar la complejidad de un modelo y puede llevar a conclusiones erróneas sobre su desempeño. Por ejemplo, una métrica como la precisión puede ser engañosa si el conjunto de datos está desequilibrado. En este caso, un modelo puede obtener una alta precisión simplemente prediciendo siempre la clase mayoritaria. Por lo tanto, es importante utilizar una variedad de métricas, como la precisión, recall, F1-Score y el coeficiente kappa de Cohen, para obtener una comprensión más completa del desempeño del modelo en diferentes aspectos. Además, el uso de varias métricas también puede ayudar a detectar posibles problemas en el modelo, como sobreentrenamiento, y proporcionar información valiosa para mejorar el modelo en futuras iteraciones.

El modelo fue entrenado a través del uso de Google Colab, con una configuración de dispositivo distinta a la descrita en la Sección 3.8. Derivado de la cantidad de datos recolectados en la base de datos, no es necesario la una configuración de hardware tan considerable como

lo utilizado en la primera etapa de entrenamiento. Los recursos necesarios para la clasificación de este problema se cubren con la suscripción de Google Colab Pro con GPU's de uso medio. Se condujeron cinco experimentos o *trials* para asegurar la generalización correcta del modelo. Estos experimentos consisten en un *5-fold-cross-validation*. En la Tabla 4.2 se muestran los promedios por métrica de cada experimento realizado, las métricas más altas se presentan en negritas.

Tabla 4.2: Métricas alcanzadas por el modelo durante cinco pruebas realizadas y el promedio de cada métrica.

	Nivel Alto					Nivel Bajo				
Trial	Acc	Prec	Recall	F1	Kappa	Acc	Prec	Recall	F1	Kappa
1	85.02 %	81.50 %	79.83 %	83.84 %	0.6539	85.02 %	84.28 %	85.75 %	84.87 %	0.6539
2	85.30 %	**89.28 %**	79.32 %	83.84 %	0.7029	85.30 %	82.55 %	**90.62 %**	86.33 %	0.7029
3	85.74 %	82.89 %	**85.74 %**	83.58 %	0.7096	85.74 %	87.98 %	86.90 %	87.32 %	0.7096
4	**87.90 %**	87.94 %	85.71 %	**86.62 %**	**0.7548**	87.90 %	88.03 %	89.89 %	88.81 %	0.7548
5	87.72 %	86.05 %	79.96 %	83.56 %	0.7055	87.72 %	82.60 %	90.35 %	86.89 %	0.7548
Promedio	85.93 %	85.53 %	82.11 %	84.29 %	0.7053	85.93 %	85.09 %	88.70 %	86.84 %	0.7053

4.1.3. Estimación de Estados de Atención

A pesar de la importancia de la atención en el rendimiento del usuario, los métodos actuales de clasificación de atención no permiten discriminar entre diferentes tipos de atención. Existe una limitación entre la estimación de atención que se ha realizado previamente con información obtenida de sesiones de seguimiento ocular, y es que se limitan a clasificar entre "niveles"de atención ambiguos, tales como bajo-medio-alto o simplemente atención y no atención. Únicamente una contribución previa a esta ha intentado diversificar entre los modelos de atención presentados. Específicamente hablando del modelo clínico de atención de Sohlberg y Mateer, Abdelrahman et al. [30] realizan una clasificación de cuatro estados de atención, utilizando seguimiento ocular e imágenes térmicas obtenidas durante las sesiones de seguimiento ocular. Ya que la información de este experimento no se encuentra disponible, específicamente la base de datos, no es posible realizar una comparativa entre este método y el propuesto por la contribución antes mencionada.

En la fase final del entrenamiento del modelo, se estiman los niveles de atención descritos en el modelo clínico de Solberg y Mateer [18]. La clasificación de los niveles de atención en categorías alta y baja pretende identificar diferentes comportamientos atencionales. Dos tipos de comportamientos se asocian al nivel de concentración bajo: la atención focalizada y la selectiva. En el nivel de concentración alto, se observa atención sostenida, dividida y alternada, siendo la última la más difícil de clasificar. El conjunto de datos utilizado para esta clasificación es el mismo de la etapa dos del entrenamiento del modelo. Existe una modificación entre las clases, ya que se pretende partir hacia lo específico, los 981 ejemplos se distribuyen en cinco clases. En la Figura 4.3 se muestra cómo es que el extractor de características basado en el modelo EfficientNet-B0 se mantiene para esta nueva clasificación y únicamente se modifica el clasificador para las cinco etiquetas de este problema de clasificación.

Figura 4.3: Tercera y última etapa de entrenamiento del modelo: estimación de estados de atención [creación propia].

La transferencia del aprendizaje se realiza para dos subentrenamientos en paralelo. Del conjunto de datos de la etapa dos, el primer subentrenamiento será el encargado de diferenciar entre los dos estados de atención del nivel bajo. Mientras que el otro subentrenamiento se entrenará para diferenciar entre los tres estados de atención del nivel alto. La experimentación realizada arroja que derivado de los pocos ejemplos encontrados del estado de atención alternada, el modelo no es capaz de procesar este comportamiento de atención, por lo que no se mostrarán resultados de la estimación de atención alternada.

Los subentrenamientos fueron realizados a través del uso de Google Colab, con una configuración de dispositivo distinta a la descrita en la Sección 3.8. Derivado de la cantidad de datos recolectados en la base de datos, no es necesario la una configuración de hardware de la primera etapa de entrenamiento. Los recursos necesarios para la clasificación de este problema se cubren con la suscripción de Google Colab Pro. Los resultados alcanzados para la clasificación de atención enfocada y atención selectiva se presentan en la Tabla 4.3, mientras que la Tabla 4.4 muestra los resultados de la estimación de atención sostenida y dividida.

Tabla 4.3: Métricas alcanzadas por el modelo en la estimación del nivel de atención bajo: atención enfocada y atención selectiva, durante cinco pruebas realizadas.

	Atención Enfocada					Atención Selectiva				
Trial	Acc	Prec	Recall	F1	Kappa	Acc	Prec	Recall	F1	Kappa
1	79.36 %	71.42 %	79.83 %	83.84 %	0.6539	79.36 %	80.00 %	85.71 %	82.75 %	0.6539
2	80.64 %	67.14 %	79.32 %	83.84 %	0.6012	80.64 %	81.25 %	92.85 %	86.66 %	0.6012
3	70.06 %	63.33 %	85.74 %	83.58 %	0.6754	70.06 %	74.46 %	83.33 %	78.65 %	0.6754
4	75.80 %	60.00 %	85.71 %	86.62 %	0.6433	75.80 %	82.97 %	84.78 %	83.87 %	0.6433
5	77.41 %	61.53 %	79.96 %	83.56 %	0.6529	77.41 %	81.25 %	86.66 %	83.87 %	0.6529
Media	77.30 %	64.68 %	82.11 %	84.29 %	0.6453	77.30 %	85.09 %	88.70 %	86.84 %	0.6453
std	4.12 %	4.61 %	3.31 %	1.31 %	0.027	4.12 %	3.26 %	3.67 %	2.91 %	0.027

Tabla 4.4: Métricas alcanzadas por el modelo en la estimación del nivel de atención alto: atención sostenida y atención dividida, durante cinco pruebas realizadas.

	Atención Sostenida					Atención Dividida				
Trial	Acc	Prec	Recall	F1	Kappa	Acc	Prec	Recall	F1	Kappa
1	94.05 %	94.05 %	100 %	96.63 %	0.6921	94.05 %	81.50 %	79.83 %	83.84 %	0.6921
2	99.00 %	98.00 %	98.98 %	98.98 %	0.7421	99.00 %	89.28 %	79.32 %	83.84 %	0.7421
3	98.01 %	98.01 %	100 %	99.00 %	0.7065	98.01 %	82.89 %	85.74 %	83.58 %	0.6754
4	95.04 %	95.04 %	100 %	97.46 %	0.6932	95.04 %	87.94 %	85.71 %	86.62 %	0.6932
5	99.00 %	99.00 %	100 %	99.49 %	0.7014	99.00 %	85.53 %	82.11 %	84.29 %	0.7014
Media	97.02 %	96.82 %	99.80 %	98.37 %	0.7008	97.02 %	85.42 %	82.54 %	84.43 %	0.7008
std	2.32 %	2.14 %	0.45 %	1.21 %	0.024	2.32 %	3.27 %	3.27 %	1.24 %	0.024

4.2. Discusión

En esta sección se presentan los resultados de los experimentos realizados. Como ya se ha mencionado, se entrenaron cincuenta modelos para garantizar la generalización, por lo que se presentan las métricas promediadas del conjunto de pruebas realizadas. Los resultados de la discusión pretenden demostrar que el método propuesto alcanza un rendimiento superior a las contribuciones del estado del arte en la primera etapa de entrenamiento: Clasificación de movimientos oculares. Para la segunda y tercera etapa de entrenamiento no se cuentan con comparativas disponibles, ya que el conjunto de datos fue generado en esta contribución no existen contribuciones externas que utilicen este conjunto como base de datos principal para las dos tareas: clasificación de niveles cognitivos y estimación de estados de atención. Por lo anterior en esta sección se presentan una serie de ensayos realizados a la metodología para demostrar la eficacia de la metodología planteada.

4.2.1. Clasificación de Movimientos Oculares

Para determinar la eficacia del modelo se llevó a cabo una serie de ensayos compuestos por una validación cruzada quíntuple. En concreto, se realizaron diez ensayos mediante el proceso de reordenar el conjunto de datos y generar nuevas validaciones cruzadas quíntuples en cada ensayo, esto se propuso para garantizar la coherencia de los resultados. Los resultados marcados con un $*$ en la Tabla 4.1 corresponden a los obtenidos en la clasificación del conjunto de datos GazeCom para fijaciones y sacadas, y consisten en la media de los resultados de los diez ensayos, que se muestran en la Tabla 4.1. Para demostrar la generalización independientemente de la naturaleza del *eye-tracker* utilizado en la generación del conjunto de datos, los resultados obtenidos en la clasificación del conjunto de datos Lund2013 también se presentan en la Tabla 4.1, marcados con un $+$. Se puede observar consistencia en los resultados a lo largo de los diez ensayos generados, y este comportamiento demuestra

estadísticamente la correcta generalización del modelo en los conjuntos de datos.

La precisión en ambas clases muestra un rendimiento robusto que oscila entre un valor mínimo de 96,50 %, con una desviación estándar de 0,00054 para todos los ensayos. Este rendimiento indica que, independientemente del orden de los datos de entrenamiento y prueba, la metodología propuesta puede predecir con precisión tanto las fijaciones como los movimientos sacádicos.

El coeficiente Kappa de Cohen muestra un mínimo de 0,9300 y un máximo de 0,9401, con una desviación estándar de 0,001. Dado que la métrica mide la fiabilidad y el acuerdo de dos evaluadores que valoran las cantidades exactas, estos resultados indican que el índice de acuerdo de 0,94 es lo suficientemente alto como para que tanto las fijaciones como los movimientos sacádicos se consideren fiables en su clasificación. Cabe destacar que la precisión de cada ensayo y el coeficiente kappa de Cohen son iguales. Puesto que la precisión es el rendimiento global de todas las clases y el coeficiente kappa de Cohen indica el acuerdo entre clases, cabe esperar un comportamiento idéntico. La precisión es coherente con ambos movimientos oculares independientemente de los datos utilizados para el entrenamiento y las pruebas, ya que los resultados muestran un comportamiento similar en ambas clases. El rendimiento de las métricas a través de las diferentes pruebas garantiza que la metodología propuesta puede clasificar con precisión los movimientos oculares y minimizar el número de falsos positivos y negativos.

Se utilizan gráficos de violín para mostrar las métricas alcanzadas a través de las diez pruebas. Se utiliza esta estrategia gráfica ya que muestra la densidad de probabilidad de los datos. La Figura 4.4 muestra la distribución completa de los valores obtenidos para cada métrica utilizada en la clasificación del conjunto de datos GazeCom. A través del análisis de la distribución de cada métrica, es posible asegurar la fiabilidad del rendimiento alcanzado por la presente contribución. En conjunto, los resultados muestran una concordancia mínima del 96 %, lo que indica la viabilidad y solidez de la contribución.

Un análisis posterior en un conjunto de diferentes modelos es realizado para demostrar la consistencia de la metodología. En concreto, se utilizan cinco modelos entrenados para la clasificación de ImageNet: ResNet50 [89], MobilenetV2 [82], EfficientNetV2 [90], DenseNet [91], y EfficientNet-B0 se entrenaron con validación cruzada con $k = 5$. Estos experimentos se llevaron a cabo en las dos estaciones de computo mencionadas en la Sección 3.8, con una media de 30 horas por experimento. En total, se necesitaron 150 horas para completar los experimentos. Los resultados demuestran el rendimiento superior del modelo EfficientNet-B0, es por ello que se utilizo como modelo base para la estrategia de transferencia de aprendizaje continua en las dos etapas de entrenamiento subsecuentes. La Figura 4.5 muestra las métricas utilizadas para determinar el rendimiento de los cinco modelos en un boxplot para una mejor visualización.

Durante el entrenamiento de un modelo CNN las gráficas que muestran las métricas de desempeño, como el accuracy, pueden mostrar patrones característicos de sobreentrenamien-

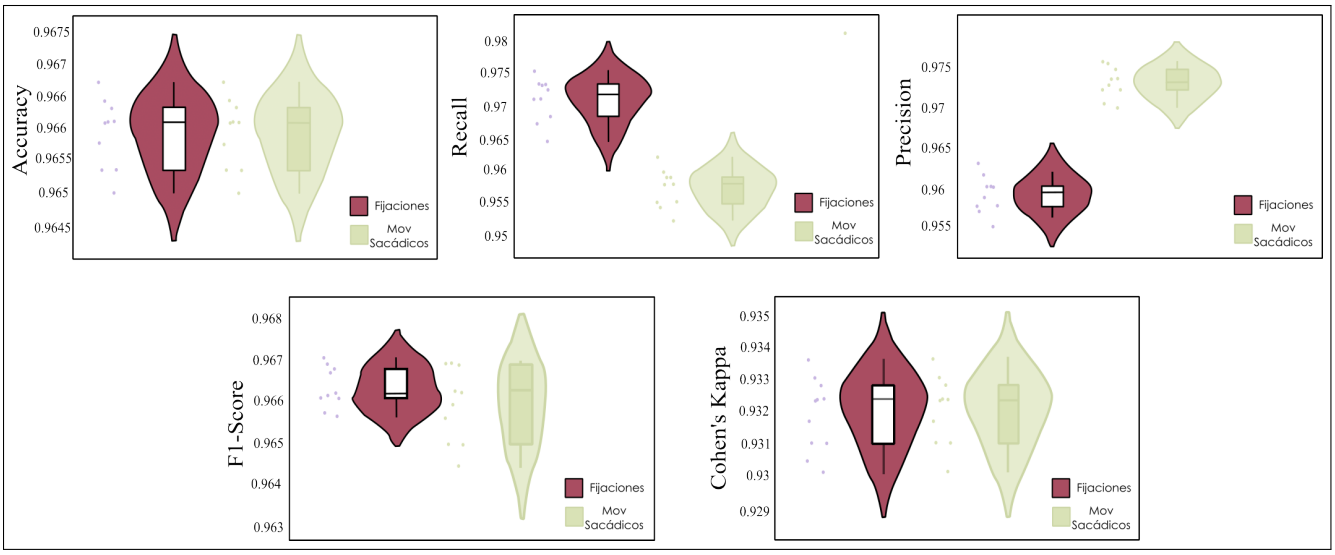


Figura 4.4: Resultados de las métricas de rendimiento en los diez ensayos ejecutados en la clasificación del conjunto de datos GazeCom.

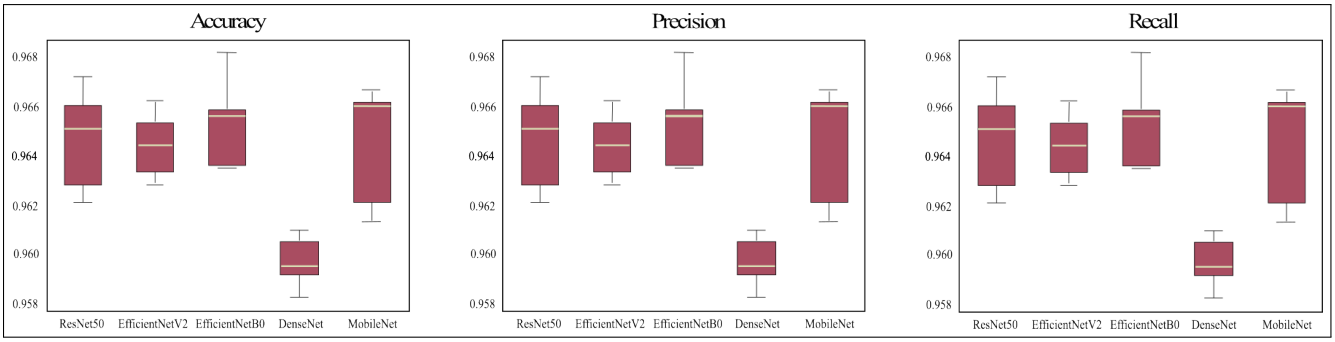


Figura 4.5: Comportamiento de los cinco modelos para tres métricas: Accuracy, Precisión y Recall.

to. En una situación de sobreentrenamiento, la accuracy en el conjunto de entrenamiento puede mejorar rápidamente y alcanzar niveles muy altos, mientras que la precisión en el conjunto de validación se estabiliza o incluso disminuye. Esto se puede observar en las gráficas, donde la curva de precisión en el conjunto de entrenamiento sube de manera pronunciada, mientras que la curva de precisión en el conjunto de validación se estanca o disminuye. En algunos casos, también puede haber una brecha significativa entre la precisión en el conjunto de entrenamiento y el conjunto de validación, lo que indica que el modelo está memorizando los datos de entrenamiento en lugar de aprender patrones generales que se puedan aplicar a nuevos datos. En general, el comportamiento de sobreentrenamiento se puede observar en cualquier métrica de desempeño.

La Figura 4.6 contiene las gráficas del desempeño del modelo durante dos folds del primer experimento. Contiene las tres principales métricas: accuracy, precision y recall. Las gráficas muestran una generalización distinta en cada fold, sin embargo el comportamiento de las metricas durante el entrenamiento y la validación no muestra un comportamiento de sobreentrenamiento. Durante el diseño del entrenamiento se implementaron estrategías que permitieran analizar exitosamente si el modelo comenzaba a mostrar un comportamiento de sobreentrenamiento y detenerlo cuando esto sucediera. Las curvas de entrenamiento y validación para cada fold no muestran una diferencia significativa, lo que nos permite visualizar una generalización correcta a través de las épocas. Con una cantidad tan reducida de información el sobreentrenamiento es una amenaza latente durante el entrenamiento del modelo, es por ello que se adoptan estrategías de evaluación distintas, como la validación cruzada y *callbacks* que permitan analizar el comportamiento del modelo época a época para determinar cuándo es el mejor momento para terminar. El desempeño mostrado por el modelo no tiene indicios que indiquen que los resultados alcanzados son fruto de la casualidad, además la ejecución de los diferentes experimentos permite visibilizar la robustez del modelo para cualesquiera sean los conjuntos de entrenamiento, validación y prueba. Esto nos permite asegurar una correcta generalización del modelo en la naturaleza de los datos.

4.2.2. Clasificación de Niveles de Atención

La Tabla 4.2 muestra los resultados de cinco pruebas realizadas en un modelo de clasificación, en las que se evaluó la accuracy (Acc), precisión (Prec), recall por clase (Recall), F1-Score y coeficiente kappa. La Figura 4.7 muestra el comportamiento del modelo durante dos folds del primer experimento realizado. Como se ha comentado previamente, se analiza el comportamiento de ambas curvas de cada métrica para determinar si existe sobreentrenamiento. El comportamiento mostrado en las gráficas no representa un indicio de sobreentrenamiento del modelo en los datos utilizados para el entrenamiento.

En la prueba 1, el modelo obtuvo una precisión del 85,02 %, una precisión por clase del 84,28 % para nivel alto y 85,75 % para el nivel alto y un recall por clase del 79,83 % para el nivel alto y 83,84 % para el nivel bajo. F1-Score fue del 83,84 % y el coeficiente kappa fue de 0.6539. En la prueba 2, el modelo obtuvo una precisión del 85,30 %, una precisión por clase del 82,55 % para nivel alto y 90,62 % para nivel bajo, y una recuperación por clase del

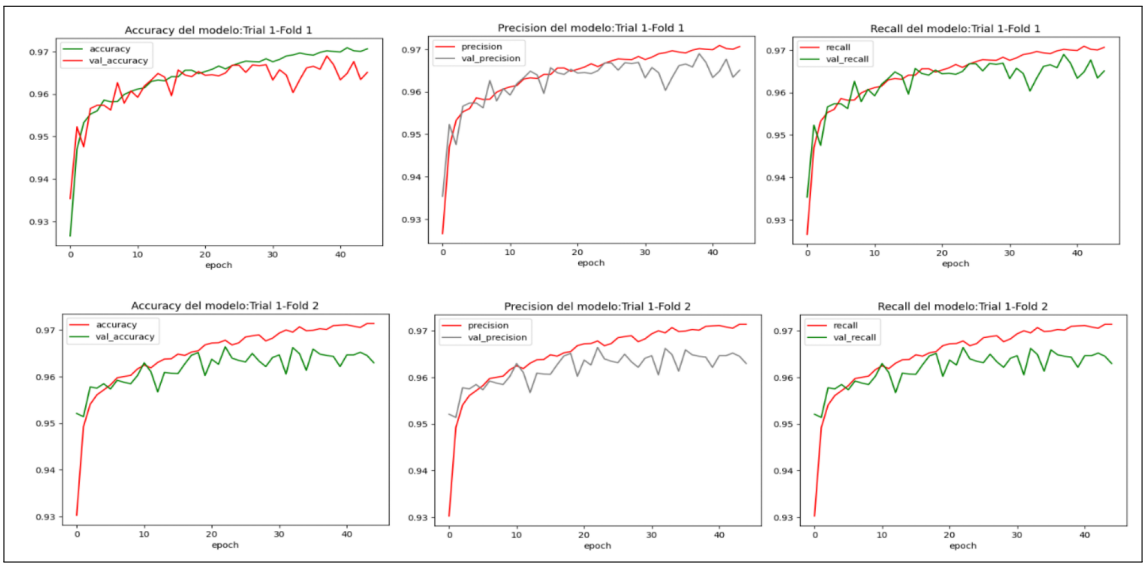


Figura 4.6: Comportamiento del modelo durante el entrenamiento para dos folds del primer experimento.

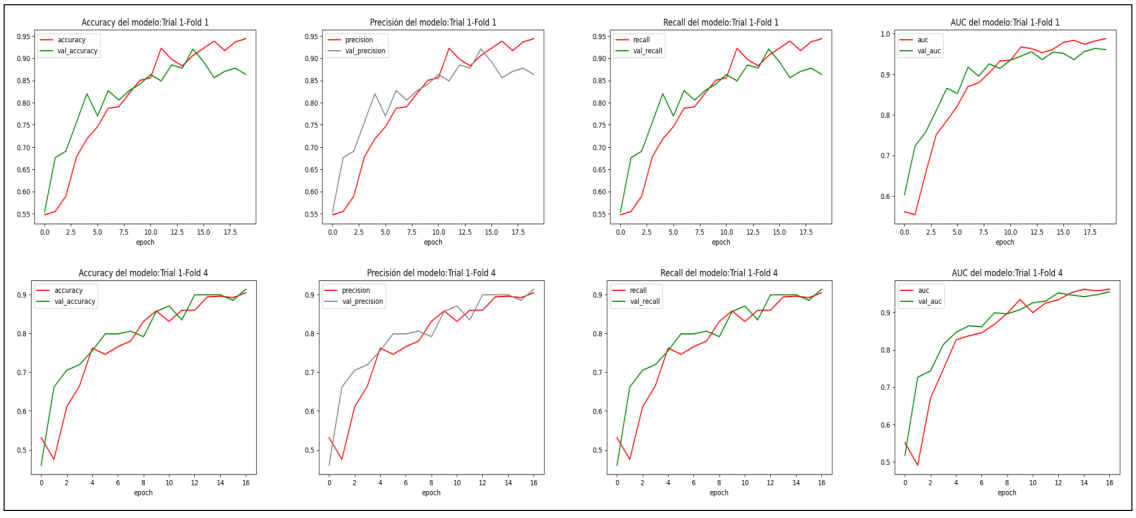


Figura 4.7: Comportamiento del modelo durante el entrenamiento en dos folds del primer experimento.

79,32 % para nivel alto y 86,33 % para nivel bajo. F1-Score fue del 83,84 % y el coeficiente kappa fue de 0,7029. En la prueba 3, el modelo obtuvo una precisión del 85,74 %, una precisión por clase del 87,98 % para el nivel alto y 86,90 % para el nivel bajo, y un recall por clase del 85,74 % para ambas clases. F1-Score fue del 83,58 % y el coeficiente kappa fue de 0,7096.

En la prueba 4, el modelo obtuvo una precisión del 87,90 %, una precisión por clase del 88,03 % para el nivel alto y 89,89 % para el nivel bajo, y una recuperación por clase del 85,71 % para el nivel alto y 86,62 % para el nivel bajo. F1-Score fue del 86,62 % y el coeficiente kappa fue de 0,7548. En la prueba 5, el modelo obtuvo una precisión del 87,72 %, una precisión por clase del 82,60 % para el nivel alto y 90,35 % para el nivel bajo, y un recall por clase del 79,96 % para el nivel alto y 86,89 % para el nivel bajo. F1-Score alcanzó 83,56 % y el coeficiente kappa fue de 0,7055. En promedio, el modelo obtuvo un accuracy del 85,93 %, una precisión por clase del 85,09 % para el nivel alto y 88,70 % para nivel bajo, y una recuperación por clase del 82,11 % para el nivel alto y 86,84 % para el nivel bajo. F1-Score fue del 84,29 % y el coeficiente kappa fue de 0,7053. Los resultados muestran que el modelo tuvo un rendimiento consistente en todas las pruebas, con una precisión y un F1-Score promedio del 85,93 % y 84,29 %, respectivamente. También se puede observar que la precisión por clase y la recuperación por clase varían según la prueba. Además, el coeficiente kappa muestra un nivel moderado de acuerdo entre las predicciones del modelo y los valores reales. En general, los resultados alcanzados durante las cinco pruebas de *5-fold-cross-validation* realizadas al modelo se muestran en la Figura 4.8

La Figura 4.7 contiene las curvas del desempeño del modelo durante el entrenamiento. Contiene las tres principales métricas: accuracy, precision y recall. Las gráficas muestran una generalización distinta en cada fold, sin embargo, el comportamiento de las cuervas durante el entrenamiento y la validación no muestran indicios de sobreentrenamiento. Como se comentó previamente, el indicio principal se observa en las cuervas de accuracy. En todos los folds existen fluctuaciones entre ambas curvas provocadas del procesamiento de la información, pero no existen señales de alarma importantes en lo que respecta a una memorización de los datos. Lo que nos permite afirmar que el modelo ha generalizado correctamente en la naturaleza del problema.

En la Figura 4.8 se muestran los diagramas de caja para un mejor entendimiento de las principales características de los resultados alcanzados. En la clasificación de niveles cognitivos de atención se alcanza un 85.93 % de exactitud en la identificación de niveles alto y bajo. Aunque son métricas prometedoras, la naturaleza del problema requiere de niveles de exactitud y precisión elevados para proporcionar información acertada sobre el comportamiento del sujeto. Tal como se encuentra el modelo se requiere de una mayor participación de un especialista durante el uso de esta herramienta. Esto no quiere decir que el modelo no sea capaz de arrojar resultados certeros, pero nuevamente la naturaleza del problema requiere de la menor cantidad de ambigüedades.

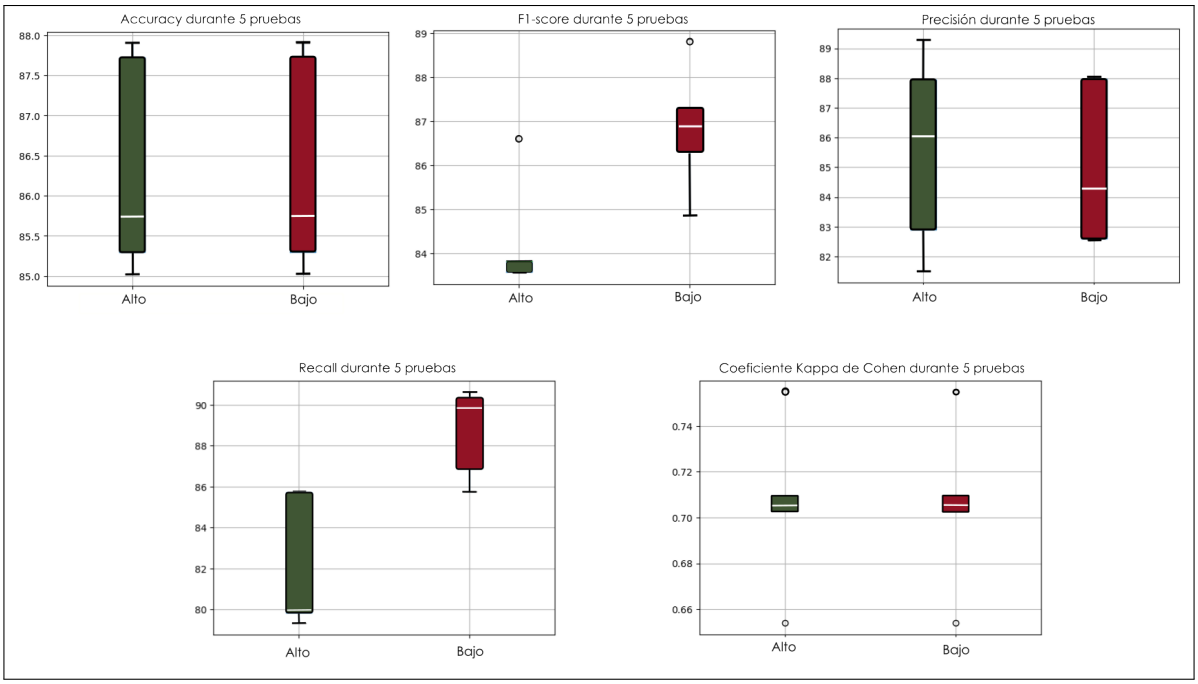


Figura 4.8: Análisis de los resultados alcanzados para la clasificación de niveles cognitivos.

4.2.3. Estimación de Estados de Atención

A pesar de combinar tres conjuntos de datos en este estudio, no hay ejemplos suficientes para la clasificación completa del nivel de atención alto. Por tanto, la estimación del modelo clínico de este estudio se limita a cuatro de los cinco niveles de atención. La experimentación reveló que las muestras de atención alternada eran las más escasas entre los conjuntos de datos. Este comportamiento puede atribuirse a la naturaleza de las pruebas administradas a los sujetos para evaluar la capacidad mental del cerebro para alternar entre estímulos visuales.

Esta etapa se sometió nuevamente a cinco experimentos, al igual que fases de entrenamiento anteriores. En estos experimentos se empleó nuevamente un enfoque de validación cruzada quíntuple, lo que permitió evaluar el rendimiento del modelo para varios conjuntos de datos de entrenamiento y prueba, garantizando una generalización adecuada entre clases y no basada únicamente en las muestras que el modelo procesó durante el entrenamiento.

Los resultados presentados en la Tabla 4.4 muestran la evaluación del modelo en cinco trials diferentes para la estimación de atención sostenida y dividida, con métricas como accuracy, precisión, recall, F1-Score y Coeficiente kappa de Cohen. Cabe destacar que los valores de asccuracy son superiores al 94 %, lo que sugiere un excelente rendimiento general del modelo. En cuanto a la precisión y recall, podemos observar que los valores varían entre los trials, siendo el valor más alto de precisión el 89,28 % para atención dividida en el ensayo dos y el valor más alto de recall 85,74 % en el trial tres. Estas diferencias pueden indicar que el modelo es sensible a algunos patrones de datos. Además, el coeficiente kappa del modelo

es de 0,6893, lo que sugiere una concordancia sustancial entre las predicciones y las etiquetas reales. Este valor indica que el modelo es coherente en sus predicciones y puede considerarse fiable.

En cuanto a la atención sostenida, el alto rendimiento del modelo viene indicado por la accuracy obtenida, que oscila entre el 94,05 % y el 99,00 %. Además, en todos los ensayos se observan valores de precisión y recall sistemáticamente elevados, siendo el valor de precisión más alto el 99,00 % del trial cinco y el valor de recuperación más alto el 100 % de los trials 1, 3 y 5. Estos valores elevados sugieren que el modelo puede ser fiable para la atención sostenida. Estos valores tan altos sugieren que el modelo puede predecir con exactitud los datos dados. Además, el coeficiente kappa de Cohen del modelo oscila entre 0,6432 y 0,7421, lo que indica una concordancia de buena a sustancial entre las predicciones y las etiquetas reales. Este comportamiento sugiere que las predicciones del modelo son coherentes y que se puede confiar en ellas para realizar predicciones precisas de nuevos datos.

La Tabla 4.3 presenta el rendimiento mostrado para la estimación de la atención selectiva y enfocada. Para la estimación de atención selectiva los resultados oscilan entre el 70,06 % y el 80,64 % de precisión. También se observa que el Coeficiente kappa de Cohen tiene un valor medio de 0,6453, lo que sugiere una buena concordancia entre las predicciones del modelo y las etiquetas reales. Cabe destacar que, aunque algunos valores de precisión y recall son relativamente bajos, el valor medio se mantiene cerca del 65 %, lo que indica que el modelo puede realizar predicciones precisas en general. Sin embargo, es esencial tener en cuenta que los resultados obtenidos son sólo una medida de la capacidad del modelo y deben interpretarse en el contexto de las limitaciones del conjunto de datos utilizado.

En cuanto a la estimación de la atención selectiva, el accuracy nuevamente oscila entre el 70,06 % y el 80,64 %, lo que indica un rendimiento moderado del modelo. Sin embargo, al observar los valores de precisión y recall, se observa que varían considerablemente entre los trials, siendo el valor más alto de precisión el 86,95 % en el trial cuatro y el valor más alto de recall es de 90,47 % en el trial uno. Estas diferencias sugieren que algunos patrones de datos son sensibles al modelo. El Coeficiente kappa de Cohen se mantiene en 0,6453, lo que indica una buena concordancia entre las predicciones y las etiquetas reales. Aunque no es un valor ideal, sugiere que las predicciones realizadas por el modelo se consideran coherentes.

El análisis de las métricas de la estimación de atención sostenida y dividida se presenta en la Figura 4.9 a través de boxplots. Este gráfico permite visualizar la distribución del conjunto de datos de cada métrica, mostrando la mediana, el rango intercuartílico, los valores extremos y los valores atípicos. El boxplot nos permite identificar si los datos están sesgados, si hay valores atípicos o si la distribución es simétrica o asimétrica. En general existe un sesgo inclinado a la estimación al estado de atención sostenida, esto en su mayoría se debe a la distribución de los datos de ambas clases durante el entrenamiento y la prueba al modelo.

El análisis de las métricas de la estimación de atención enfocada y selectiva se presenta en

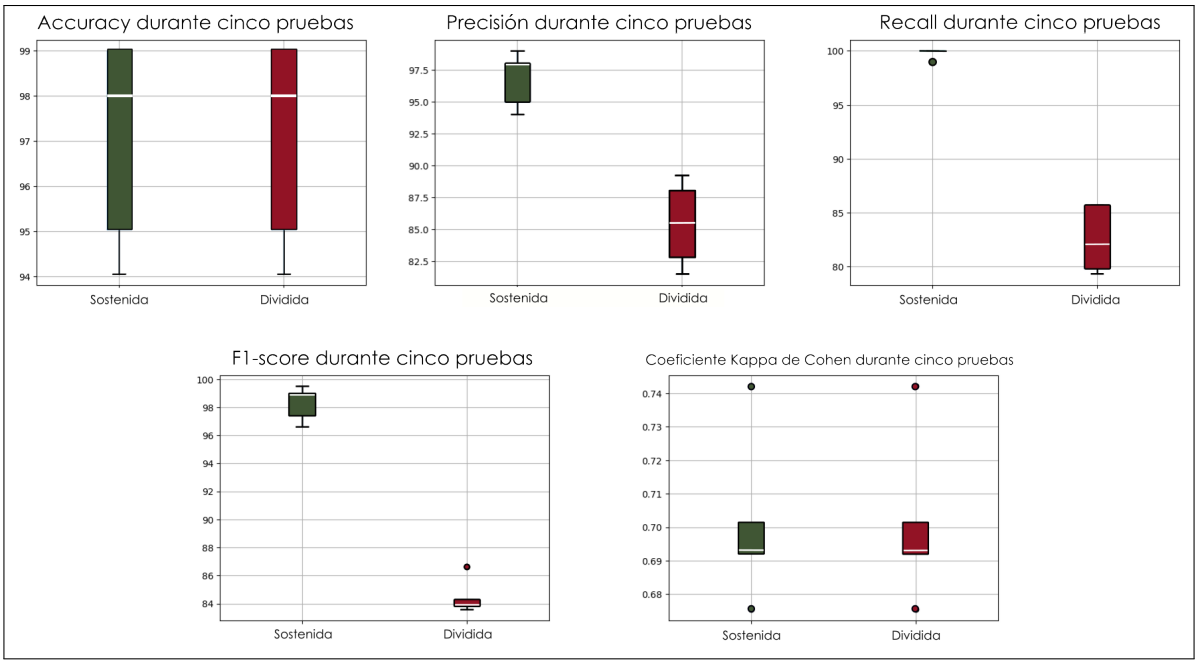


Figura 4.9: Análisis de las métricas de la estimación de sostenida y dividida.

la Figura 4.10 a través de boxplots nuevamente. Este gráfico permite visualizar la distribución del conjunto de datos de cada métrica, mostrando la mediana, el rango intercuartílico, los valores extremos y los valores atípicos. En general en esta estimación las métricas no se encuentran tan sesgadas a un estado de atención en particular y salvo el coeficiente kappa de Cohen, las métricas no presentan valores atípicos entre los cinco experimentos. Esto permite asegurar una correcta generalización del modelo en ambos estados de atención.

Ya se ha mencionado anteriormente que no se pudo completar la estimación de los cinco niveles del modelo de atención clínica debido a la falta de muestras para el entrenamiento del modelo. Lamentablemente, en las bases de datos públicas utilizadas en este estudio no se incluye una prueba que permita a los sujetos demostrar un comportamiento alternativo entre dos estímulos visuales. Se obtuvieron resultados prometedores para la estimación de la atención sostenida y selectiva, ya que se observó un comportamiento coherente en las distintas pruebas que confirmaba el procesamiento correcto del modelo en los patrones relacionados con estos dos niveles. Sin embargo, los resultados para la atención dividida y enfocada indican una generalización media en los patrones de estos estados de atención. En comparación con los niveles anteriormente mencionados, se observó una caída en el rendimiento del modelo de una media del 20 %. Este comportamiento puede entenderse debido a la distribución de clases en el conjunto de datos de entrenamiento. En general, se muestra una confianza moderada en la estimación de los niveles de atención a partir de los resultados obtenidos.

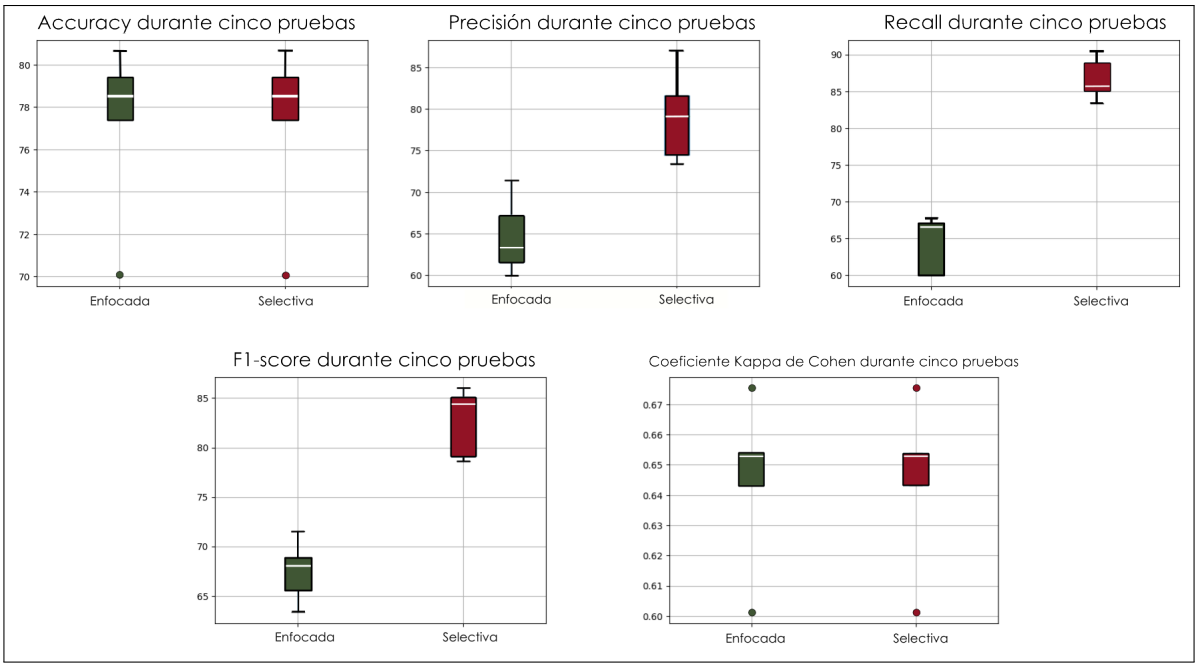


Figura 4.10: Análisis de las métricas de la estimación de atención enfocada y selectiva.

4.3. Impacto

En México, alrededor del cinco por ciento de los niños y adolescentes presentan Trastorno por Déficit de Atención e Hiperactividad (TDAH), pero su detección y diagnóstico tarda de tres a cinco años porque su comportamiento se confunde con el de menores criados de manera inadecuada e hiperactivos. El TDAH se puede detonar durante el embarazo por tabaquismo, alcoholismo o consumo de drogas. También pueden desarrollarlo los menores que presentan asfixia perinatal, bajo peso al nacer, nacimiento prematuro o cuyas madres enfrentaron disfunción familiar durante el embarazo. Otros factores de riesgo son la edad materna, es decir, tener hijos antes de los 18 años o después de los 35 [92]. A pesar de tener tan alto impacto en la calidad de vida de los menores que lo padecen, en la mayoría de los países el TDAH no se diagnostica en edades tempranas [93]; lo anterior supone que los menores tendrían que ser diagnosticados antes de los seis años, condición indispensable para que desarrollen las habilidades requeridas al enfrentar las exigencias que les habrá de imponer la etapa escolar [94].

4.3.1. Impacto Social

El trastorno por déficit de atención con hiperactividad (TDAH) es un trastorno neurológico común en la infancia que afecta el rendimiento académico y el comportamiento social de los niños. Cuando se diagnostica adecuadamente y se proporciona el tratamiento adecuado, los niños con TDAH pueden lograr resultados académicos y sociales positivos. Sin

embargo, si un niño es mal diagnosticado con TDAH, pueden surgir consecuencias graves y perjudiciales. Por ejemplo, el niño puede recibir un tratamiento inadecuado, que puede incluir la prescripción de medicamentos psicoestimulantes que no son necesarios o la falta de tratamiento para otros trastornos subyacentes que pueden estar afectando su salud mental. Además, el estigma asociado con el TDAH puede afectar la autoestima y el bienestar emocional del niño, lo que puede tener un impacto negativo en su vida a largo plazo [95]. Es crucial que se realicen diagnósticos precisos del TDAH para evitar estas consecuencias perjudiciales y garantizar que los niños reciban el tratamiento adecuado que necesitan para prosperar en su vida académica y social [96].

Para diagnosticar el TDAH, los profesionales de la salud pueden utilizar una combinación de métodos de evaluación, que pueden incluir [97]:

1. Evaluación clínica: un profesional de la salud, como un psiquiatra o psicólogo clínico, puede realizar una entrevista clínica exhaustiva con el paciente y sus padres o cuidadores para recopilar información sobre los síntomas actuales y pasados, la historia familiar y los antecedentes médicos y de desarrollo.

2. Evaluación psicológica: Puede incluir entrevistas con el paciente y sus familiares, así como pruebas psicológicas y de atención para evaluar la presencia de síntomas de TDAH.

3. Evaluación conductual: Los profesionales de la salud pueden observar el comportamiento del paciente en diferentes situaciones, como en la escuela o en casa, para evaluar la presencia de síntomas de TDAH.

4. Evaluación neurológica: Los profesionales de la salud pueden realizar pruebas neurológicas para evaluar la función cerebral y descartar otras afecciones que puedan estar causando los síntomas.

5. Evaluación de la salud física: Los profesionales de la salud pueden realizar pruebas de salud física para descartar otras afecciones médicas que puedan estar causando los síntomas.

En general, el diagnóstico del TDAH se realiza mediante la combinación de estos métodos de evaluación. Es importante destacar que no existe una prueba única que pueda diagnosticar el TDAH de manera precisa, por lo que se recomienda una evaluación completa y detallada por parte de un profesional de la salud capacitado para hacer un diagnóstico adecuado.

Detectar los síntomas relacionados con el TDAH puede ser un desafío para los profesionales de la salud debido a la variedad de síntomas que pueden ser observados en los niños, así como a la falta de pruebas diagnósticas claras. Además, algunos síntomas del TDAH, como la falta de atención y la hiperactividad, pueden ser comunes en los niños en general, lo que dificulta aún más la identificación precisa del trastorno.

La detección de síntomas de trastornos de atención en general, como TDAH, en las escuelas públicas puede ser un desafío debido a una serie de factores. En primer lugar, los profesores pueden tener una carga de trabajo muy alta y no tener tiempo suficiente para observar cuidadosamente a cada estudiante en el aula y notar cualquier comportamiento que pueda indicar un trastorno de atención. Además, los profesores pueden no estar capacitados para reconocer los síntomas del TDAH y otros trastornos de atención, lo que puede retrasar el diagnóstico y el tratamiento adecuado.

Los estados de atención son fundamentales para el desarrollo y la realización de las tareas cotidianas, y cualquier alteración en ellos puede tener consecuencias negativas para la vida de una persona. En particular, los trastornos y patologías relacionados con el TDAH están estrechamente ligados a problemas en el estado de atención, como la falta de concentración, la hiperactividad y la impulsividad [98].

La investigación ha demostrado que estos síntomas pueden afectar significativamente la vida académica y social de una persona, lo que a su vez puede tener un impacto negativo en su autoestima y bienestar emocional a largo plazo. Por lo tanto, es crucial que se realicen diagnósticos precisos del TDAH y que se proporcione un tratamiento adecuado para ayudar a las personas a manejar estos síntomas y alcanzar su máximo potencial [99].

Los estados de atención también pueden ser un indicador de otros trastornos y patologías, como el trastorno del espectro autista, la depresión y la ansiedad. Por lo tanto, es importante que los profesionales de la salud presten atención a los estados de atención y consideren su relación con otros trastornos y patologías en el diagnóstico y tratamiento de estas condiciones. La presente contribución pretende brindar una herramienta que permita a especialistas o profesores realizar una primera estimación sobre estados de atención durante pruebas cognitivas realizadas a sujetos para determinar si existe alguna patología o trastorno. La metodología propuesta alcanza resultados prometedores que pueden abrir la puerta a un diagnostico preciso y eficaz por parte de especialistas, el modelo entrenado y probado arroja altos porcentajes de exactitud y representan confianza en la estimación de estados de atención.

4.4. Publicaciones

Esta investigación cuenta con una publicación presentada en XIX Congreso Internacional de Ingeniería titulada: Visual attention in images: estimating attention levels with eye-tracking and deep learning.

4.5. Trabajo Futuro

Los modelos entrenados en esta contribución tienen el potencial para su implementación en forma de herramienta. Esta herramienta tendría tanto la clasificación de eventos de movimientos oculares, clasificación de niveles de atención y la estimación de estados de atención. Este trabajo fue diseñado con conjuntos de datos de diferentes dispositivos de registro ocular de manera que el modelo no sea dependiente de un solo dispositivo, marca o frecuencia de muestreo. Esta estrategia proveerá de robustez cuando nueva información sea procesada para su análisis. Se pretende diseñar esta herramienta para usuarios especialistas que deseen aplicar un examen cognitivo a usuarios con dispositivos de seguimiento ocular.

La metodología presentada en esta contribución puede ser utilizada para nuevos conjuntos de datos relacionados a pruebas cognitivas y seguimiento ocular. Se plantea la hipótesis sobre la naturaleza de los modelos, una mayor cantidad de información extraída de diferentes usuarios tendría un mejor desempeño. Esta contribución se ve limitada por el número de sujetos de prueba de los conjuntos de datos disponibles, de tener acceso a conjuntos más grandes y diversos, la clasificación de niveles de atención y la estimación de estados presentaría mejores resultados. Como trabajo futuro se podría explorar la hipótesis de que, con una mayor diversidad de usuarios, los resultados podrían ser alentadores.

CAPÍTULO 5

Conclusiones

Esta contribución propone el uso de información extraída de sesiones de seguimiento ocular durante pruebas cognitivas para la estimación de estados de atención. Uno de los principales objetivos de esta investigación fue establecer métricas de seguimiento ocular con los diferentes estados de atención. Contribuciones previas se han limitado a la clasificación de niveles de atención simples, como alto o bajo. Los niveles de atención no tienen una descripción universal, cada autor la torna a interpretación propia. Los estados de atención a estimar en esta investigación consisten en un modelo clínico establecido previamente, en donde se describen cinco estados de atención, que brindan información precisa sobre la carga cognitiva del sujeto.

Varios padecimientos o patologías tienen problemas de atención como parte de los síntomas. Especialistas que realizan el diagnostico de estas patologías podrían incluir como parte de sesiones de análisis una herramienta que permita visualizar el comportamiento ocular y una estimación sobre el estado de atención ejecutado por el sujeto. Modelos como el presentado en esta investigación implicarían un apoyo a los especialistas para emitir diagnósticos completos a sujetos con padecimientos relacionados a problemas de atención. Esta contribución pretende además minimizar el escenario necesario para la estimación de estados de atención. Propuestas anteriores presentan la combinación de distintos dispositivos para registrar características diversas del sujeto durante la prueba realizada, mientras que la metodología propuesta solo requiere del archivo de la sesión de seguimiento ocular en *csv*.

La metodología planteada consiste en un enfoque de procesamiento con aprendizaje profundo y es la primera en la literatura en proponer el procesamiento de la información en imágenes. A través de una transformación a un dominio bidimensional se utiliza una técnica de estimación de movimiento: flujo óptico. La técnica de flujo óptico denso se utiliza para la transformación de sesiones de seguimiento ocular completas, resultando en imágenes con la translación de la mirada por la pantalla representadas en magnitud y dirección a través de un modelo de color. El procesamiento de los modelos CNN con la información en un domino tan rico en información permitió trabajar con una limitación importante: datos. Como se

mencionó anteriormente, una de las limitaciones más importantes de esta contribución recae en la información disponible. Este problema fue abordado a través de la combinación de diferentes conjuntos de datos disponibles en la literatura y transferencia de aprendizaje del modelo en cuestión. Contribuciones previas han creado bases de datos completas, sin embargo, ninguna fue compartida con los autores de esta investigación. La poca difusión de la información consiste en una limitación en el desempeño del modelo que estima los diferentes estados de atención.

La base de datos maestra se encuentra limitada por el número de muestras disponibles para cada estado de atención y derivado de ello los experimentos no arrojan las métricas de desempeño esperadas. Los experimentos conducidos en esta investigación sugieren una correcta estimación de cuatro estados de atención del modelo clínico con un promedio de exactitud del 80 %. Los pocos ejemplos disponibles para el entrenamiento del modelo no fueron suficientes para la estimación del estado de atención alternada, dada la naturaleza de las pruebas de los diferentes conjuntos de datos. Un trabajo futuro podría ser integrar una mayor cantidad de ejemplos para la estimación del modelo clínico completo, esto supondría la presentación de los cinco estados de atención y arrojaría información completa a especialistas que deseen incluir el análisis del comportamiento ocular durante sesiones con sujetos. Dado que el método de obtención de movimientos oculares en general no es invasivo, esto podría mejorar la interacción de los sujetos durante su evaluación cognitiva.

CAPÍTULO 6

Anexos

6.1. Artículo Presentado

CONiiN
XXII CONGRESO INTERNACIONAL DE INGENIERÍA

Visual attention in images: estimating attention levels with eye-tracking and deep learning.

Alea Bello-Díaz
Facultad de Ingeniería
Universidad Autónoma de Querétaro
C. U. Cerro de las Campanas s/n,
Querétaro, 76010, Querétaro, México

Sebastián Salazar-Colores
Centro de Investigaciones en Óptica
A. C.
León, 37150, Guanajuato, México

M. A. Aceves-Fernández*
Facultad de Ingeniería
Universidad Autónoma de Querétaro
C. U. Cerro de las Campanas s/n,
Querétaro, 76010, Querétaro, México

*corresponding author: marco.aceves@uaq.mx

***Abstract*— The investigation of attention levels is deemed a critical task in various research areas, including psychology, neurology, psychiatry, and others. However, complete studies on the spectrum of attention are yet to be presented. Previous studies have been limited to determining the mere presence or absence of attention, which precludes the accurate estimation of subjects' cognitive load using such attention classification. To address this shortcoming, a novel clinical model comprising five levels of attention is proposed in this study. A comprehensive description of the subject's cognitive state is provided by this model, facilitating a more nuanced understanding of attention levels. The estimation of these five levels is carried out by utilizing physical properties of eye movements through a deep learning and transfer learning methodology. The study findings suggest that the proposed methodology provides a precise generalization for four of the five levels of attention. Consequently, the model has potential applications in various contexts, including the diagnosis of attention deficit hyperactivity disorder and the measurement of cognitive intervention outcomes.**

***Keywords*—deep learning; eye-tracking; attention; estimation**

I. Introduction

Estimating visual attention levels is considered a critical aspect of human-computer interaction research [1], [2]. Precise information on a user's visual attention is provided by eye-tracking; however, its analysis and classification can be challenging due to its complexity. Deep learning (DL) techniques are a promising solution to this problem since they can learn complex patterns and process large amounts of data, making them ideal for classifying visual attention levels.

Detailed information on eye fixations and their duration is provided by eye-tracking sensors, which can be combined with DL techniques to analyze eye movement dynamics and classify neural activity during cognitive tasks [3], [4]. In addition, insights into neurological disorders and other attention-related processes can also be obtained from eye movement patterns [4], [5]. Therefore, developing robust systems that can accurately determine a person's attentional state from eye-tracking data is crucial.

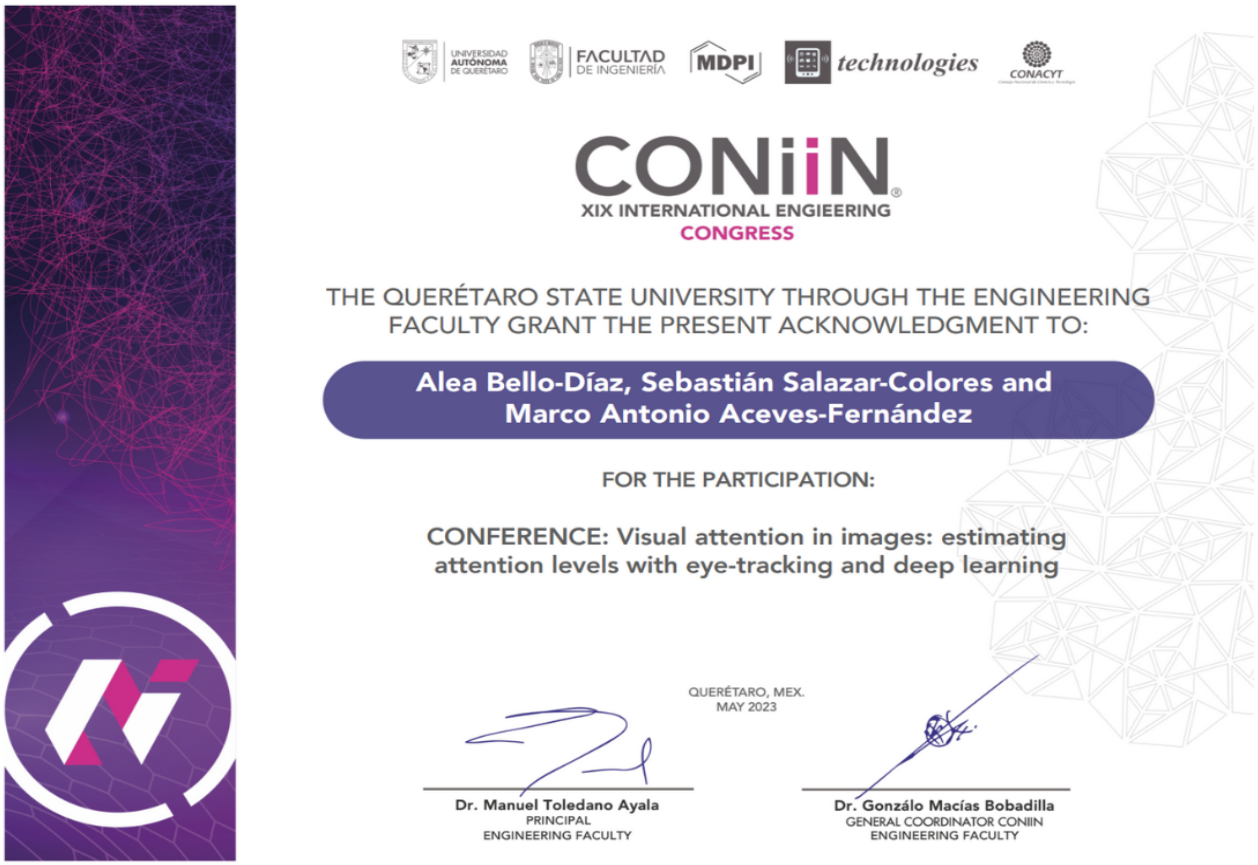
Eye movements are essential for visual perception, allowing us to gather information about the world [5]. Three primary types of eye movements are fixations, saccades, and smooth pursuit. Fixations are short pauses during which the eye remains relatively still and focused on a particular point of interest [6]. These pauses can last anywhere from a few hundred milliseconds to several seconds [5]. Saccades, on the other hand, are rapid eye movements that shift the gaze from one point to another. They are responsible for bringing new visual information into the fovea, the central part of the retina responsible for high-resolution vision [1], [5]. Finally, smooth pursuit movements are used to track moving objects. During the smooth pursuit, the eye moves smoothly to follow a moving object to keep it focused on the fovea[5], [7].

Valuable insights can be obtained by analyzing eye movement velocity and duration regarding how visual information is processed [1], [8]. For example, research has demonstrated that attention and interest in a specific object may be connected to the time to fixate on it. Cognitive processes such as attention, memory, and decision-making can be indicated by the velocity and duration of saccades[2], [6], [9]. On the other hand, the speed and precision of smooth pursuits can be analyzed to evaluate visual system functionality, particularly in diagnosing visual and neurological disorders.

Attention levels based on eye-tracking metrics are aimed to be estimated using new domain and DL techniques in this contribution. The lack of datasets is one of the primary challenges in estimating attention with eye-tracking. To address this issue, three publicly available datasets labeled with the three main eye movements: fixations, saccades, and smooth pursuits, are combined in this contribution. The physical properties of eye movements and the nature of visual stimuli are analyzed to classify attention levels as low or high.

The Transfer Learning (TL) approach forms the core of the proposed methodology, which leverages the pre-trained EfficientNet-B0 [7] model developed for a classification task with a more significant number of examples: the classification of eye movement events. This strategy is seen as critical in addressing the primary challenge of the research: the lack of information. By combining DL techniques and eye-tracking data, accurate and efficient models can be created to predict a person's attention level, with potential applications in neuroscience research and other domains.

Constancia de Participación en el XIX Congreso Internacional de Ingeniería



UNIVERSIDAD AUTÓNOMA DE QUERÉTARO | FACULTAD DE INGENIERÍA | MDPI | technologies | CONACYT

CONiiN
XIX INTERNATIONAL ENGIEERING CONGRESS

THE QUERÉTARO STATE UNIVERSITY THROUGH THE ENGINEERING FACULTY GRANT THE PRESENT ACKNOWLEDGMENT TO:

Alea Bello-Díaz, Sebastián Salazar-Colores and Marco Antonio Aceves-Fernández

FOR THE PARTICIPATION:

CONFERENCE: Visual attention in images: estimating attention levels with eye-tracking and deep learning

QUERÉTARO, MEX.
MAY 2023

Dr. Manuel Toledano Ayala
PRINCIPAL
ENGINEERING FACULTY

Dr. Gonzálo Macías Bobadilla
GENERAL COORDINATOR CONIIN
ENGINEERING FACULTY

6.2. Constancias Manejo Lengua Extranjera

Constancia Examen Manejo de la Lengua

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
FACULTAD DE LENGUAS Y LETRAS

A QUIEN CORRESPONDA:

La que suscribe, Directora de la Facultad de Lenguas y Letras, hace **C O N S T A R** que

BELLO DIAZ ALEA FERNANDA

Presentó el **Examen de Manejo de la Lengua** efectuado el día diez de noviembre de dos mil veintiuno, en el cual obtuvo la siguiente calificación:

8

Se extiende la presente a petición de la parte interesada, para los fines escolares y legales que le convengan, en el Campus Aeropuerto de la Universidad Autónoma de Querétaro, el día veinticinco de noviembre de dos mil veintiuno.

A t e n t a m e n t e,
"Enlazar Culturas por la Palabra"

DRA. ADELINA VELÁZQUEZ HERRERA

AVH/japa*CL*FLL-C.-2271

SOMOS UAQ
EDUCAR CRECER CONSOLIDAR

Campus Aeropuerto, Anillo Vial Fray Junípero Serra S/N, Querétaro, Qro. C.P. 76140
Tel. 442 192 12 00 Dirección Ext. 61010, Secretaría Administrativa Ext.61300, Posgrado Ext. 61140, Licenciatura Ext.61070, Centro de Lenguas Ext.61050, Secretaría Académica Ext.61100 y Planeación Ext.61110

Constancia Examen de Comprensión de Textos en Inglés

UNIVERSIDAD AUTÓNOMA DE QUERÉTARO
FACULTAD DE LENGUAS Y LETRAS

A QUIEN CORRESPONDA:

La que suscribe, Directora de la Facultad de Lenguas y Letras, hace **C O N S T A R** que

BELLO DIAZ ALEA FERNANDA

Presentó y acreditó el **Examen de Comprensión de Textos en Inglés** efectuado el día veintisiete de febrero de dos mil veintitrés.

Se extiende la presente a petición de la parte interesada, para los fines escolares y legales que le convengan, en el Campus Aeropuerto de la Universidad Autónoma de Querétaro, el día nueve de marzo de dos mil veintitrés.

A t e n t a m e n t e,
"Enlazar Culturas por la Palabra"

DRA. ADELINA VELÁZQUEZ HERRERA

AVH/daa*CL*FLL-C.-616

Bibliografía

[1] M. Dorr, T. Martinetz, K. R. Gegenfurtner, and E. Barth, "Variability of eye movements when viewing dynamic natural scenes," *Journal of Vision*, vol. 10, pp. 28–28, Aug. 2010.

[2] L. Larsson, M. Nystrom, and M. Stridh, "Detection of saccades and postsaccadic oscillations in the presence of smooth pursuit," *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 2484–2493, Sept. 2013.

[3] I. Agtzidis, M. Startsev, and M. Dorr, "360-degree video gaze behaviour: A ground-truth data set and a classification algorithm for eye movements," in *Proceedings of the 27th ACM International Conference on Multimedia (MM '19)*, ACM, 2019.

[4] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Sept. 2020. arXiv:1905.11946 [cs, stat].

[5] O. T. Chen, P. C. Chen, and Y. T. Tsai, "Attention estimation system via smart glasses," *2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2017*, 2017.

[6] A. T. Duchowski, *Eye Tracking Methodology*. Cham: Springer International Publishing, 2017.

[7] M. A. Just and P. A. Carpenter, "A theory of reading: from eye fixations to comprehension.," *Psychological review*, vol. 87 4, pp. 329–54, 1980.

[8] B. Zablotsky, L. I. Black, M. J. Maenner, L. A. Schieve, M. L. Danielson, R. H. Bitsko, S. J. Blumberg, M. D. Kogan, and C. A. Boyle, "Prevalence and Trends of Developmental Disabilities among Children in the United States: 2009–2017," *Pediatrics*, vol. 144, 10 2019. e20190811.

[9] S. De Silva, S. Dayarathna, G. Ariyarathne, D. Meedeniya, S. Jayarathna, A. M. P. Michalek, and G. Jayawardena, "A rule-based system for adhd identification using eye movement data," in *2019 Moratuwa Engineering Research Conference (MERCon)*, pp. 538–543, 2019.

[10] J. M. Swanson and N. D. Volkow, "Lessons from the 1918 flu pandemic: a novel etiologic subtype of adhd?," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 60, pp. 1–2, Jan. 2021.

[11] Y. Zhao, Z. Jiang, S. Guo, P. Wu, Q. Lu, Y. Xu, L. Liu, S. Su, L. Shi, J. Que, Y. Sun, Y. Sun, J. Deng, S. Meng, W. Yan, K. Yuan, S. Sun, L. Yang, M. Ran, T. R. Kosten, J. Strang, Y. Lu, G. Huang, L. Lu, Y. Bao, and J. Shi, "Association of symptoms of attention deficit and hyperactivity with problematic internet use among university students in wuhan, china during the covid-19 pandemic," *Journal of Affective Disorders*, vol. 286, pp. 220–227, May 2021.

[12] L. Hantsoo, S. Kornfield, M. C. Anguera, and C. N. Epperson, "Inflammation: a proposed intermediary between maternal stress and offspring neuropsychiatric risk," *Biological Psychiatry*, vol. 85, pp. 97–106, Jan. 2019.

[13] M. Johnson and J.-E. Kim, "The effect of task complexity on eye movement and multitasking performance in students with and without adhd," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 64, pp. 786–790, Dec. 2020.

[14] P. Deans, L. O'Laughlin, B. Brubaker, N. Gay, and D. Krug, "Use of eye movement tracking in the differential diagnosis of attention deficit hyperactivity disorder (Adhd) and reading disability," *Psychology*, vol. 01, no. 04, pp. 238–246, 2010.

[15] R. Engbert and R. Kliegl, "Microsaccades uncover the orientation of covert attention," *Vision Research*, vol. 43, pp. 1035–1045, Apr. 2003.

[16] D. P. Munoz, I. T. Armstrong, K. A. Hampton, and K. D. Moore, "Altered control of visual fixation and saccadic eye movements in attention-deficit hyperactivity disorder," *Journal of Neurophysiology*, vol. 90, pp. 503–514, July 2003.

[17] M. Abercrombie, "Perception and communication," *Education + Training*, vol. 8, pp. 264–269, June 1966.

[18] M. M. Sohlberg and C. A. Mateer, "Effectiveness of an attention-training program," *Journal of Clinical and Experimental Neuropsychology*, vol. 9, pp. 117–130, Apr. 1987.

[19] S. Van Der Stigchel, N. N. J. Rommelse, J. B. Deijen, C. J. A. Geldof, J. Witlox, J. Oosterlaan, J. A. Sergeant, and J. Theeuwes, "Oculomotor capture in ADHD," *Cognitive Neuropsychology*, vol. 24, pp. 535–549, July 2007.

[20] P.-H. Tseng, I. G. M. Cameron, G. Pari, J. N. Reynolds, D. P. Munoz, and L. Itti, "High-throughput classification of clinical populations from natural viewing eye movements," *Journal of Neurology*, vol. 260, pp. 275–284, Jan. 2013.

[21] A. Belle, R. H. Hargraves, and K. Najarian, "An automated optimal engagement and attention detection system using electrocardiogram," *Computational and Mathematical Methods in Medicine*, vol. 2012, pp. 1–12, 2012.

[22] N.-H. Liu, C.-Y. Chiang, and H.-C. Chu, "Recognizing the degree of human attention using eeg signals from mobile sensors," *Sensors*, vol. 13, pp. 10273–10286, Aug. 2013.

[23] S.-M. Yang, C.-M. Chen, and C.-M. Yu, "Assessing the attention levels of students by using a novel attention aware system based on brainwave signals," in *2015 IIAI 4th International Congress on Advanced Applied Informatics*, (Okayama, Japan), pp. 379–384, IEEE, July 2015.

[24] S. Aliakbaryhosseinabadi, E. N. Kamavuako, N. Jiang, D. Farina, and N. Mrachacz-Kersting, "Classification of EEG signals to identify variations in attention during motor task execution," *Journal of Neuroscience Methods*, vol. 284, pp. 27–34, June 2017.

[25] C. K. Toa, K. S. Sim, and S. C. Tan, "Electroencephalogram-based attention level classification using convolution attention memory neural network," *IEEE Access*, vol. 9, pp. 58870–58881, 2021.

[26] X. Wang, X. Li, J. Zhu, Z. Xu, K. Ren, W. Zhang, X. Liu, and K. Yu, "A local similarity-preserving framework for nonlinear dimensionality reduction with neural networks," in *Database Systems for Advanced Applications: 26th International Conference, DASFAA 2021, Taipei, Taiwan, April 11–14, 2021, Proceedings, Part II*, (Berlin, Heidelberg), p. 376–391, Springer-Verlag, 2021.

[27] T. A. Funkhouser and C. H. Séquin, "Adaptive display algorithm for interactive frame rates during visualization of complex virtual environments," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques - SIGGRAPH '93*, (Not Known), pp. 247–254, ACM Press, 1993.

[28] M. J. Mohammadi-Aragh, J. E. Ball, and D. Jaison, "Using wavelets to categorize student attention patterns," in *2016 IEEE Frontiers in Education Conference (FIE)*, (Erie, PA, USA), pp. 1–8, IEEE, Oct. 2016.

[29] J. Zaletelj and A. Košir, "Predicting students' attention in the classroom from Kinect facial and body features," *EURASIP Journal on Image and Video Processing*, vol. 2017, p. 80, Dec. 2017.

[30] Y. Abdelrahman, A. A. Khan, J. Newn, E. Velloso, S. A. Safwat, J. Bailey, A. Bulling, F. Vetere, and A. Schmidt, "Classifying attention types with thermal imaging and eye tracking," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, pp. 1–27, Sept. 2019.

[31] M. Startsev, I. Agtzidis, and M. Dorr, "1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits," *Behavior Research Methods*, vol. 51, pp. 556–572, Apr. 2019.

[32] I. Agtzidis, M. Startsev, and M. Dorr, "A ground-truth data set and a classification algorithm for eye movements in 360-degree videos," in *Proceedings of the 27th ACM*

International Conference on Multimedia, pp. 1007–1015, Oct. 2019. arXiv:1903.06474 [cs].

[33] K. P. Murphy, *Machine learning: a probabilistic perspective*. Adaptive computation and machine learning series, Cambridge, MA: MIT Press, 2012.

[34] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. Adaptive computation and machine learning, Cambridge, Massachusetts: The MIT Press, 2016.

[35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017.

[36] Z. Tu, W. Xie, D. Zhang, R. Poppe, R. C. Veltkamp, B. Li, and J. Yuan, "A survey of variational and CNN-based optical flow techniques," *Signal Processing: Image Communication*, vol. 72, pp. 9–24, Mar. 2019.

[37] X.-C. Yin, Z.-Y. Zuo, S. Tian, and C.-L. Liu, "Text detection, tracking and recognition in video: a comprehensive survey," *IEEE Transactions on Image Processing*, vol. 25, pp. 2752–2773, June 2016.

[38] Y.-H. Tsai, M.-H. Yang, and M. J. Black, "Video segmentation via object flow," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Las Vegas, NV, USA), pp. 3899–3908, IEEE, June 2016.

[39] R. V. H. M. Colque, C. Caetano, M. T. L. De Andrade, and W. R. Schwartz, "Histograms of optical flow orientation and magnitude and entropy to detect anomalous events in videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, pp. 673–682, Mar. 2017.

[40] T. Poggio and W. Reichardt, "Visual control of orientation behaviour in the fly: Part II. Towards the underlying neural interactions," *Quarterly Reviews of Biophysics*, vol. 9, pp. 377–438, Aug. 1976.

[41] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, Aug. 1981.

[42] G. Farneback, "Fast and accurate motion estimation using orientation tensors and parametric motion models," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 1, (Barcelona, Spain), pp. 135–139, IEEE Comput. Soc, 2000.

[43] M. Frueh, T. Kuestner, M. Nachbar, D. Thorwarth, A. Schilling, and S. Gatidis, "Self-supervised learning for automated anatomical tracking in medical image data with minimal human labeling effort," *Computer Methods and Programs in Biomedicine*, vol. 225, p. 107085, Oct. 2022.

[44] I. Agtzidis, M. Startsev, and M. Dorr, "In the pursuit of (Ground) truth: a hand-labelling tool for eye movements recorded during dynamic scene viewing," in *2016 IEEE Second Workshop on Eye Tracking and Visualization (ETVIS)*, (Baltimore, MD, USA), pp. 65–68, IEEE, Oct. 2016.

[45] L. Itti, G. Rees, and J. K. Tsotsos, eds., *Neurobiology of attention*. Amsterdam ; Boston: Elsevier Academic Press, 2005. OCLC: ocm58724292.

[46] C. Valuch, P. König, and U. Ansorge, "Memory-guided attention during active viewing of edited dynamic scenes," *Journal of Vision*, vol. 17, p. 12, Jan. 2017.

[47] S. T. Chung, G. Kumar, R. W. Li, and D. M. Levi, "Characteristics of fixational eye movements in amblyopia: Limitations on fixation stability and acuity?," *Vision Research*, vol. 114, pp. 87–99, Sept. 2015.

[48] B. Parhizi, M. R. Daliri, and M. Behroozi, "Decoding the different states of visual attention using functional and effective connectivity features in fMRI data," *Cognitive Neurodynamics*, vol. 12, pp. 157–170, Apr. 2018.

[49] R. P. G. Van Gompel, ed., *Eye movements: a window on mind and brain*. Amsterdam ; Boston: Elsevier, 1st ed ed., 2007. OCLC: ocm82671634.

[50] I. T. Hooge and C. J. Erkelens, "Peripheral vision and oculomotor control during visual search," *Vision Research*, vol. 39, pp. 1567–1575, Apr. 1999.

[51] H. J. Müller and A. Von Mühlenen, "Attentional tracking and inhibition of return in dynamic displays," *Perception & Psychophysics*, vol. 58, pp. 224–249, Mar. 1996.

[52] S. R. Simon, M. Meunier, L. Piettre, A. M. Berardi, C. M. Segebarth, and D. Boussaoud, "Spatial Attention and Memory Versus Motor Preparation: Premotor Cortex Involvement as Revealed by fMRI," *Journal of Neurophysiology*, vol. 88, pp. 2047–2057, Oct. 2002.

[53] E. Kowler, E. Anderson, B. Dosher, and E. Blaser, "The role of attention in the programming of saccades," *Vision Research*, vol. 35, pp. 1897–1916, July 1995.

[54] D. P. Munoz and S. Everling, "Look away: the anti-saccade task and the voluntary control of eye movement," *Nature Reviews Neuroscience*, vol. 5, pp. 218–228, Mar. 2004.

[55] N. Unsworth and M. K. Robison, "Pupillary correlates of lapses of sustained attention," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 16, pp. 601–615, Aug. 2016.

[56] C. L. Wiggs and A. Martin, "Properties and mechanisms of perceptual priming," *Current Opinion in Neurobiology*, vol. 8, pp. 227–233, Apr. 1998.

[57] A. Shechter and D. L. Share, "Keeping an Eye on Effort: A Pupillometric Investigation of Effort and Effortlessness in Visual Word Recognition," *Psychological Science*, vol. 32, pp. 80–95, Jan. 2021.

[58] L. L. Di Stasi, M. B. McCamy, A. Catena, S. L. Macknik, J. J. Cañas, and S. Martinez-Conde, "Microsaccade and drift dynamics reflect mental fatigue," *European Journal of Neuroscience*, vol. 38, pp. 2389–2398, Aug. 2013.

[59] K. Nishida, Y. Morishima, M. Yoshimura, T. Isotani, S. Irisawa, K. Jann, T. Dierks, W. Strik, T. Kinoshita, and T. Koenig, "EEG microstates associated with salience and frontoparietal networks in frontotemporal dementia, schizophrenia and Alzheimer's disease," *Clinical Neurophysiology*, vol. 124, pp. 1106–1114, June 2013.

[60] J. Zachary Jacobson and P. Dodwell, "Saccadic eye movements during reading," *Brain and Language*, vol. 8, pp. 303–314, Nov. 1979.

[61] B. Fischer and E. Ramsperger, "Human express saccades: extremely short reaction times of goal directed eye movements," *Experimental Brain Research*, vol. 57, no. 1, 1984.

[62] S. G. Lisberger, "Visual Guidance of Smooth-Pursuit Eye Movements: Sensation, Action, and What Happens in Between," *Neuron*, vol. 66, pp. 477–491, May 2010.

[63] J. L. Stubbs, S. L. Corrow, B. R. Kiang, J. C. Corrow, H. L. Pearce, A. Y. Cheng, J. J. S. Barton, and W. J. Panenka, "Working memory load improves diagnostic performance of smooth pursuit eye movement in mild traumatic brain injury patients with protracted recovery," *Scientific Reports*, vol. 9, p. 291, Jan. 2019.

[64] R. J. Krauzlis and S. G. Lisberger, "Simple spike responses of gaze velocity Purkinje cells in the floccular lobe of the monkey during the onset and offset of pursuit eye movements," *Journal of Neurophysiology*, vol. 72, pp. 2045–2050, Oct. 1994.

[65] K. Fukushima, J. Fukushima, T. Warabi, and G. R. Barnes, "Cognitive processes involved in smooth pursuit eye movements: behavioral evidence, neural substrate and clinical correlation," *Frontiers in Systems Neuroscience*, vol. 7, 2013.

[66] R. J. Leigh and D. S. Zee, *The Neurology of Eye Movements*. Oxford University Press, 5 ed., June 2015.

[67] J. Fan, B. D. McCandliss, T. Sommer, A. Raz, and M. I. Posner, "Testing the Efficiency and Independence of Attentional Networks," *Journal of Cognitive Neuroscience*, vol. 14, pp. 340–347, Apr. 2002.

[68] D. A. Pollen, "On the Neural Correlates of Visual Perception," *Cerebral Cortex*, vol. 9, pp. 4–19, Jan. 1999.

[69] S. S. Sherigar, A. H. Gamsa, and K. Srinivasan, "Oculomotor deficits in attention deficit hyperactivity disorder: a systematic review and meta-analysis," *Eye*, Oct. 2022.

[70] A. Serra, C. G. Chisari, and M. Matta, "Eye Movement Abnormalities in Multiple Sclerosis: Pathogenesis, Modeling, and Treatment," *Frontiers in Neurology*, vol. 9, p. 31, Feb. 2018.

[71] C.-C. Wu, B. Cao, V. Dali, C. Gagliardi, O. J. Barthelemy, R. D. Salazar, M. Pomplun, A. Cronin-Golomb, and A. Yazdanbakhsh, "Eye movement control during visual pursuit in Parkinson's disease," *PeerJ*, vol. 6, p. e5442, Aug. 2018.

[72] S. S. Stevens, J. T. Wixted, E. A. Phelps, and L. Davachi, eds., *Stevens' handbook of experimental psychology and cognitive neuroscience*. New York: John Wiley & Sons, Inc, fourth edition ed., 2018.

[73] G. K. Thaker, M. T. Avila, E. L. Hong, D. R. Medoff, D. E. Ross, and H. M. Adami, "A model of smooth pursuit eye movement deficit associated with the schizophrenia phenotype," *Psychophysiology*, vol. 40, pp. 277–284, Mar. 2003.

[74] J. G. Franco, J. de Pablo, A. M. Gaviria, E. Sepúlveda, and E. Vilella, "Smooth pursuit eye movements and schizophrenia: Literature review," *Archivos de la Sociedad Española de Oftalmología (English Edition)*, vol. 89, pp. 361–367, Sept. 2014.

[75] S. Fletcher-Watson, S. Leekam, V. Benson, M. Frank, and J. Findlay, "Eye-movements reveal attention to social information in autism spectrum disorder," *Neuropsychologia*, vol. 47, pp. 248–257, Jan. 2009.

[76] C. S. Green and D. Bavelier, "Action video game modifies visual selective attention," *Nature*, vol. 423, pp. 534–537, May 2003.

[77] H. Strasburger, I. Rentschler, and M. Juttner, "Peripheral vision and pattern recognition: A review," *Journal of Vision*, vol. 11, pp. 13–13, Dec. 2011.

[78] S. Caldani, R. Delorme, A. Moscoso, M. Septier, E. Acquaviva, and M. P. Bucci, "Improvement of Pursuit Eye Movement Alterations after Short Visuo-Attentional Training in ADHD," *Brain Sciences*, vol. 10, p. 816, Nov. 2020.

[79] H. Akima, "A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures," *Journal of the ACM*, vol. 17, pp. 589–602, Oct. 1970.

[80] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and M. A. Azim, "Transfer learning: a friendly introduction," *Journal of Big Data*, vol. 9, p. 102, Oct. 2022.

[81] D. J. Berg, S. E. Boehnke, R. A. Marino, D. P. Munoz, and L. Itti, "Free viewing of dynamic stimuli by humans and monkeys," *Journal of Vision*, vol. 9, pp. 19–19, May 2009.

[82] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," Mar. 2019. arXiv:1801.04381 [cs].

[83] M. Harms, "Book review," *Physiotherapy*, vol. 96, p. 82, Mar. 2010.

[84] R. Delgado and X.-A. Tibau, "Why Cohen's Kappa should be avoided as performance measure in classification," *PLOS ONE*, vol. 14, p. e0222916, Sept. 2019.

[85] R. Zemblys, D. C. Niehorster, and K. Holmqvist, "gazeNet: End-to-end eye-movement event detection with deep neural networks," *Behavior Research Methods*, vol. 51, pp. 840–864, Apr. 2019.

[86] C. Elmadjian, C. Gonzales, and C. H. Morimoto, "Eye Movement Classification with Temporal Convolutional Networks," in *Pattern Recognition. ICPR International Workshops and Challenges* (A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, and R. Vezzani, eds.), vol. 12663, pp. 390–404, Cham: Springer International Publishing, 2021.

[87] C. Elmadjian, C. Gonzales, R. L. D. Costa, and C. H. Morimoto, "Online eye-movement classification with temporal convolutional networks," *Behavior Research Methods*, Oct. 2022.

[88] Z. Zhong, H. Fang, H. Zhang, and S. Wu, "Eye Movement Events Detection with KNN-GA and Prior Knowledge," in *2021 International Conference on Computer Engineering and Application (ICCEA)*, (Kunming, China), pp. 468–472, IEEE, June 2021.

[89] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2016.

[90] M. Tan and Q. V. Le, "Efficientnetv2: smaller models and faster training," 2021.

[91] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[92] S. d. Salud, "035. Cinco por ciento de la población infantil y adolescente presenta TDA."

[93] A. M. Hamed, A. J. Kauer, and H. E. Stevens, "Why the Diagnosis of Attention Deficit Hyperactivity Disorder Matters," *Frontiers in Psychiatry*, vol. 6, Nov. 2015.

[94] V. A. Harpin, "The effect of ADHD on the life of an individual, their family, and community from preschool to adult life," *Archives of Disease in Childhood*, vol. 90, pp. i2–i7, Feb. 2005.

[95] S. Gnanavel, P. Sharma, P. Kaushal, and S. Hussain, "Attention deficit hyperactivity disorder and comorbidity: A review of literature," *World Journal of Clinical Cases*, vol. 7, pp. 2420–2426, Sept. 2019.

[96] S. dosReis, C. L. Barksdale, A. Sherman, K. Maloney, and A. Charach, "Stigmatizing experiences of parents of children with a new diagnosis of adhd," *Psychiatric Services*, vol. 61, pp. 811–816, Aug. 2010.

[97] Subcommittee on Attention-Deficit/Hyperactivity Disorder, Steering Committee on Quality Improvement and Management, "Adhd: clinical practice guideline for the diagnosis, evaluation, and treatment of attention-deficit/hyperactivity disorder in children and adolescents," *Pediatrics*, vol. 128, pp. 1007–1022, Nov. 2011.

[98] R. A. Barkley, "Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD.," *Psychological Bulletin*, vol. 121, pp. 65–94, Jan. 1997.

[99] A. Jangmo, A. Stålhandske, Z. Chang, Q. Chen, C. Almqvist, I. Feldman, C. M. Bulik, P. Lichtenstein, B. D'Onofrio, R. Kuja-Halkola, and H. Larsson, "Attention-deficit/hyperactivity disorder, school performance, and effect of medication," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 58, pp. 423–432, Apr. 2019.