



**UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA**



ANÁLISIS DE DATOS DE EYE-TRACKER DURANTE TEXTOS DIGITALES

POR

Ignacio Andrés Fritis Zacconi

Informe de Memoria de Título presentada a la Facultad de Ingeniería de la Universidad de Concepción
para optar al título profesional de Ingeniero Civil Biomédico

Profesor Guía
Esteban Pino Quiroga
Mabel Urrutia Martínez

Agosto 2025
Concepción
(Chile)

© 2025 Ignacio Andrés Fritis Zacconi

© 2025 Ignacio Andrés Fritis Zacconi

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.

Agradecimientos

Agradezco el financiamiento otorgado por la Agencia Nacional de Investigación y Desarrollo (ANID) mediante el Proyecto Fondecyt Regular n° 1241145 (Etapa 2025), titulado: “Entrenando la comprensión lectora online mediante una plataforma virtual”.

A mis padres, por su apoyo constante y por estar siempre presentes. A mi hermano Jorge, por acompañarme en este camino y brindarme su apoyo en cada momento. A Rosario, por su paciencia, cariño y guía en los momentos más difíciles.

Finalmente, expreso mi gratitud a mi profesor guía, Esteban J. Pino, y a mi profesora co-guía, Mabel Urrutia, por su orientación, apoyo y valiosas sugerencias que ayudaron en este trabajo.

Resumen

Esta memoria presenta un estudio basado en un proyecto centrado en el análisis de la lectura digital mediante el uso de tecnología de seguimiento ocular (Eye-Tracking). Se evaluó el comportamiento visual de 41 participantes universitarios, previamente clasificados como lectores hábiles o menos hábiles en comprensión lectora, mientras leían textos digitales estructurados jerárquicamente o en red, y con niveles de dificultad normal o difícil.

Las métricas extraídas incluyeron duración promedio de fijaciones, longitud de sacadas, número de fijaciones por minuto, cantidad de regresiones y tiempo total de lectura. Estas fueron analizadas mediante pruebas estadísticas (t-test, ANOVA bifactorial y modelos lineales mixtos), obteniendo diferencias significativas entre los tipos de lectores en varias de estas métricas. En particular, los lectores hábiles mostraron patrones más eficientes, como fijaciones más cortas y sacadas más largas. Respecto a la dificultad del texto, en las lecturas difíciles aumentaron el tiempo de lectura y la cantidad de fijaciones. No hubo una interacción significativa entre el tipo de lector y la dificultad del texto.

Además, como complemento, se implementaron modelos de Machine Learning y Redes Neuronales para clasificar a los participantes según su tipo de lector, a partir de movimientos oculares durante una lectura. Se entrenaron múltiples modelos (Random Forest, SVC, XGBoost, MLP, entre otros), optimizados con búsqueda bayesiana mediante Optuna y validados con validación cruzada estratificada. El mejor rendimiento se obtuvo con XGBoost, alcanzando un F1-score de 0.753.

Los resultados confirman que las métricas de Eye-Tracking reflejan diferencias lectoras significativas entre lectores hábiles y de bajo desempeño, y que estas métricas pueden ser utilizadas para clasificar a los participantes según su capacidad lectora en entornos digitales.

Abstract

This thesis presents a study focused on digital reading analysis using Eye-Tracking technology. The visual behavior of 41 university students was evaluated, previously classified as either skilled or low-performing readers in reading comprehension, while they read digital texts structured hierarchically or as networks, and with normal or difficult levels of complexity.

The extracted metrics included average fixation duration, saccade length, fixations per minute, number of regressions, and total reading time. These were analyzed using statistical tests (t-test, two-way ANOVA, and mixed linear models), revealing significant differences between reader types for several of these measures. Specifically, skilled readers showed more efficient patterns, such as shorter fixations and longer saccades. Difficult texts increased reading time and the number of fixations. No significant interaction was found between reader type and text difficulty.

In addition, Machine Learning and Neural Network models were implemented to classify participants based on their reading performance using eye movement data. Various models were trained (Random Forest, SVC, XGBoost, MLP, among others), optimized using Bayesian search with Optuna, and evaluated with stratified cross-validation. The best performance was achieved with XGBoost, reaching an F1-score of 0.753.

The results confirm that Eye-Tracking metrics reveal meaningful differences between skilled and low-performing readers and can be used to classify individuals according to their reading ability in digital environments.

Tabla de contenidos

Índice de Tablas	IX
Índice de Figuras.....	X
Capítulo 1. Introducción	1
1.1 INTRODUCCIÓN GENERAL.....	1
1.2 OBJETIVOS	1
1.2.1 <i>Objetivo general</i>	1
1.2.2 <i>Objetivos específicos</i>	2
1.3 ALCANCE Y LIMITACIONES	2
1.3.1 <i>Alcance</i>	2
1.3.2 <i>Limitaciones</i>	2
1.4 METODOLOGÍA.....	3
1.5 TEMARIO	3
Capítulo 2. Marco Teórico	5
2.1 INTRODUCCIÓN	5
2.2 PRINCIPIOS DEL EYE-TRACKING	5
2.2.1 <i>Definiciones</i>	6
2.2.2 <i>Herramientas de Análisis Visual</i>	7
2.2.3 <i>Comportamiento lector: métricas y comprensión</i>	9
2.2.4 <i>Entornos de lectura digital: atención y estructuras</i>	9
2.2.5 <i>Tipos de Eye-Trackers</i>	10
2.2.6 <i>Estudios actuales con Eye-Tracking</i>	11
2.2.7 <i>EyeOn Elite</i>	12
2.2.8 <i>Software Complementario</i>	14
2.3 PRUEBAS ESTADÍSTICAS APLICADAS.....	15
2.4 MODELOS DE MACHINE LEARNING PARA LA CLASIFICACIÓN DE LECTORES	16
2.4.1 <i>Fundamentos del Machine Learning</i>	16

2.4.2	<i>Modelos tradicionales basados en árboles de decisión</i>	16
2.4.3	<i>Optimización de hiperparámetros mediante búsqueda bayesiana</i>	16
2.4.4	<i>Red Neuronal Multicapa</i>	17
2.4.5	<i>Conceptos clave en Redes Neuronales</i>	17
2.4.6	<i>Evaluación y métrica utilizada</i>	18
Capítulo 3. Desarrollo		19
3.1	INTRODUCCIÓN GENERAL	19
3.2	DISEÑO EXPERIMENTAL Y CONDICIONES DE PRESENTACIÓN	19
3.3	CONDICIONES DE CONTRABALANCEO EXPERIMENTAL	21
3.4	CRITERIOS PARA LA SELECCIÓN DE PARTICIPANTES	21
3.5	ADQUISICIÓN DE DATOS MEDIANTE EYE-TRACKING	22
3.6	PROCESAMIENTO DE LOS DATOS	23
3.7	CÁLCULO DE MÉTRICAS VISUALES	24
3.8	ANÁLISIS ESTADÍSTICO	25
3.9	MODELOS DE MACHINE LEARNING Y REDES NEURONALES	27
3.10	DISCUSIÓN	29
Capítulo 4. Resultados		31
4.1	VISUALIZACIÓN DE DISTRIBUCIONES	31
4.2	RESULTADOS ANÁLISIS ESTADÍSTICO	33
4.2.1	<i>Comparaciones por grupo (t-test)</i>	34
4.2.2	<i>ANOVA bifactorial</i>	34
4.2.3	<i>Modelos lineales mixtos</i>	35
4.2.4	<i>Resumen general de resultados</i>	35
4.3	RESULTADOS DE LOS MODELOS DE MACHINE LEARNING Y REDES NEURONALES	35
4.3.1	<i>Balanceo de clases</i>	36
4.3.2	<i>Rendimiento de los modelos</i>	36
4.4	GRÁFICOS DE IMPORTANCIAS DE MÉTRICAS	37
Capítulo 5. Discusión y Conclusiones		42
5.1	DISCUSIÓN DE LOS RESULTADOS DEL ANÁLISIS ESTADÍSTICO	42
5.1.1	<i>Análisis de las Métricas de Lectura</i>	42
5.1.2	<i>Comparación entre Tipos de Texto</i>	42

5.2	DISCUSIÓN DE LOS RESULTADOS DE MODELOS DE MACHINE LEARNING Y REDES NEURONALES	43
5.2.1	<i>Comparación de Desempeño de Modelos</i>	43
5.2.2	<i>Interpretación General</i>	43
5.3	CONCLUSIONES	44
5.4	LÍNEAS FUTURAS DE INVESTIGACIÓN.....	45
Glosario		46
Referencias		47

Índice de Tablas

3-1	Características de los textos utilizados en el experimento.	20
3-2	Condiciones de contrabalanceo experimental	21
3-3	Promedio de caracteres por palabra en los textos utilizados.	24
4-4	Valores p de las pruebas t de Student para cada métrica comparando buenos y malos lectores.	34
4-5	Valores p del Análisis de Varianza (ANOVA) bifactorial por tipo de lector, dificultad del texto e interacción.	34
4-6	Resultados de los Modelos Lineales Mixtos (MLM) con interacción tipo lector × dificultad texto	35
4-7	Resumen general de significancia estadística por métrica	35
4-8	Resultados de modelos ML y RN con sample weights	36
4-9	Resultados de modelos ML y RN con SMOTE	37
4-10	Resultados de modelos ML y RN sin balanceo	37

Índice de Figuras

2.1	Ejemplo de la pupila de una persona a distintos niveles de dilatación, tomado por un eye tracker.	6
2.2	Ejemplo de fijaciones y movimientos sacádicos.	7
2.3	Ejemplo de mapa de calor.	8
2.4	Ejemplo de scanpath.	8
2.5	Tablet EyeOn Elite de EyeTech.	13
3.1	Participante leyendo un texto en la tablet EyeOn Elite.	23
4.1	Duración promedio de fijaciones según tipo de lector y dificultad del texto.	31
4.2	Tiempo total de lectura según tipo de lector y dificultad del texto.	32
4.3	Cantidad de fijaciones por minuto según tipo de lector y dificultad del texto.	32
4.4	Longitud media de sacadas según tipo de lector y dificultad del texto.	33
4.5	Cantidad de regresiones según tipo de lector y dificultad del texto.	33
4.6	Importancia de características del modelo XGBoost.	38
4.7	Distribución de la desviación estándar de la duración de fijaciones según tipo de lector. .	39
4.8	Relación entre desviación estándar de duración de fijaciones y cantidad de fijaciones por minuto.	40
4.9	Relación entre longitud media y duración media de movimientos sacádicos.	41

Capítulo 1. Introducción

1.1 Introducción general

En la actualidad, la lectura digital ha cambiado la manera en que las personas interactúan con la información. A medida que los textos en línea y los hipertextos se vuelven más comunes, hay una mayor necesidad de estudiar cómo estas nuevas formas de lectura afectan la comprensión y el comportamiento lector, particularmente en lectores con diferentes niveles de habilidad.

Este proyecto analiza los patrones de lectura en dos modalidades de textos digitales: una modalidad de tipo tradicional, en la que los hipervínculos despliegan definiciones o imágenes en formato pop-up sin abandonar la página principal; y una modalidad que simula la navegación en línea, en la que los hipervínculos redirigen a otras páginas con contenido complementario. Mediante el uso de una tablet EyeOn Elite de EyeTech, se registran los movimientos oculares de los participantes con el fin de analizar métricas como fijaciones, movimientos sacádicos y regresiones, y explorar su relación con la comprensión lectora.

El proyecto compara los patrones de lectura entre dos grupos de participantes: lectores con un buen nivel de comprensión lectora y lectores con desempeño más bajo. Se espera que las diferencias en las métricas visuales registradas mediante Eye-Tracking (Seguimiento Ocular, ET) ofrezcan información relevante sobre cómo ambos grupos interactúan con los textos digitales, y cómo estas interacciones visuales influyen en su capacidad para procesar y comprender la información presentada.

1.2 Objetivos

1.2.1 Objetivo general

Analizar los patrones de comportamiento visual durante la lectura digital mediante un ET y análisis estadísticos, con el fin de relacionar estas métricas con la capacidad lectora de los participantes.

1.2.2 Objetivos específicos

- Obtener datos de ET durante la lectura digital, la cual incluía textos con distintos niveles de dificultad y diferentes tipos de estructura.
- Relacionar los datos obtenidos del ET con las actividades de lectura y habilidad lectora.
- Realizar análisis estadísticos para extraer características y relacionarlas con la capacidad lectora.

1.3 Alcance y limitaciones

1.3.1 Alcance

El alcance de esta memoria comprende al diseño, implementación y análisis de un experimento orientado a estudiar el comportamiento lector en entornos digitales utilizando tecnología de ET. El estudio se centró en la construcción y validación de textos digitales con estructuras jerárquica y en red, ambos con distintos niveles de dificultad, y su aplicación en condiciones controladas con contrabalanceo experimental.

Se analizaron los movimientos oculares de 41 participantes clasificados previamente según su nivel de comprensión lectora, utilizando pruebas estandarizadas. El trabajo incluyó la recolección de datos visuales, su procesamiento y el análisis estadístico de métricas como duración y número de fijaciones, longitud de sacadas y número de regresiones. Este estudio se limitó al análisis transversal de los datos obtenidos en una única sesión experimental por participante.

1.3.2 Limitaciones

Entre las posibles limitaciones del estudio se encuentra la variabilidad en las condiciones individuales de los participantes, como el nivel de interés en los textos o la fatiga durante la lectura, que podrían influir en los resultados. Además, el uso de una Tablet con ET integrado puede presentar ciertos tipos de problemas, como la calibración adecuada para todos los participantes y la precisión de los registros de movimientos oculares en condiciones de variación de la postura o movimiento de cabeza. La reducida cantidad de participantes puede limitar la generalización de los resultados, aunque se espera que los patrones observados sean representativos de las diferencias entre lectores hábiles y menos hábiles en entornos digitales.

1.4 Metodología

Para llevar a cabo este estudio, se utilizó una tablet EyeOn Elite. Esta tablet cuenta con un ET integrado que registra la mirada del lector a una frecuencia de muestreo de 42 Hz, permitiendo un análisis del comportamiento visual durante la lectura.

Antes del experimento principal, los participantes fueron evaluados mediante dos pruebas estandarizadas de comprensión lectora: Lectum e Interamericana. Los puntajes obtenidos en estas pruebas permitieron clasificarlos en dos grupos: lectores hábiles y lectores menos hábiles. Esta clasificación fue utilizada como base para comparar los patrones de comportamiento visual durante la lectura digital. Posteriormente, se midieron los movimientos oculares de los participantes mientras leían cuatro textos (uno con estructura jerárquica y dificultad normal, otro con estructura en red y dificultad normal, y dos textos adicionales con estas mismas estructuras pero con un nivel de dificultad mayor), utilizando el ET para capturar la mirada del participante en coordenadas X e Y.

El paquete de softwares Mangold Vision fue utilizado para diseñar el experimento, presentar los textos, registrar y analizar los datos de movimientos oculares. Los datos recolectados fueron exportados desde el software con el propósito de llevar a cabo un análisis estadístico que permitiera identificar patrones de comportamiento visual asociados a las diferencias en la capacidad lectora entre lectores hábiles y menos hábiles.

Se llevó a cabo un proceso de filtrado y preprocesamiento de los datos para identificar y corregir cualquier variable fuera de lo común que pudiera afectar el análisis. Para el manejo de los datos obtenidos mediante el ET, se utilizó Python con el fin de optimizar y reducir el tamaño de la base de datos, haciéndola más manejable para su análisis. Asimismo, Python fue la herramienta principal para realizar el análisis estadístico, utilizando la biblioteca *Scipy*.

Por último se utilizaron modelos de Machine Learning (Aprendizaje Automático, ML) y Redes Neuronales (RN) para clasificar a los participantes según su capacidad lectora a partir de sus movimientos oculares, basándose en las métricas obtenidas del ET.

1.5 Temario

- **Capítulo 1. Introducción:** Introducción de manera general indicando la metodología, objetivos y alcances del proyecto.

- **Capítulo 2. Marco Teórico:** En esta sección se tiene el marco teórico, en el cual se proporciona un contexto integral sobre diversos aspectos relevantes del ET, su funcionamiento, utilizaciones y estudios actuales de este.
- **Capítulo 3. Proyecto:** En este capítulo se presenta el diseño e implementación del experimento de lectura digital mediante ET, incluyendo la descripción de los textos, el procedimiento de recolección de datos y el análisis posterior realizado con base en las métricas visuales obtenidas.
- **Capítulo 4. Resultados:** En este capítulo se presentan los resultados obtenidos del experimento, incluyendo gráficos y tablas que ilustran las métricas visuales analizadas.
- **Capítulo 5. Discusión y Conclusiones:** En este capítulo se discuten los resultados obtenidos, se reflexiona sobre su significado en el contexto del estudio y se sugieren posibles líneas de investigación futura.

Capítulo 2. Marco Teórico

2.1 Introducción

El ET se ha consolidado como una herramienta clave para el estudio del comportamiento humano en diversos contextos, desde la comprensión lectora hasta la navegación en entornos digitales. Los avances recientes en la tecnología han permitido que esta técnica sea más precisa y accesible [1], incrementando su uso en investigaciones de psicología, educación, marketing y neurociencias [2–4]. En este capítulo, se explorarán los fundamentos teóricos del ET, su evolución histórica, las tecnologías actuales y sus aplicaciones en la investigación.

2.2 Principios del Eye-Tracking

Los movimientos oculares durante la lectura no son continuos, sino que se caracterizan por una alternancia entre fijaciones y movimientos sacádicos. Durante las fijaciones, los ojos permanecen relativamente inmóviles mientras el cerebro procesa la información visual. En cambio, los movimientos sacádicos son desplazamientos rápidos que permiten saltar de un punto de fijación a otro [3–5]. Este patrón resulta especialmente relevante en tareas como la lectura, donde los ojos se mueven de palabra en palabra, pero solo ciertas regiones del texto reciben atención focalizada [4].

Para registrar estos movimientos, los ET más utilizados se basan en la técnica de reflexión corneal y el seguimiento de la pupila mediante cámaras infrarrojas. Este método emplea cámaras especializadas que proyectan luz infrarroja invisible al ojo humano, la cual genera un reflejo en la córnea conocido como reflejo de Purkinje [6, 7]. Este reflejo facilita el seguimiento preciso de la posición ocular [3], como se muestra en la Figura 2.1.

A partir de la posición relativa entre la pupila y el reflejo corneal, el dispositivo calcula el denominado Punto de Regard (POR), es decir, el punto exacto del espacio visual hacia donde se dirige la mirada del usuario [8]. Las cámaras del ET capturan imágenes del ojo en tiempo real, y mediante software especializado se estiman las coordenadas precisas de la mirada, permitiendo identificar las Areas of Interest (Áreas de Interés, AOIs) dentro del estímulo presentado [2].

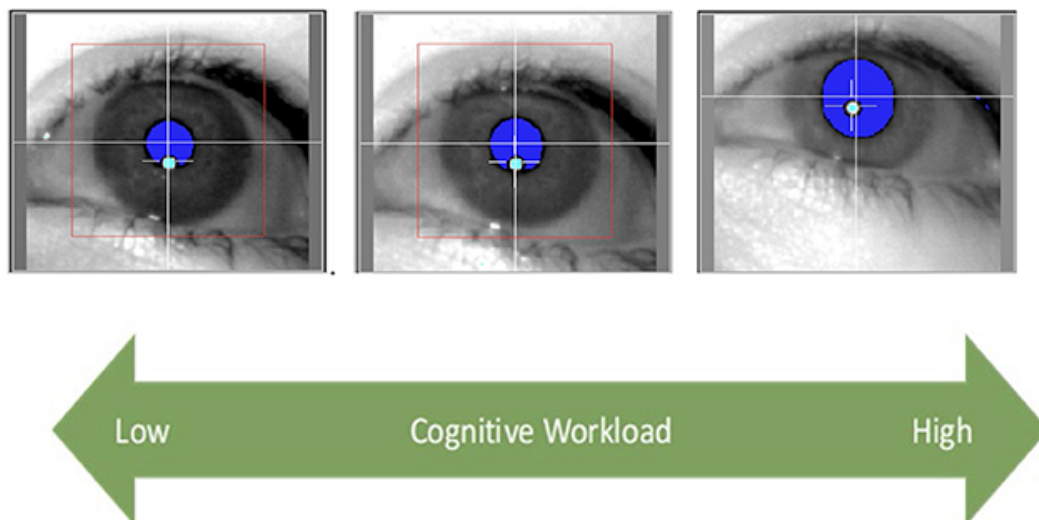


Figura 2.1: Ejemplo de la pupila de una persona a distintos niveles de dilatación, tomado por un eye tracker. [9].

2.2.1 Definiciones

A continuación, se presentan algunas definiciones clave relacionadas con el ET:

- **Estímulos visuales:** Representan los elementos presentados al participante durante un experimento, como palabras, imágenes, videos o páginas web. Los datos obtenidos por el ET se mapean en relación con estos estímulos, permitiendo identificar AOIs en elementos específicos del texto o pantalla [2, 3, 5]. Estas áreas son fundamentales para analizar la interacción entre la mirada y los estímulos visuales.
- **Fijaciones:** Momentos en los que los ojos permanecen estáticos sobre un punto específico. Las fijaciones indican las áreas donde se concentra la atención visual y estas son cruciales para entender la cantidad de esfuerzo cognitivo que un área específica requiere. Por ejemplo, palabras difíciles o conceptos complejos tienden a provocar fijaciones más largas [3, 4]. En la Figura 2.2 se pueden observar las fijaciones como puntos dentro del AOIs.
- **Movimientos sacádicos:** Desplazamientos rápidos de los ojos entre fijaciones. Estos movimientos son esenciales para la lectura, ya que permiten al lector saltar de una palabra a otra. La longitud y velocidad de estos movimientos pueden indicar la fluidez lectora y la familiaridad con el contenido [3, 4]. En la Figura 2.2 se pueden observar los movimientos sacádicos como líneas que conectan las fijaciones.

- **Regresiones:** Movimientos oculares que regresan a un punto previamente fijado. Las regresiones son comunes en la lectura, especialmente cuando el lector necesita aclarar o revisar información. Un alto número de regresiones puede indicar dificultades de comprensión o falta de fluidez [3, 4].
- **Duración de la fijación inicial:** Es el tiempo que los ojos permanecen en un punto por primera vez antes de moverse hacia otra área. Este valor puede ser especialmente útil para analizar la atención inicial a palabras clave o elementos visuales destacados [4].
- **Duración total de la fijación:** Suma del tiempo dedicado a una palabra o elemento visual durante múltiples fijaciones. Es útil para evaluar la importancia percibida de un estímulo [3].

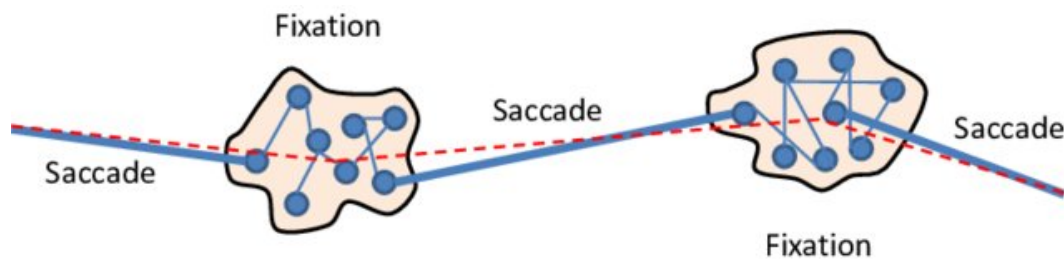


Figura 2.2: Ejemplo de fijaciones y movimientos sacádicos [10].

Las fijaciones y movimientos sacádicos son patrones fundamentales del movimiento ocular [7, 11]. La duración de las fijaciones y el tamaño y velocidad de las sacadas dependen del tipo de tarea realizada. Por ejemplo, durante la lectura, las fijaciones tienden a ser más largas y las sacadas más cortas que en tareas de exploración visual como búsqueda de objetos [3, 4].

2.2.2 Herramientas de Análisis Visual

Los datos registrados por los ET pueden representarse mediante herramientas visuales que facilitan su interpretación. Entre las herramientas más utilizadas se encuentran:

- **Mapas de Calor:** Son representaciones gráficas que muestran las áreas donde el usuario ha pasado más tiempo mirando [12]. Las zonas de mayor atención se visualizan con colores cálidos (rojo), mientras que las zonas con menor atención aparecen en tonos verdes [2, 5]. Esta representación gráfica con los cambios de tonalidad en los colores se puede ver en la Figura 2.3.

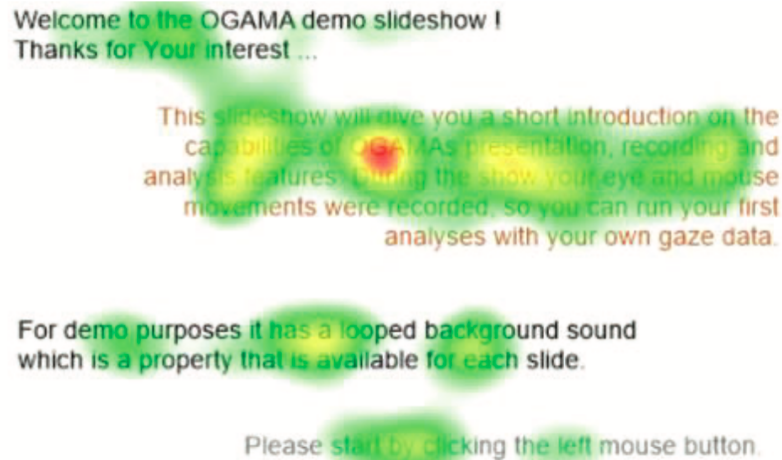


Figura 2.3: Ejemplo de mapa de calor [5].

- **Scanpaths:** Son representaciones del recorrido que sigue la mirada del usuario a lo largo de un estímulo visual [13]. Los scanpaths permiten visualizar el orden en que se exploraron diferentes partes del estímulo [2,5]. En la Figura 2.4 se puede observar el recorrido que se forma por la mirada del usuario.

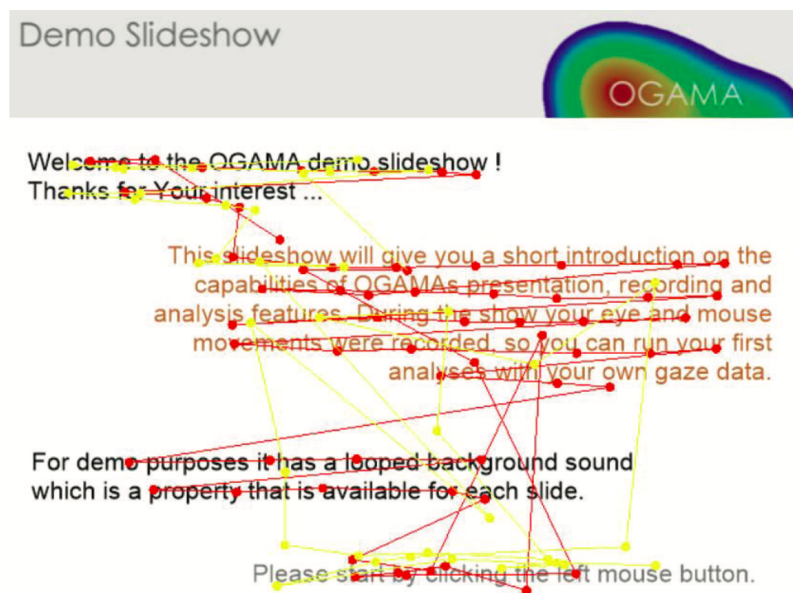


Figura 2.4: Ejemplo de scanpath [5].

Estas herramientas no solo permiten representar visualmente los patrones de atención, sino que también se utilizan como base para analizar tareas cognitivas complejas como la lectura.

2.2.3 Comportamiento lector: métricas y comprensión

El ET permite medir el nivel de comprensión lectora. Las estrategias de lectura varían dependiendo del propósito, el tipo de texto y la habilidad lectora del individuo [14]. Los ET pueden identificar comportamientos como las regresiones oculares, que suelen indicar dificultades en la comprensión [3].

El ET permite investigar cómo los lectores procesan palabras a nivel individual. Los investigadores pueden medir el tiempo de fijación en palabras específicas y cómo este varía según su familiaridad, longitud o contexto. Los lectores hábiles realizan menos fijaciones y movimientos sacádicos más rápidos, mientras que los lectores menos hábiles suelen hacer más regresiones y fijaciones más largas [4, 14].

Un estudio reciente destaca la utilidad del ET para analizar tanto procesos cognitivos como emocionales mediante métricas como la duración de fijaciones, las sacadas, el tiempo hasta la primera fijación (TFF) y la dilatación pupilar. Estas métricas permiten inferir estados como la carga cognitiva, la atención visual, la toma de decisiones, así como emociones como ansiedad o valencia afectiva. Por ejemplo, una mayor duración de fijaciones puede indicar una mayor carga cognitiva o procesamiento emocional profundo, mientras que cambios en el diámetro pupilar pueden reflejar estados de activación emocional. El estudio resalta la importancia de combinar métricas oculares con otros métodos como cuestionarios o señales fisiológicas para una interpretación más robusta [15].

2.2.4 Entornos de lectura digital: atención y estructuras

En entornos digitales, los elementos no textuales (imágenes, hipervínculos, anuncios) influyen significativamente en los patrones de lectura [16]. Los mapas de calor generados por ET permiten identificar cómo los lectores distribuyen su atención en páginas web [17], o como se visualiza en el scanpath (Figura 2.4), estos patrones también informan cómo se distribuye la atención en entornos digitales.

En el contexto de la lectura digital, los textos pueden organizarse siguiendo distintas estructuras hipertextuales, siendo las más representativas la **estructura jerárquica** y la **estructura en red**. Estas estructuras no solo modifican la forma en que la información es presentada, sino que también inciden directamente en la manera en que los lectores navegan por los contenidos y procesan la información.

La estructura jerárquica se caracteriza por ofrecer una navegación dirigida, basada en nodos conectados desde un menú principal, lo cual facilita una ruta de lectura más lineal y predecible. Esta estructura suele beneficiar a lectores menos hábiles, ya que se asemeja más a los textos tradicionales impresos y exige menos habilidades de autorregulación [18].

Por otro lado, la estructura en red permite una navegación no lineal, en la que los enlaces se interconectan libremente sin un orden predefinido. Esta estructura pone a prueba la capacidad de los lectores para tomar decisiones estratégicas, identificar la relevancia de la información y construir su propio recorrido textual, lo que puede favorecer a lectores hábiles, pero resultar desorientador para quienes poseen una menor habilidad de comprensión [18].

Diversos estudios han mostrado que estas estructuras afectan tanto las estrategias de navegación (por ejemplo, cantidad de enlaces visitados, número de regresos a la página principal) como los resultados de comprensión lectora, siendo la estructura jerárquica generalmente más accesible para lectores con menores habilidades, mientras que la estructura en red exige un mayor control cognitivo y habilidad de planificación [19,20].

2.2.5 Tipos de Eye-Trackers

- **ET de Mesa:** Este tipo de dispositivo se instala frente al participante, generalmente debajo de una pantalla. Es ideal para estudios en entornos controlados, como experimentos de lectura o navegación web. Un ejemplo es el EyeLink Portable Duo, que ofrece una alta precisión y una frecuencia de muestreo de hasta 2000 Hz [21].
- **ET Integrados:** Se encuentran incorporados en dispositivos de consumo, como laptops, teléfonos inteligentes o gafas de Realidad Virtual (VR). Aunque menos precisos que los dispositivos dedicados, son más accesibles y permiten aplicaciones masivas, como análisis de usabilidad de interfaces [22]. Ejemplos incluyen los sistemas de Tobii Pro Glasses 3, los integrados en dispositivos como Meta Quest [23], y también el EyeOn Elite, que combina un ET con una tablet para facilitar la comunicación y el acceso a la tecnología para personas con discapacidades [24].
- **ET Basados en Webcam:** Utilizan cámaras estándar para estimar la mirada del usuario, generalmente con menor precisión. Son populares en aplicaciones donde el costo es un factor clave, como la investigación de mercados o el análisis de atención en plataformas en línea [25].

2.2.6 Estudios actuales con Eye-Tracking

Eye-Tracking en la medicina

El ET se utiliza para evaluar la expertise visual en el diagnóstico médico. Estudios recientes han mostrado que los médicos expertos, como los cardiólogos, tienen patrones de movimientos oculares más eficientes que los novatos. Estos estudios analizan fijaciones, movimientos sacádicos y trayectorias oculares para mejorar la formación médica, proporcionando retroalimentación sobre las estrategias visuales más efectivas en la interpretación de imágenes médicas como Electrocardiograma (ECG) o resonancias magnéticas. Además, el ET se ha usado en la rehabilitación médica para monitorear el progreso de pacientes con lesiones cerebrales, ayudando a mejorar su comunicación no verbal [26–28].

Eye-Tracking en la educación

El ET también es útil para analizar la carga cognitiva de los estudiantes durante la lectura [29]. Permite a los investigadores medir la dilatación de las pupilas, como se observa en la Figura 2.1, y como se correlaciona con la dificultad percibida al leer textos complejos [9]. También es utilizado para comprender como el movimiento de los ojos ayuda en el pensamiento matemático [30].

Realidad Virtual y Aumentada

Los dispositivos de ET integrados en *headsets* de VR, como Meta Quest, mejoran la interacción usuario-dispositivo al rastrear la mirada para navegar menús o activar funciones [23]. Además, esta tecnología se está integrando en herramientas de aprendizaje para medir la atención y evaluar la carga cognitiva [22, 31].

Eye-Tracking en el estudio de la lectura

El ET también se utiliza para investigar la comprensión lectora. Estudios recientes han medido cómo los lectores procesan la información textual y cómo varía su atención según la dificultad del texto [9, 14, 32]. También, en otro estudio se destaca el uso de ET en psicología del lenguaje para analizar cómo las personas procesan textos no manipulados experimentalmente, permitiendo nuevos modelos

computacionales de comprensión del lenguaje. Esto fomenta estudios translingüísticos y comparativos sobre diferencias individuales entre lectores [33]

Uso del Eye-Tracking en la predicción de la comprensión lectora

En un estudio reciente se investigó cómo las métricas obtenidas mediante ET pueden servir como predictores de la comprensión lectora. Para ello, se evaluó a 79 adultos con tres pruebas distintas de comprensión lectora, mientras se registraban sus movimientos oculares. Estas pruebas variaban en sus demandas cognitivas e incluían lectura silenciosa y lectura en voz alta. Durante la tarea, se midieron parámetros como el número y duración de fijaciones, la cantidad de movimientos sacádicos y las regresiones oculares. Los investigadores aplicaron modelos estadísticos de regresión lineal para analizar la relación entre estas métricas y los puntajes obtenidos en las pruebas. Los resultados mostraron que las variables extraídas del ET explicaban una proporción significativamente mayor de la variabilidad en la comprensión lectora, en comparación con predictores tradicionales como la velocidad de lectura. Este hallazgo resalta el potencial del ET como herramienta más precisa y sensible para evaluar la comprensión lectora que los métodos comúnmente utilizados [14].

Combinación de modelos de lenguaje y Eye-Tracking para clasificación del estado neural

Otro estudio combinó ET con Electroencefalograma (EEG) y modelos de lenguaje de gran escala como GPT para analizar la comprensión lectora a nivel de palabra. Se identificaron estados mentales mientras los participantes leían palabras con diferentes niveles de relevancia semántica en relación con una palabra clave. Las palabras de alta relevancia recibieron significativamente más fijaciones, lo que sugiere que las métricas de ET pueden ofrecer una visión profunda del procesamiento semántico durante la lectura [34].

2.2.7 EyeOn Elite

El EyeOn Elite es un sistema portátil de ET desarrollado por EyeTech Digital Systems, diseñado específicamente para facilitar la comunicación y el acceso a la tecnología a personas con discapacidades motoras o del habla. Este dispositivo integra una pantalla táctil de 12.5 pulgadas, un sistema de ET incorporado y sistema operativo Windows 10 o 11, todo en un solo hardware, lo que lo convierte en una

herramienta versátil tanto para usuarios como para terapeutas y educadores [24].

El dispositivo se monta fácilmente sobre una base ajustable o en sillas de ruedas, y su ET está configurado a una frecuencia de muestreo de 42 Hz. Está pensado para ser compatible con software de comunicación aumentativa y alternativa (CAA) como Snap Core First, Grid 3 y otros, permitiendo a los usuarios interactuar con su entorno digital de forma intuitiva [24].

Su portabilidad, facilidad de instalación y compatibilidad con software especializado lo hacen una herramienta útil para estudios en contextos reales, como entornos educativos o clínicos, donde se requiere movilidad y rápida puesta en marcha.

La siguiente figura muestra el dispositivo EyeOn Elite.



Figura 2.5: Tablet EyeOn Elite de EyeTech.

En este estudio se utilizó la Tablet EyeOn Elite de EyeTech, su diseño portátil permite la recolección de datos en contextos móviles o clínicos sin requerir equipamiento adicional. La calibración se realiza a través del software Mangold Vision, y el sistema ofrece buena precisión incluso en condiciones de iluminación variables.

2.2.8 Software Complementario

El sistema EyeOn Elite fue utilizado en conjunto con el paquete de softwares Mangold Vision, estos están especializados para la recolección y análisis de datos de ET. Esta plataforma se compone de tres módulos principales que trabajan de forma integrada para facilitar todo el flujo de trabajo experimental: diseño de estímulos, recolección de datos y análisis de resultados.

Mangold Vision Designer

Este módulo permite diseñar las presentaciones experimentales, definiendo los estímulos visuales que serán mostrados al participante, tales como imágenes, textos, videos o páginas web. Su interfaz permite crear múltiples presentaciones y asignarlas a distintos participantes del estudio, lo que es especialmente útil en proyectos con grandes volúmenes de datos [35].

Mangold Vision Recorder

El módulo Recorder ejecuta los estímulos diseñados previamente y registra los datos de ET en tiempo real. El investigador debe seleccionar el proyecto y el participante correspondiente antes de comenzar la prueba. Este software controla el funcionamiento del ET y asegura que la presentación de los estímulos esté sincronizada con la adquisición de datos [35].

Mangold Vision Analyzer

Finalmente, el módulo Analyzer permite revisar y analizar los datos obtenidos durante la sesión. Proporciona herramientas para visualizar mapas de calor, AOIs, trayectorias oculares y otras métricas. Además, permite importar datos de ET de terceros, lo que lo convierte en una herramienta versátil para la investigación en comportamiento visual [35].

Una característica importante del Analyzer es su capacidad para aplicar métodos de agregación de fijaciones. Esta funcionalidad permite consolidar múltiples puntos de mirada en eventos de fijación mediante parámetros como el tamaño del clúster espacial y la ventana temporal. Al ajustar estos valores, es posible suavizar los datos y obtener métricas más estables, pero también se reduce la resolución temporal del ET.

2.3 Pruebas Estadísticas Aplicadas

En este estudio se aplicaron diversas pruebas estadísticas para analizar las diferencias entre condiciones experimentales en tareas de lectura. La elección de cada prueba dependió de las características de los datos, especialmente en tres aspectos fundamentales:

- **Normalidad de las distribuciones:** Se refiere a si los datos siguen una distribución normal (en forma de campana), el cual tiene importancia para muchas pruebas estadísticas paramétricas.
- **Homogeneidad de varianzas:** Implica que los distintos grupos a comparar presentan una variabilidad similar en sus datos. Esto es importante para pruebas como el t-test.
- **Estructura del diseño experimental:** Hace referencia a cómo están organizadas las variables independientes y cómo se agrupan o repiten las observaciones (por ejemplo, si hay medidas repetidas dentro de un mismo participante o si los factores son cruzados). Esto determinó el uso de modelos como el Análisis de Varianza (ANOVA) bifactorial o los Modelos Lineales Mixtos (MLM).

Para evaluar estos aspectos, se realizaron las siguientes pruebas estadísticas:

- **Prueba de Shapiro-Wilk:** Utilizada para verificar la normalidad de los datos. Esta prueba es adecuada para tamaños de muestra pequeños o moderados y fue fundamental para decidir entre el uso de pruebas paramétricas o no paramétricas [36].
- **Prueba de Levene:** Aplicada para evaluar la homogeneidad de varianzas entre grupos. Su resultado determinó si era apropiado utilizar pruebas t paramétricas [36].
- **Prueba t de Student:** Se utilizó para comparar las medias de dos grupos independientes cuando se cumplían los supuestos de normalidad y homogeneidad de varianzas [36].
- **ANOVA bifactorial:** Se empleó para analizar el efecto conjunto de dos factores independientes sobre una variable dependiente, así como sus posibles interacciones. Esta prueba es útil para evaluar si los efectos de un factor cambian según los niveles del otro [37].
- **MLM:** Este modelo estadístico se utilizó para tener en cuenta tanto efectos fijos (como dificultad del texto o tipo de lector) como efectos aleatorios (como el código del participante). Los MLM permiten analizar datos con estructuras jerárquicas o medidas repetidas, siendo especialmente útiles cuando los datos están agrupados o hay dependencia intra-sujeto [38].

Estas pruebas proporcionaron una base robusta para la interpretación de los datos recogidos mediante ET y permitieron evaluar diferencias significativas entre condiciones experimentales con distintos niveles de complejidad.

2.4 Modelos de Machine Learning para la Clasificación de Lectores

Con el fin de explorar enfoques complementarios a los métodos estadísticos tradicionales, en este estudio se implementaron modelos de ML y RN para predecir la habilidad lectora de los participantes a partir de métricas extraídas del ET.

2.4.1 Fundamentos del Machine Learning

El ML permite construir modelos que aprenden patrones directamente a partir de los datos. En tareas de clasificación, como en este caso, el objetivo es predecir una etiqueta (lector hábil o menos hábil) en función de un conjunto de características. En este estudio, se utilizaron como variables predictoras distintas métricas visuales registradas durante la lectura. [39]

2.4.2 Modelos tradicionales basados en árboles de decisión

Se entrenaron modelos como Random Forest, Gradient Boosting y otros algoritmos de árboles, los cuales son adecuados para datos tabulares con relaciones no lineales. Estos modelos fueron optimizados mediante búsqueda bayesiana usando la biblioteca Optuna, y evaluados con validación cruzada estratificada por lector (StratifiedGroupKFold), lo que permitió preservar la independencia entre participantes en los conjuntos de entrenamiento y prueba. [40]

2.4.3 Optimización de hiperparámetros mediante búsqueda bayesiana

Los hiperparámetros son valores configurables del modelo que no se ajustan automáticamente durante el entrenamiento, sino que deben establecerse de antemano. Ejemplos comunes incluyen la tasa de aprendizaje, el número de capas ocultas y el tamaño del lote. La elección adecuada de estos valores puede tener un impacto significativo en el rendimiento final del modelo.

Para mejorar el rendimiento de los modelos, se aplicó una técnica de optimización de hiperparámetros conocida como *búsqueda bayesiana*, implementada a través de la biblioteca Optuna. A diferencia de métodos tradicionales como la búsqueda exhaustiva (grid search) o aleatoria (random search), la búsqueda bayesiana construye un modelo probabilístico del espacio de hiperparámetros y lo utiliza para seleccionar de forma inteligente los valores más prometedores.

Este enfoque permite explorar de manera más eficiente el espacio de búsqueda, reduciendo el número de iteraciones necesarias para encontrar configuraciones óptimas. En este estudio, se utilizaron funciones objetivo que incluían validación cruzada estratificada por grupo y la maximización de la métrica F1 ponderado. [41]

2.4.4 Red Neuronal Multicapa

Además de los modelos tradicionales, se implementó una Multilayer Perceptron (Red Neuronal Multicapa, MLP) utilizando PyTorch. Este tipo de red es capaz de modelar relaciones complejas entre las variables de entrada y la etiqueta de clase. [42]

2.4.5 Conceptos clave en Redes Neuronales

Para mejorar el rendimiento y la generalización de la MLP, se emplearon técnicas comunes en el entrenamiento de modelos:

- **Dropout:** Técnica de regularización que consiste en desactivar aleatoriamente un porcentaje de neuronas durante el entrenamiento. Esto evita que el modelo se sobreajuste a los datos de entrenamiento y mejora su capacidad de generalizar [42].
- **Normalización por lotes (Batch Normalization):** Método que normaliza las salidas de cada capa intermedia para mantener una distribución estable de los valores a lo largo del entrenamiento. Esto acelera la convergencia y permite usar tasas de aprendizaje más altas [42].
- **Validación cruzada:** Estrategia que consiste en dividir el conjunto de datos en varios pliegues o grupos. El modelo se entrena en algunos pliegues y se valida en otros, rotando el proceso. Esto permite evaluar la capacidad de generalización del modelo de forma robusta.

2.4.6 Evaluación y métrica utilizada

Los modelos fueron evaluados usando la métrica de F1 ponderado, calculada a nivel de lectura. Esto permitió balancear el impacto de cada clase en el rendimiento general del modelo, especialmente en presencia de clases desbalanceadas.

Capítulo 3. Desarrollo

3.1 Introducción general

Este capítulo describe el diseño e implementación de un experimento de lectura digital mediante ET, cuyo objetivo fue comparar los patrones de comportamiento visual entre lectores con distinto nivel de comprensión lectora. Para ello, se utilizaron textos con diferentes niveles de dificultad y estructuras de navegación (jerárquica y en red), presentados bajo cuatro condiciones de contrabalanceo.

A continuación, se detallan las decisiones experimentales, el procesamiento de los textos, el procedimiento de recolección de datos y el análisis posterior realizado con base en las métricas visuales obtenidas.

3.2 Diseño experimental y condiciones de presentación

El experimento fue implementado utilizando el software Mangold Vision Designer y se utilizó Mangold Vision Recorder para grabar el experimento en una tablet EyeOn Elite. A cada participante se le asignó una de las cuatro condiciones de contrabalanceo, y debió leer cuatro textos consecutivos según la estructura (jerárquica o en red) y el nivel de dificultad (normal o difícil) establecidos en la condición asignada. La lectura se realizó en un entorno controlado, con iluminación adecuada y sin distracciones externas.

Los textos seleccionados para el experimento fueron diseñados para representar los niveles de dificultad y tipos de estructura de navegación. La dificultad de los textos fue evaluada utilizando la plataforma legible.es, la cual proporciona métricas basadas en los índices de INFLESZ y Fernández-Huerta. Los textos difíciles presentaban menor puntuación en estas escalas, indicando mayor complejidad léxica y sintáctica.

Para asegurar una presentación estandarizada, los textos fueron convertidos a archivos HTML y presentados en la tablet EyeOn Elite mediante el software Mangold Vision Recorder. Los textos de estructura jerárquica fueron diseñados para ser leídos de forma lineal y contenían hipervínculos que, al hacer clic, desplegaban definiciones o imágenes en formato pop-up sin abandonar la página principal. En contraste, los textos de estructura en red permitían una navegación no lineal, donde los hipervínculos llevaban a

otras páginas con información complementaria, incluyendo definiciones, imágenes o videos.

Los hipervínculos fueron seleccionados mediante un modelo GPT, configurado para identificar palabras con baja familiaridad y alta sofisticación léxica. En los textos jerárquicos, el 70 % de los enlaces correspondía a definiciones y el 30 % a imágenes. En los textos en red, ese 30 % se repartía entre imágenes y videos. Este diseño buscó simular escenarios de lectura digital interactiva.

A continuación, se detallan las características de cada texto, incluyendo su dificultad, estructura y número de hipervínculos.

Tabla 3-1: Características de los textos utilizados en el experimento.

Texto	Dificultad	Estructura	Nº hipervínculos
Cómo crear un club de lectura con tu familia	Normal	Jerárquico	25
El tiempo frente a las pantallas ha aumentado: así puedes reconducirlo hacia la lectura	Normal	Red	30
Bases neurológicas de la lectura: descubrimientos recientes	Difícil	Jerárquico	18
¿Está la habilidad metafórica literalmente en el cerebro?	Difícil	Red	20

Al final de cada texto, se incluyeron tres preguntas de selección múltiple diseñadas para evaluar la comprensión lectora del participante. Estas preguntas fueron guardadas en formato CSV.

Las preguntas se diseñaron para evaluar distintos niveles de comprensión lectora. Cada conjunto incluía una pregunta explícita (cuya respuesta se encontraba directamente en el texto), una pregunta inferencial (que requería deducir información a partir de múltiples fragmentos del texto), y una pregunta implícita (que apelaba a la interpretación del lector basada en el contexto y conocimientos previos).

Para la creación de las imágenes asociadas a los hipervínculos en los textos jerárquicos, se utilizó la herramienta de generación de imágenes DALL·E, la cual permitió generar ilustraciones contextualizadas a partir de cada palabra seleccionada, asegurando coherencia visual con el contenido. En cambio, para los textos en red, se utilizaron imágenes de dominio público y videos de YouTube.

3.3 Condiciones de contrabalanceo experimental

Para controlar los efectos del orden de presentación y minimizar posibles sesgos de aprendizaje o fatiga, se establecieron cuatro condiciones de contrabalanceo. Cada condición define un orden distinto en que los participantes leen los textos, variando tanto el nivel de dificultad inicial como el tipo de estructura de lectura.

Tabla 3-2: Condiciones de contrabalanceo experimental

Condición	Orden de lectura (Dificultad, Estructura)
C1	Normal Jerárquico → Normal En red → Difícil Jerárquico → Difícil En red
C2	Normal En red → Normal Jerárquico → Difícil En red → Difícil Jerárquico
C3	Difícil Jerárquico → Difícil En red → Normal Jerárquico → Normal En red
C4	Difícil En red → Difícil Jerárquico → Normal En red → Normal Jerárquico

3.4 Criterios para la selección de participantes

La muestra del estudio estuvo compuesta por 41 participantes, seleccionados de manera voluntaria entre estudiantes universitarios de segundo año en adelante. Para garantizar la validez de los datos recolectados, se establecieron los siguientes criterios de inclusión:

- Visión normal o corregida (uso de lentes o lentes de contacto).
- Estudiantes universitarios de pregrado a partir del segundo año.

Los criterios de exclusión incluyeron:

- Alumnos de universitarios de postgrado.
- Alumnos de educación básica o media.
- Antecedentes neurológicos o trastornos que pudieran afectar el comportamiento visual o cognitivo

Todos los participantes fueron informados previamente sobre los objetivos del estudio y firmaron un consentimiento informado, conforme a los lineamientos éticos de la institución.

Además, antes de participar en el experimento, se solicitó a cada estudiante completar dos pruebas estandarizadas de comprensión lectora: Lectum e Interamericana. Los puntajes obtenidos en estas evaluaciones fueron utilizados para clasificar a los participantes en dos grupos: lectores hábiles y lectores menos hábiles. Esta clasificación sirvió como base para los análisis comparativos posteriores sobre el comportamiento visual durante la lectura.

Debido a problemas de calidad en las grabaciones del ET, se excluyeron 8 participantes, quedando finalmente 31 sujetos válidos para el análisis, de los cuales 14 correspondieron a lectores hábiles y 19 lectores menos hábiles.

3.5 Adquisición de datos mediante Eye-Tracking

Durante la lectura, el sistema registró en tiempo real las coordenadas de la mirada (ejes X e Y) y el tiempo relativo de cada fijación, con una frecuencia de muestreo de 42 Hz.

Aunque el fabricante del sistema reporta una frecuencia de muestreo de 42 Hz, se realizó una verificación utilizando los datos crudos exportados desde Mangold Vision Analyzer. Al analizar los registros sin agregación, se observó una frecuencia cercana a los 38 Hz, con un intervalo de muestreo de aproximadamente 26 ms. Esta estimación se obtuvo dividiendo el número total de muestras por la duración total de una lectura, resultando en una frecuencia aproximada de 38,1 Hz a partir de un total de 36.457 muestras para una lectura de 15 minutos y 57 segundos.

La diferencia entre la frecuencia reportada por el fabricante y la observada podría atribuirse a procesos internos del software o a condiciones específicas durante la captura de datos.

Por otro lado, los archivos procesados por Mangold mediante métodos de agregación de fijaciones (utilizando cluster espacial y ventanas temporales) presentaron frecuencias significativamente menores, en torno a los 3 Hz. Esto se debe a que los datos fueron consolidados en eventos más relevantes para el análisis cognitivo, reduciendo el número total de registros a 2.213 muestras en la misma lectura de 15 minutos y 57 segundos, con un intervalo de muestreo de 350 ms.

Se optó por utilizar los datos exportados desde Mangold Vision Analyzer aplicando un método de agregación de fijaciones, ya que, de esta manera, se obtiene una representación más estable y significativa del comportamiento visual, reduciendo la influencia de pequeñas oscilaciones o artefactos del ET. El uso de clústeres y ventanas temporales permitió condensar múltiples muestras en eventos de fijación más robustos, facilitando la comparación entre participantes y mejorando la interpretación de los resultados

en términos de comprensión lectora.

Tras la lectura de cada texto, los participantes respondieron tres preguntas de comprensión lectora.

La sesión completa duró aproximadamente 50 minutos por participante. Se realizó un proceso de calibración de nueve puntos antes de cada texto para asegurar la precisión del ET. Al finalizar la recolección de datos, se utilizó Mangold Vision Analyzer para exportar los registros en formato CSV.

Para facilitar la navegación por los textos sin interferir con el ET, se conectó un mouse externo a la tablet. Esto se debió a que el uso de las manos para hacer scroll directamente sobre la pantalla táctil bloqueaba el sensor infrarrojo del ET, impidiendo un registro preciso de la mirada. El uso del mouse permitió preservar la integridad de los datos sin comprometer la experiencia interactiva del participante, esto se puede observar en la Figura 3.1.



Figura 3.1: Participante leyendo un texto en la tablet EyeOn Elite.

3.6 Procesamiento de los datos

Para calcular los parámetros utilizados en el análisis de comprensión lectora, se desarrolló un script en Python que procesó los datos exportados desde Mangold Vision Analyzer. El procesamiento se realizó

sobre archivos previamente agrupados mediante un método de agregación de fijaciones, configurado con un tamaño de clúster espacial adaptado a la dificultad del texto: 100 píxeles para textos de dificultad normal (correspondiente al largo promedio de una palabra de cinco letras) y 120 píxeles para textos de dificultad alta (promedio de seis letras por palabra). Esta diferencia en la longitud promedio puede observarse en la Tabla 3-3.

La elección de estos valores se basó en la resolución de la pantalla de la tablet (1920×1080 píxeles). Se midió el largo promedio de una palabra de cinco letras, obteniendo aproximadamente 100 píxeles, y se ajustó proporcionalmente para los textos difíciles. Además, se empleó una ventana temporal de 4ms para la agregación de fijaciones en el dominio del tiempo.

Tabla 3-3: Promedio de caracteres por palabra en los textos utilizados.

Tipo de Texto	Promedio de caracteres por palabra	Tamaño cluster (píxeles)
Normal Jerárquico	4.970	100
Normal En red	4.938	100
Difícil Jerárquico	5.555	120
Difícil En red	5.813	120

3.7 Cálculo de métricas visuales

El script calculó métricas como la duración promedio de fijaciones, la cantidad de fijaciones por minuto, la longitud media de movimientos sacádicos (en píxeles), el número de regresiones y tiempo total de lectura. Estas variables se utilizaron para caracterizar y comparar el comportamiento lector de los participantes, en función de su nivel de desempeño en comprensión lectora.

Las métricas se calcularon de la siguiente forma:

- **Duración promedio de fijaciones:** Se calculó como el promedio de la duración de todas las fijaciones registradas durante la lectura de un texto.

Siendo D_i la duración en milisegundos de la i -ésima fijación y N el número total de fijaciones,

$$\text{Promedio de duración de fijaciones (ms)} = \frac{1}{N} \sum_{i=1}^N D_i$$

- **Tiempo total de lectura:** Se obtuvo sumando la duración de todas las fijaciones registradas durante la lectura de un texto.

Con T_{ini} el instante de inicio de la primera fijación y T_{fin} el instante de inicio de la última,

$$T_{\text{total (ms)}} = T_{\text{fin}} - T_{\text{ini}}, \quad \text{Tiempo total de lectura (min)} = \frac{T_{\text{total (ms)}}}{60\,000}.$$

- **Cantidad de fijaciones por minuto:** Se obtuvo dividiendo el número total de fijaciones por la duración total de la lectura (en minutos).

$$\text{Cantidad de fijaciones por minuto} = \frac{N}{\text{Tiempo total de lectura (min)}}.$$

- **Longitud de movimiento sacádicos:** Se estimó midiendo la distancia euclidiana en píxeles entre cada par de fijaciones consecutivas, considerando las coordenadas (X, Y) de la mirada.

Definiendo $\Delta x_i = x_i - x_{i-1}$ y $\Delta y_i = y_i - y_{i-1}$ como la diferencia de posición de la mirada en los ejes X e Y,

$$\text{Longitud media de sacadas} = \frac{1}{N-1} \sum_{i=2}^N \sqrt{(\Delta x_i)^2 + (\Delta y_i)^2}$$

- **Número de regresiones:** Se contabilizaron las sacadas con dirección hacia atrás (es decir, hacia una posición horizontal menor que la anterior) dentro de una misma línea.

Se considera regresión todo movimiento sacadico con $\Delta x_i < -50$ px y $|\Delta y_i| < 50$ px.

$$\text{Cantidad de regresiones} = \sum_{i=2}^N \mathbf{1}[(\Delta x_i < -50) \wedge (|\Delta y_i| < 50)].$$

3.8 Análisis estadístico

El análisis estadístico de los datos se realizó utilizando Python, con el objetivo de identificar diferencias significativas en las métricas visuales entre los distintos grupos de lectores (hábles vs. menos hábiles) y entre la dificultad del texto.

Pruebas de supuestos

Previo a la comparación de grupos, se aplicaron pruebas para verificar el cumplimiento de los supuestos requeridos por las pruebas paramétricas:

- **Normalidad:** Se utilizó la prueba de Shapiro-Wilk para evaluar si las distribuciones de cada métrica visual se ajustaban a una distribución normal.
- **Homogeneidad de varianzas:** Se aplicó la prueba de Levene para determinar si las varianzas entre grupos eran estadísticamente equivalentes.

Comparación entre grupos

Según el cumplimiento de los supuestos, se aplicaron las siguientes pruebas:

- **T-test de Student:** Para comparar medias entre grupos independientes cuando se cumplían normalidad y homogeneidad de varianzas.
- **Mann-Whitney U:** Prueba no paramétrica utilizada cuando no se cumplía el supuesto de normalidad.

Análisis multifactorial y modelos mixtos

Además de las comparaciones simples entre grupos, se utilizaron modelos que permiten evaluar la influencia simultánea de múltiples factores, así como controlar la variabilidad entre participantes.

- **ANOVA bifactorial:** Se aplicó un análisis de varianza con dos factores categóricos: tipo de lector (hábles o menos hábiles) y dificultad del texto (normal o difícil), incluyendo su interacción. La fórmula empleada fue:

$$\text{métrica} \sim C(\text{tipo_lector}) + C(\text{tipo_texto}) + C(\text{tipo_lector}):C(\text{tipo_texto})$$

Esto permitió analizar tanto los efectos principales como la interacción entre factores sobre cada métrica visual. La interacción es particularmente relevante, ya que indica si el efecto del tipo de lector cambia dependiendo de la dificultad del texto.

- **MLM:** Para abordar la estructura jerárquica de los datos —con múltiples observaciones por participante— se aplicaron MLM utilizando la biblioteca `statsmodels`. El modelo base incluyó como variable dependiente cada métrica visual, considerando el tipo de lector y dificultad del texto como efecto fijo y el código del participante como efecto aleatorio:

$$\text{métrica} \sim \text{tipo_lector} * \text{tipo_texto} + (1 \mid \text{código})$$

Finalmente, se interpretaron los coeficientes estimados y sus valores p para los efectos fijos, y se validaron los supuestos de los modelos (normalidad de los residuos y homogeneidad de varianzas) mediante inspección gráfica y tests complementarios cuando fue necesario.

3.9 Modelos de Machine Learning y Redes Neuronales

Con el objetivo de explorar un enfoque predictivo complementario al análisis estadístico, se implementaron modelos de ML y RN para clasificar a los participantes según su nivel lector, utilizando como entrada métricas extraídas del ET durante la lectura.

Características utilizadas

Los modelos se entrenaron utilizando las siguientes variables como características predictoras:

- Promedio de duración de fijaciones (ms)
- Desviación estándar de duración de fijaciones (ms)
- Tiempo total de lectura (min)
- Cantidad de fijaciones por minuto
- Longitud media de movimientos sacádicos (px)
- Desviación estándar de la longitud de sacadas (px)
- Duración media de sacadas (ms)
- Desviación estándar de la duración de sacadas (ms)
- Cantidad de regresiones

Cada instancia correspondió a la lectura de un texto por un participante, y las etiquetas (lector hábil o menos hábil) fueron asignadas en base a los puntajes obtenidos en las pruebas Lectum e Interamericana.

El conjunto de datos final estuvo conformado por 131 lecturas realizadas por 31 participantes, de los cuales 14 fueron clasificados como lectores hábiles y 19 como lectores menos hábiles. En total, se analizaron 55 lecturas correspondientes a lectores hábiles y 76 a lectores menos hábiles. Una lectura de un lector hábil fue descartada debido a problemas de calidad en la grabación del ET.

Modelos tradicionales basados en árboles

Para entrenar modelos de clasificación tradicionales, se utilizó un pipeline de procesamiento implementado en Python. Los modelos evaluados incluyeron Random Forest, Gradient Boosting y Support Vector Machines (SVM). La base de datos se dividió utilizando validación cruzada estratificada por grupo, a través del método `StratifiedGroupKFold`, asegurando que los datos correspondientes a un mismo participante (identificado por su código) no estuvieran simultáneamente en los conjuntos de entrenamiento y prueba.

La base de datos consistía en una fila por lectura de texto, con 9 métricas visuales como características y una etiqueta binaria asociada al tipo de lector. Para cada pliegue de validación, se realizó una optimización de hiperparámetros utilizando búsqueda bayesiana con la biblioteca `Optuna`. Se ejecutaron 100 *trials* por modelo, donde en cada uno se evaluaba una configuración distinta de hiperparámetros. La función maximizó el F1-score ponderado, utilizando un estimador definido dentro de un pipeline de `scikit-learn`.

Red neuronal multicapa

La MLP fue implementada en PyTorch. El dataset completo fue estructurado en tensores, donde cada instancia representaba las 9 métricas visuales asociadas a una lectura. La separación de datos se realizó siguiendo los mismos grupos utilizados para los modelos tradicionales: validación cruzada con `StratifiedGroupKFold`, asegurando que no hubiera fuga de información entre entrenamiento y prueba.

La arquitectura del modelo fue definida como una secuencia de capas lineales intercaladas con funciones de activación ReLU, capas de normalización por lotes (`BatchNorm1d`) y Dropout para evitar el sobreajuste. La función de pérdida utilizada fue `CrossEntropyLoss`, y el optimizador fue Adam. Se utilizaron 100 *trials* con Optuna para explorar distintas configuraciones de hiperparámetros, incluyendo la tasa de aprendizaje, el número de capas ocultas, el número de neuronas por capa y la tasa de

dropout. En cada trial se entrenaba el modelo y se evaluaba el F1 ponderado en validación, con el objetivo de seleccionar la mejor arquitectura.

El modelo fue entrenado con `early stopping`, evaluando el F1-score ponderado en un conjunto de validación interna en cada pliegue. Al finalizar, se promediaron los resultados de F1 ponderado sobre los distintos pliegues para obtener el rendimiento final del clasificador.

Estrategias para manejar el desbalance de clases

Ambos enfoques —modelos tradicionales y RN— enfrentaron el desafío de una distribución de clases desbalanceada (14 lectores hábiles y 19 lectores menos hábiles). Para mitigar este problema, se aplicaron dos estrategias:

- **SMOTE (Synthetic Minority Over-sampling Technique):** permitió generar ejemplos sintéticos de la clase minoritaria, mejorando el aprendizaje en contextos con pocos datos.
- **Sample weights:** se asignaron mayores pesos a las instancias de la clase minoritaria durante el entrenamiento, ajustando su influencia en la función de pérdida.

Estas técnicas fueron aplicadas en distintos experimentos tanto con los modelos de árboles como con la MLP, y contribuyeron a mejorar la sensibilidad y el equilibrio en la clasificación final.

3.10 Discusión

En este capítulo se detalló el diseño experimental, la metodología de recolección de datos y las métricas utilizadas para analizar el comportamiento lector de los participantes. El uso del ET permitió capturar las fijaciones durante la lectura digital de textos con diferente nivel de dificultad y estructura de navegación.

El uso de la agregación de fijaciones resultó ser útil para obtener datos más representativos del comportamiento lector real, evitando interpretar microfluctuaciones como eventos cognitivos significativos. Esta decisión metodológica implica, sin embargo, una pérdida de resolución temporal que podría afectar el análisis de eventos rápidos como sacadas breves entre palabras muy próximas. Por lo tanto, aunque se logró robustez, también se introdujo un filtro que podría invisibilizar diferencias sutiles entre lectores hábiles.

El diseño experimental incluyó textos jerárquicos y en red, con dificultad normal y difícil, distribuidos a través de cuatro condiciones de contrabalanceo. Esta estructura permitió controlar los efectos del orden de lectura y minimizar posibles sesgos por fatiga o aprendizaje. Además, se aseguraron condiciones estandarizadas de lectura mediante el uso de un entorno controlado, calibraciones previas y la utilización de un mouse externo para no interferir con el ET.

Se implementaron métricas visuales clave, como la duración promedio de fijaciones, la cantidad de fijaciones por minuto, la longitud de las sacadas, el número de regresiones y el tiempo total de lectura. Estas métricas fueron seleccionadas por su relevancia en estudios de comprensión lectora, ya que permiten establecer diferencias entre patrones de lectura más eficientes y aquellos que implican mayores dificultades de procesamiento.

El análisis estadístico posterior permitirá evaluar si estas métricas varían significativamente entre los grupos de lectores, así como si la dificultad del texto impactan el comportamiento visual. En conjunto, los datos recolectados y procesados en este capítulo proporcionan una base sólida para explorar las relaciones entre el comportamiento lector y la comprensión en entornos digitales interactivos.

El uso de modelos de ML y RN ofrece una perspectiva complementaria al análisis estadístico tradicional, permitiendo explorar patrones complejos en los datos. Estos modelos se entrenaron con las métricas visuales obtenidas, buscando clasificar a los participantes según su nivel lector.

Capítulo 4. Resultados

En este capítulo se presentan los resultados del estudio, acompañados de análisis estadísticos, gráficos y tablas que los ilustran de manera clara. Asimismo, se discuten sus implicancias en relación con la problemática abordada.

4.1 Visualización de distribuciones

A continuación, se presentan los diagramas de caja (boxplots) correspondientes a cada métrica visual, comparando la distribución de los lectores hábiles como de los menos hábiles por dificultad del texto. Estas visualizaciones permiten observar la distribución de los datos y evidenciar las diferencias entre los grupos evaluados.

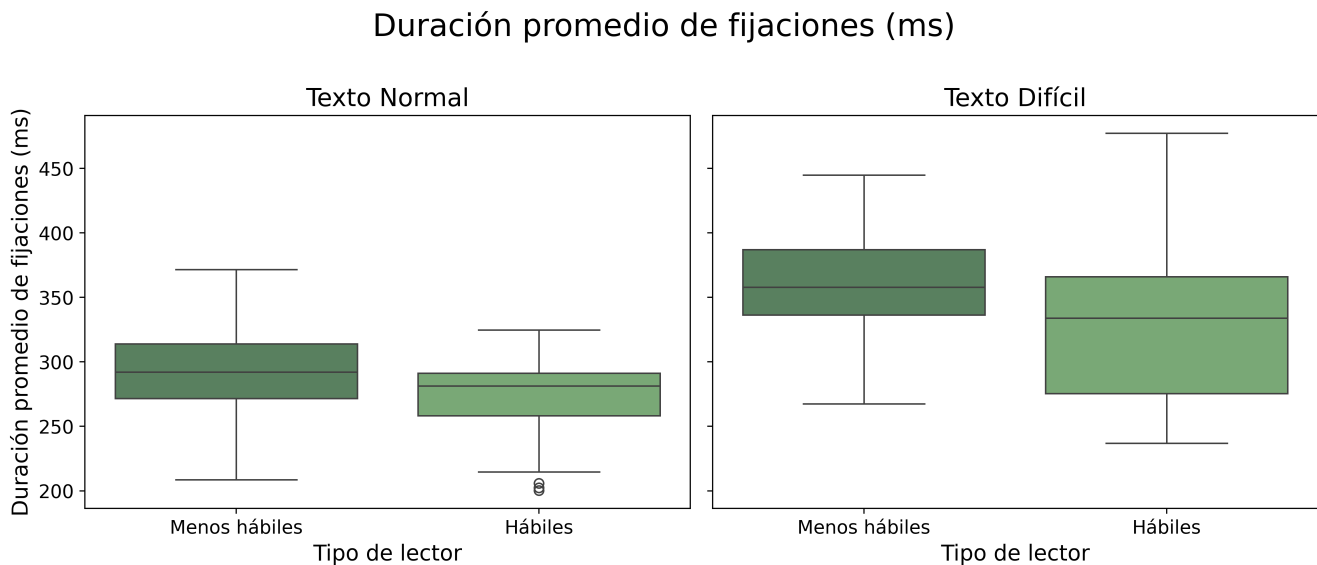


Figura 4.1: Duración promedio de fijaciones según tipo de lector y dificultad del texto.

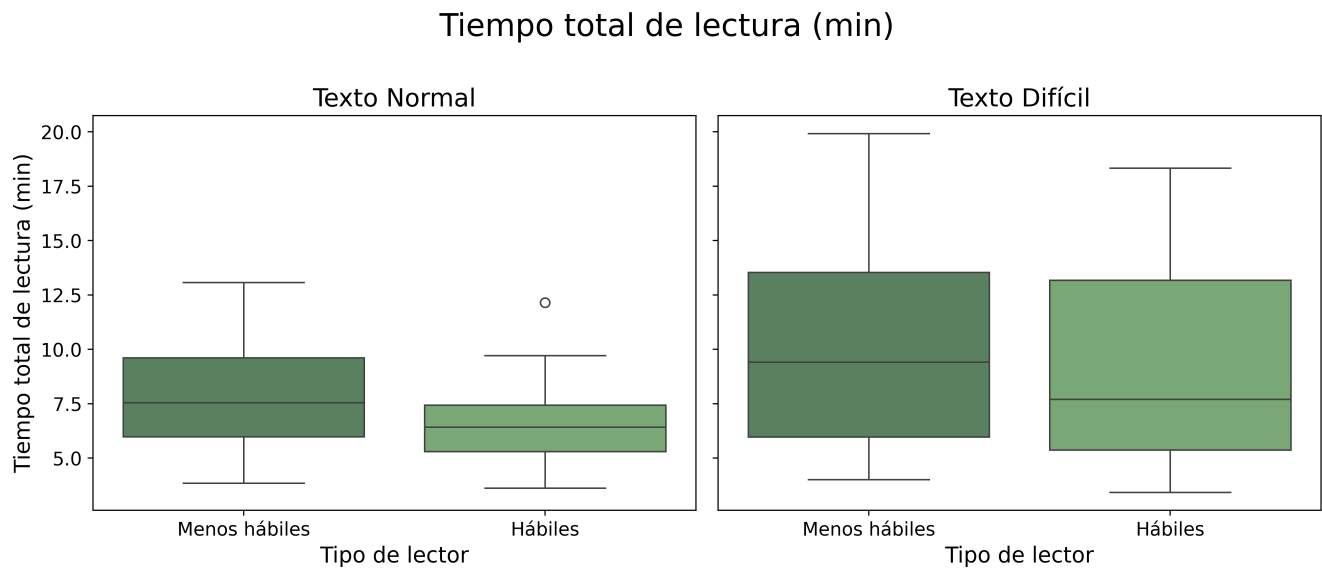


Figura 4.2: Tiempo total de lectura según tipo de lector y dificultad del texto.

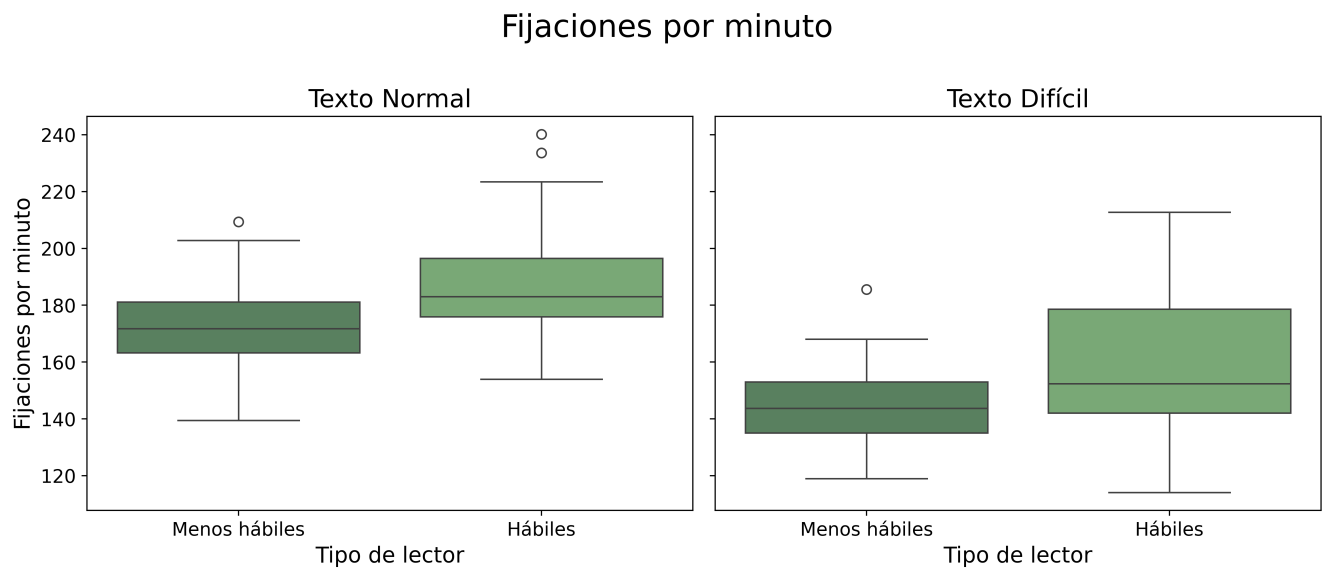


Figura 4.3: Cantidad de fijaciones por minuto según tipo de lector y dificultad del texto.

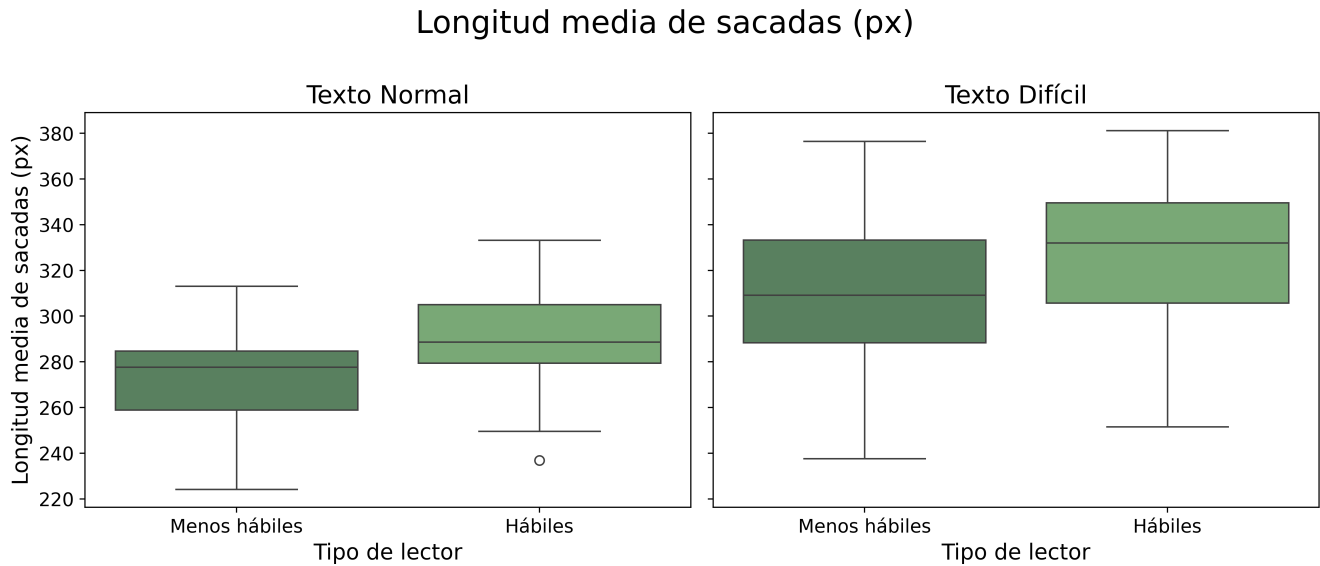


Figura 4.4: Longitud media de sacadas según tipo de lector y dificultad del texto.

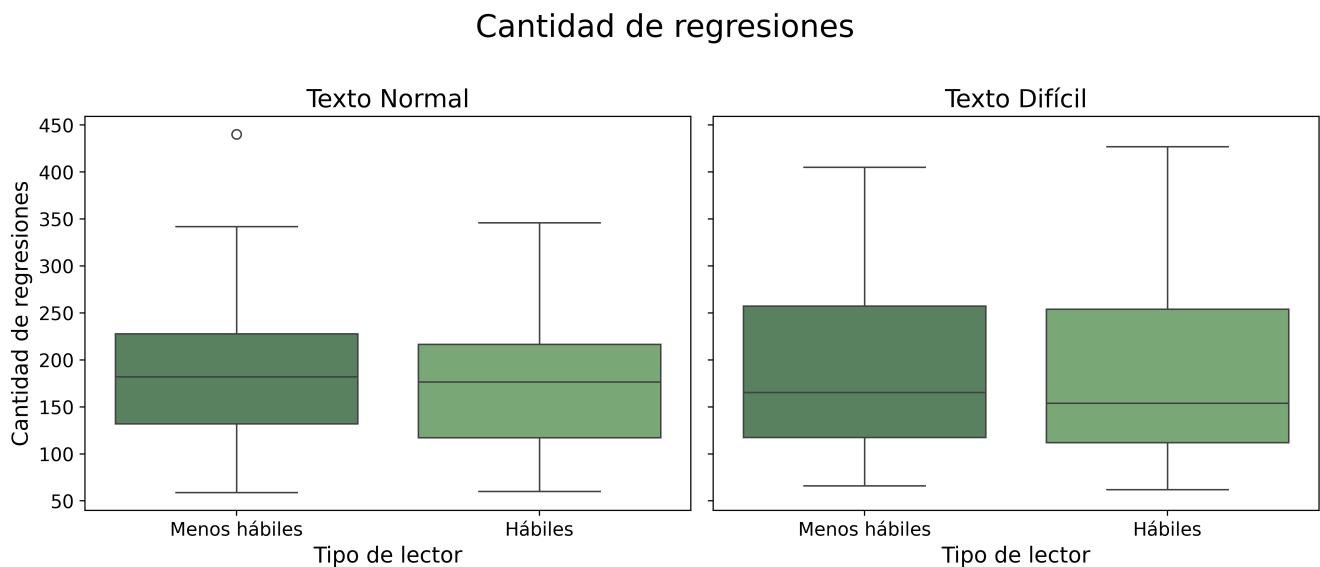


Figura 4.5: Cantidad de regresiones según tipo de lector y dificultad del texto.

4.2 Resultados Análisis Estadístico

A continuación, se presentan los resultados de las pruebas estadísticas aplicadas a las métricas visuales registradas durante la lectura. Se evaluaron diferencias entre lectores hábiles y lectores menos hábiles, así como efectos de la dificultad del texto (normal o difícil) y su interacción con el tipo de lector.

4.2.1 Comparaciones por grupo (t-test)

Para evaluar diferencias entre lectores hábiles y lectores menos hábiles, se aplicaron pruebas t para muestras independientes en aquellas métricas que cumplieron con los supuestos de normalidad (Shapiro-Wilk) y homogeneidad de varianzas (Levene). A continuación se presentan los valores p obtenidos en las pruebas t de Student para cada métrica:

Tabla 4-4: Valores p de las pruebas t de Student para cada métrica comparando buenos y malos lectores.

Métrica	p-valor
Duración promedio de fijaciones	0.002*
Tiempo total de lectura	0.053
Fijaciones por minuto	<0.001*
Longitud media de sacadas	0.001*
Cantidad de regresiones	0.347

* p <0.05 (significativo)

4.2.2 ANOVA bifactorial

Se realizaron análisis de varianza bifactoriales para cada métrica, considerando como factores el tipo de lector (hábles vs. menos hábiles) y la dificultad del texto (normal vs. difícil). A continuación, se resumen los resultados obtenidos:

Tabla 4-5: Valores p del ANOVA bifactorial por tipo de lector, dificultad del texto e interacción.

Métrica	Lector (p)	Texto (p)	Interacción (p)
Duración promedio de fijaciones	<0.001*	<0.001*	0.454
Tiempo total de lectura	0.046*	<0.001*	0.875
Fijaciones por minuto	<0.001*	<0.001*	0.922
Longitud media de sacadas	<0.001*	<0.001*	0.966
Cantidad de regresiones	0.352	0.730	0.888

* p <0.05 (significativo)

4.2.3 Modelos lineales mixtos

Se ajustaron MLM que incluyeron la interacción entre tipo de lector y dificultad del texto. Estos modelos permiten controlar la variabilidad entre participantes (efecto aleatorio) y evaluar simultáneamente los efectos principales y su interacción.

Tabla 4-6: Resultados de los MLM con interacción tipo lector × dificultad texto

Métrica	Tipo lector (p)	Dificultad texto (p)	Interacción (p)
Duración promedio de fijaciones	0.012*	<0.001*	0.198
Tiempo total de lectura	0.131	0.020*	0.874
Fijaciones por minuto	0.008*	<0.001*	0.854
Longitud media de sacadas	0.014*	<0.001*	0.980
Cantidad de regresiones	0.493	0.897	0.871

* p <0.05 (significativo)

4.2.4 Resumen general de resultados

Tabla 4-7: Resumen general de significancia estadística por métrica

Métrica	T-test	ANOVA		MLM	
		Lector	Texto	Lector	Texto
Duración promedio de fijaciones	Sí	Sí	Sí	Sí	Sí
Tiempo total de lectura	No	Sí	Sí	No	Sí
Fijaciones por minuto	Sí	Sí	Sí	Sí	Sí
Longitud media de sacadas	Sí	Sí	Sí	Sí	Sí
Cantidad de regresiones	No	No	No	No	No

4.3 Resultados de los Modelos de Machine Learning y Redes Neuronales

Además de los análisis estadísticos, se entrenaron modelos de clasificación supervisada para predecir el tipo de lector a partir de métricas visuales extraídas durante la lectura. Se utilizaron tanto modelos basa-

dos en árboles de decisión como RN, evaluando su rendimiento mediante validación cruzada estratificada por participante.

4.3.1 Balanceo de clases

Dado que el conjunto de datos estaba desequilibrado (55 lecturas lectores hábiles y 76 de lectores menos hábiles), se aplicaron dos estrategias para mitigar este problema:

- Aplicación de **SMOTE** para generar ejemplos sintéticos de la clase minoritaria.
- Utilización de **sample weights** durante el entrenamiento.

4.3.2 Rendimiento de los modelos

A continuación, se presentan los resultados de los modelos de ML y RN bajo tres estrategias de entrenamiento: uso de sample weights, aplicación de SMOTE y sin balanceo de clases. Cada tabla resume el mejor desempeño alcanzado por cada modelo en términos de F1-score y otras métricas relevantes.

Tabla 4-8: Resultados de modelos ML y RN con sample weights

Modelo	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.654	0.695	0.686	0.654
Decision Tree	0.637	0.631	0.631	0.645
Random Forest	0.704	0.727	0.689	0.698
Gradient Boosting	0.708	0.691	0.696	0.710
XGBoost	0.747	0.767	0.754	0.753
SVC (linear)	0.722	0.752	0.755	0.726
SGD Classifier	0.685	0.664	0.661	0.681
MLP (RN)	0.696	0.704	0.705	0.699

Tabla 4-9: Resultados de modelos ML y RN con SMOTE

Modelo	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.661	0.717	0.701	0.657
Decision Tree	0.634	0.630	0.624	0.636
Random Forest	0.680	0.674	0.674	0.680
Gradient Boosting	0.703	0.686	0.686	0.703
XGBoost	0.714	0.734	0.718	0.714
SVC (linear)	0.698	0.721	0.731	0.702
SGD Classifier	0.714	0.705	0.689	0.706
MLP (RN)	0.683	0.691	0.677	0.687

Tabla 4-10: Resultados de modelos ML y RN sin balanceo

Modelo	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.692	0.780	0.732	0.684
Decision Tree	0.710	0.704	0.691	0.707
Random Forest	0.719	0.745	0.713	0.714
Gradient Boosting	0.710	0.692	0.692	0.713
XGBoost	0.729	0.758	0.741	0.727
SVC (linear)	0.729	0.782	0.762	0.728
SGD Classifier	0.697	0.722	0.694	0.697
MLP (RN)	0.696	0.704	0.705	0.699

4.4 Gráficos de importancias de métricas

A continuación, se muestra el gráfico de importancia de características del modelo XGBoost, el cual permite identificar las métricas visuales que más influyen en la clasificación del tipo de lector.

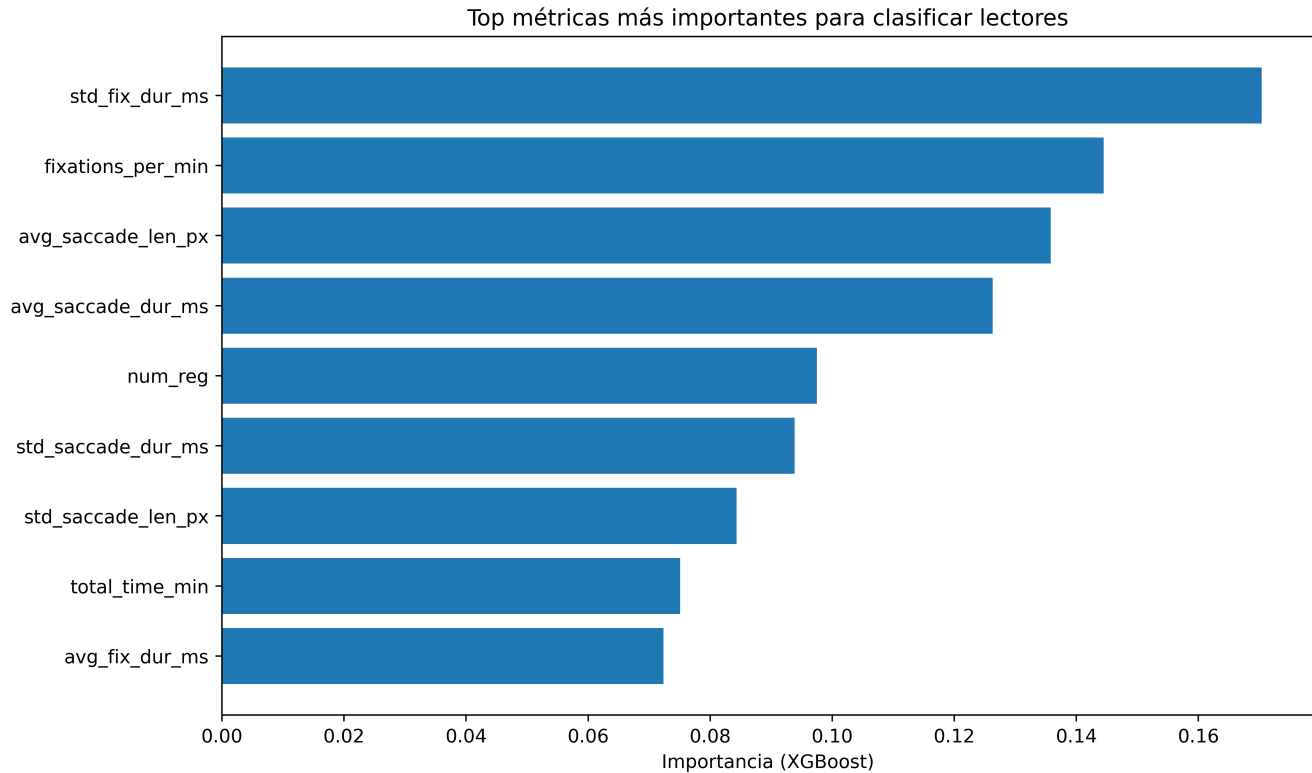


Figura 4.6: Importancia de características del modelo XGBoost.

En la Figura 4.6 se presentan los nombres abreviados utilizados para cada métrica, los cuales corresponden a las siguientes descripciones:

- avg_fix_dur_ms: Promedio de duración de fijaciones (ms)
- std_fix_dur_ms: Desviación estándar de duración de fijaciones (ms)
- total_time_min: Tiempo total de lectura (min)
- fixations_per_min: Cantidad de fijaciones por minuto
- avg_saccade_len_px: Longitud media de movimientos sacádicos (px)
- std_saccade_len_px: Desviación estándar de longitud de movimientos sacádicos (px)
- avg_saccade_dur_ms: Duración media de movimientos sacádicos (ms)
- std_saccade_dur_ms: Desviación estándar de duración de movimientos sacádicos (ms)
- num_reg: Cantidad de regresiones

A partir del análisis de importancia de características, se identificó que la métrica más influyente en la clasificación del tipo de lector fue la desviación estándar de la duración de fijaciones. Su relevancia supera notablemente al resto de las métricas evaluadas, por lo que se representó su distribución diferenciada por tipo de lector en la Figura 4.7.

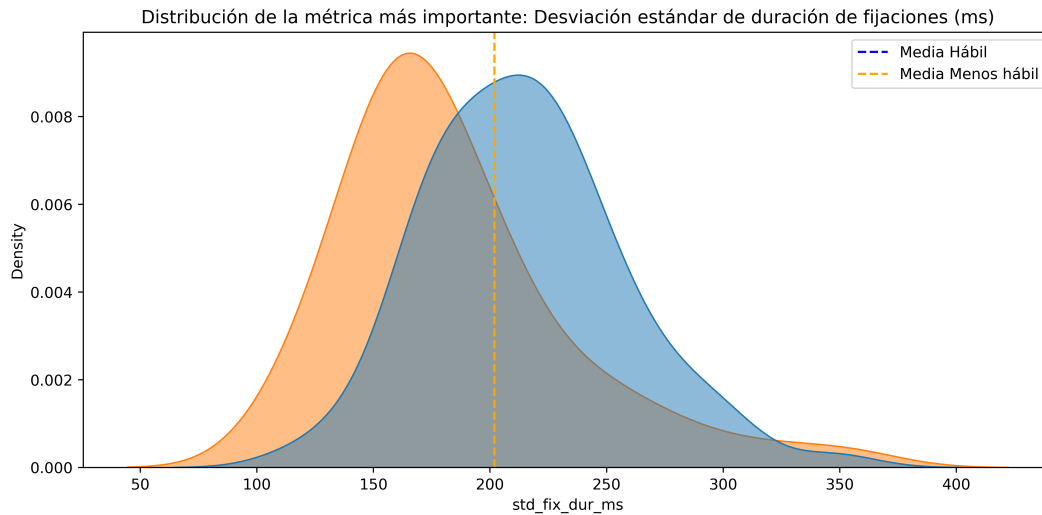


Figura 4.7: Distribución de la desviación estándar de la duración de fijaciones según tipo de lector.

Asimismo, se analizaron combinaciones entre las métricas más importantes. En la Figura 4.8 se presenta un gráfico de dispersión entre las dos métricas más relevantes: desviación estándar de la duración de fijaciones y cantidad de fijaciones por minuto.

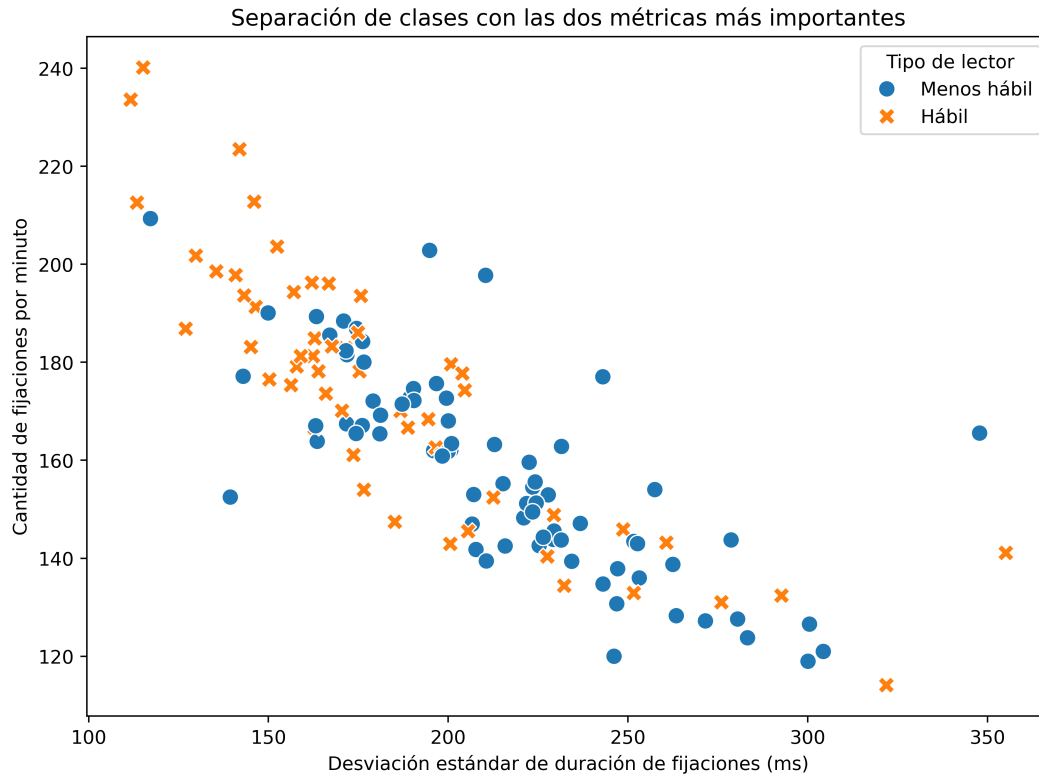


Figura 4.8: Relación entre desviación estándar de duración de fijaciones y cantidad de fijaciones por minuto.

Por otro lado, la Figura 4.9 muestra la relación entre la longitud media de movimientos sacádicos y la duración media de dichos movimientos, correspondientes a las métricas en tercera y cuarta posición en importancia.

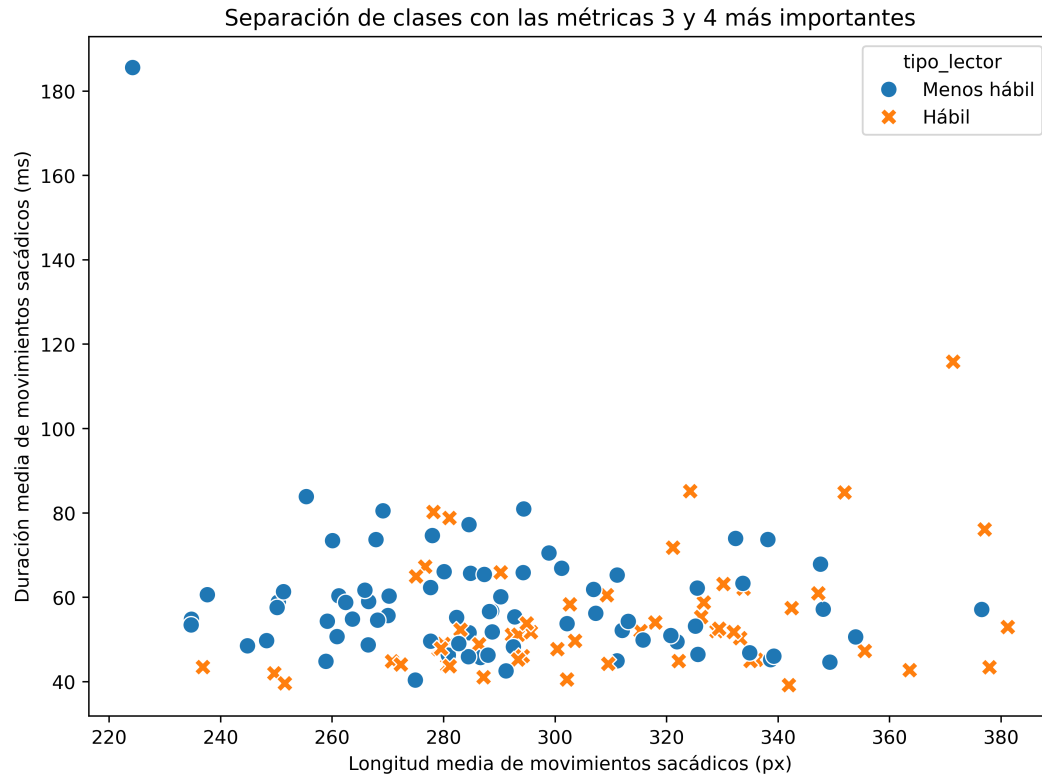


Figura 4.9: Relación entre longitud media y duración media de movimientos sacádicos.

Capítulo 5. Discusión y Conclusiones

En este capítulo se presentan las conclusiones generales del trabajo realizado, así como una discusión sobre los resultados obtenidos y las implicaciones de los mismos. Se reflexiona sobre la importancia de los hallazgos en el contexto del estudio y se sugieren posibles líneas de investigación futura.

5.1 Discusión de los Resultados del Análisis Estadístico

5.1.1 Análisis de las Métricas de Lectura

El análisis de las métricas de lectura, mediante pruebas *t* de Student y ANOVA bifactorial, reveló diferencias significativas entre los tipos de lectores. En promedio, los lectores hábiles presentaron una menor duración promedio de fijaciones (298,4 ms) en comparación con los lectores menos hábiles (328,5 ms; $p = 0.002$), una mayor longitud media de sacadas (310 px vs. 290 px; $p = 0.001$) y una mayor cantidad de fijaciones por minuto (174.3 vs. 157.8; $p < 0.001$). Estos valores promedio sugieren que los lectores hábiles muestran un estilo de lectura más fluido y eficiente, posiblemente asociado a una mejor decodificación y planificación visual.

Estos hallazgos fueron respaldados por los MLM, en los cuales el tipo de lector se mantuvo como un predictor significativo en varias métricas, incluso al controlar el efecto aleatorio del participante. En conjunto, estos resultados son coherentes con investigaciones previas que han documentado diferencias sistemáticas en los movimientos oculares entre lectores hábiles y menos hábiles en entornos digitales.

5.1.2 Comparación entre Tipos de Texto

Los resultados también evidenciaron que la dificultad del texto influye de manera significativa en las métricas de lectura. En promedio, los textos difíciles generaron una mayor duración de las fijaciones (348.6 ms frente a 283.4 ms; $p < 0.001$) y un menor número de fijaciones por minuto (150.9 vs. 178.4; $p < 0.001$), en comparación con los textos normales. Estos efectos podrían atribuirse a una mayor carga cognitiva, posiblemente relacionada con la estructura hipertextual de los textos difíciles o el uso de un vocabulario más especializado.

No obstante, tanto en el análisis ANOVA como en los MLM con interacción, no se encontró un efecto significativo de la interacción entre el tipo de lector y la dificultad del texto. Esto sugiere que, aunque ambos factores afectan el desempeño lector por separado, su combinación no genera un efecto adicional relevante. En otras palabras, lectores hábiles y menos hábiles parecen responder de forma similar frente a diferentes niveles de complejidad textual.

Por otro lado, la cantidad de regresiones no presentó diferencias significativas en función del tipo de lector ni de la dificultad del texto. Esto podría deberse a falta de datos suficientes para detectar diferencias sutiles en este aspecto, o que la desigualdad de los grupos de lectores no permitió una evaluación adecuada de este parámetro.

5.2 Discusión de los Resultados de Modelos de Machine Learning y Redes Neuronales

5.2.1 Comparación de Desempeño de Modelos

Se entrenaron distintos modelos de ML y RN para predecir el tipo de lector a partir de las métricas extraídas. Se consideraron tres configuraciones: uso de pesos por sample weights, uso de SMOTE, y un enfoque sin ninguna técnica de balanceo.

En general, los modelos con mejor desempeño fueron los basados en XGBoost y SVC, obteniendo un F1-score cercano a 0.75 en las mejores configuraciones. En el caso de los modelos con sample weights, el clasificador XGBoost logró un F1-score de 0.753, mientras que con SMOTE alcanzó 0.714 y sin técnicas de balanceo obtuvo 0.727. Por su parte, SVC logró un F1-score de 0.728 sin balanceo y 0.726 con sample weights.

La RN MLP logró un desempeño competitivo con un F1-score de 0.699 utilizando sample weights, aunque no superó consistentemente a los modelos de ML tradicionales. Esto podría explicarse por la cantidad limitada de datos o la mayor sensibilidad de las RN a la configuración de hiperparámetros.

5.2.2 Interpretación General

Los resultados sugieren que las métricas extraídas a partir del ET permiten una clasificación razonablemente precisa del tipo de lector. La precisión obtenida por los modelos mejor entrenados supera el

70 %, lo que indica un potencial para el uso de estas técnicas en aplicaciones educativas o diagnósticas. El uso de técnicas de balanceo como SMOTE o sample weights ayudó a mejorar el rendimiento en algunos modelos, aunque no de forma uniforme.

En general, el modelo de XGBoost con sample weights fue el que presentó el mejor compromiso entre precisión, recall y F1-score. Esto refuerza la utilidad de enfoques basados en árboles de decisión para problemas con datos tabulares derivados de mediciones fisiológicas como el ET.

En relación con la importancia de las métricas utilizadas en la clasificación, la desviación estándar de la duración de fijaciones y la cantidad de fijaciones por minuto fueron las variables más influyentes según el modelo XGBoost. Tal como se muestra en la Figura 4.7, la métrica más importante —la desviación estándar de la duración de fijaciones— permite una clara separación entre lectores hábiles y menos hábiles. La menor variabilidad en la duración de fijaciones observada en los lectores hábiles sugiere un estilo de lectura más eficiente y focalizado, posiblemente asociado a una mejor comprensión del texto.

En el gráfico de dispersión de la Figura 4.8, se aprecia que los lectores hábiles tienden a presentar una mayor cantidad de fijaciones por minuto y una menor variabilidad en su duración, lo que refuerza la hipótesis de una mayor eficiencia en el procesamiento visual del contenido textual.

Por otro lado, la Figura 4.9 muestra que los lectores hábiles tienden a tener una mayor longitud media de sacadas en comparación con los menos hábiles, aunque ambos grupos mantienen una duración media de sacadas similar. Esto sugiere que los lectores hábiles realizan movimientos oculares más largos en el mismo tiempo, lo que podría estar vinculado a una mejor planificación visual y a una mayor capacidad para anticipar el contenido del texto.

5.3 Conclusiones

Los resultados obtenidos en esta Memoria proporcionan información sobre cómo los patrones de lectura digital pueden diferir entre lectores hábiles y menos hábiles. Las métricas de lectura, como la duración de fijaciones y la longitud de sacadas, se relacionan con la capacidad lectora, lo que sugiere que el ET puede ser una herramienta efectiva para evaluar habilidades lectoras en entornos digitales. Además, los modelos de ML y RN entrenados con estas métricas lograron clasificar a los participantes con un grado razonable de precisión, lo que abre la puerta a futuras aplicaciones en el ámbito educativo y clínico.

5.4 Líneas Futuras de Investigación

A partir de los hallazgos de esta Memoria, se sugieren varias líneas futuras de investigación:

- Ampliar el conjunto de datos con más participantes y así lograr resultados más robustos y generalizables.
- Explorar el uso de diferentes arquitecturas de RN y técnicas de regularización para mejorar el desempeño de los modelos.
- Desarrollar un sistema de captura que registre la palabra específica que el lector está visualizando en cada momento, con el objetivo de extraer métricas de lectura más detalladas y contextualizadas. Para ello, se podría emplear un ET EyeLink Portable Duo, que permite alcanzar frecuencias de muestreo de hasta 2000 Hz.
- Profundizar en el análisis de las métricas de lectura para identificar patrones más sutiles que puedan diferenciar aún más a los lectores hábiles de los menos hábiles.
- Realizar estudios longitudinales para analizar cómo evolucionan los patrones de lectura a lo largo del tiempo tras un entrenamiento específico.
- Comparar el rendimiento del sistema en diferentes dispositivos (por ejemplo, pantallas pequeñas vs. grandes, tablets vs. notebooks), para explorar la generalización del modelo en distintos contextos de lectura digital.
- Comparación entre el tipo de estructura de texto (jerárquica vs. en red) y su impacto en la comprensión lectora, utilizando métricas de ET para evaluar cómo estas estructuras afectan el comportamiento lector.

Glosario

ANOVA Análisis de Varianza

AOIs Areas of Interest (Áreas de Interés)

ECG Electrocardiograma

EEG Electroencefalograma

ET Eye-Tracking (Seguimiento Ocular)

ML Machine Learning (Aprendizaje Automático)

MLM Modelos Lineales Mixtos

MLP Multilayer Perceptron (Red Neuronal Multicapa)

POR Punto de Regard

RN Redes Neurales

VR Realidad Virtual

Referencias

- [1] B. T. Carter and S. G. Luke, “Best practices in eye tracking research,” *International Journal of Psychophysiology*, vol. 155, pp. 49–62, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167876020301458>
- [2] A. T. Duchowski, *Eye tracking methodology: Theory and practice*, 3rd ed. Springer Science & Business Media, 2017. [Online]. Available: <https://link.springer.com/book/10.1007/978-1-84628-609-4>
- [3] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press, 2015. [Online]. Available: <https://global.oup.com/academic/product/eye-tracking-9780198738596>
- [4] K. Rayner, “Eye movements and attention in reading, scene perception, and visual search,” *The Quarterly Journal of Experimental Psychology*, vol. 62, no. 8, pp. 1457–1506, 2009. [Online]. Available: <https://doi.org/10.1080/17470210902816461>
- [5] E. Estévez, “Análisis comparativo de sistemas de tracking Óptico para el análisis de patrones de lectura de la web por sus usuarios,” *Universidad Politécnica de Madrid*, 2022. [Online]. Available: https://oa.upm.es/71428/1/TFG_EKATERINA_ESTEVEZ_IVANOVA.pdf
- [6] G. Gredebäck, S. Johnson, and C. von Hofsten, “Eye tracking in infancy research,” *Developmental Neuropsychology*, vol. 35, no. 1, pp. 1–19, 2009. [Online]. Available: <https://doi.org/10.1080/87565640903325758>
- [7] E. Álvarez García, “Lo que esconden tus ojos: la metodología eye-tracking aplicada al estudio del lenguaje,” *Estudios de Lingüística del Español*, vol. 45, p. pp. 205–239, jun. 2022. [Online]. Available: <https://bop.unibe.ch/elies/article/view/8856>
- [8] Z. Sharafi, B. Sharif, and Y. G. Guéhéneuc, “A practical guide on conducting eye tracking studies in software engineering,” *Empir Software Eng*, 2020.
- [9] M. JE and H. MC, “Uncovering the mental effort of reading using eye-tracking technology,” *Frontiers for Young Minds*, 2024.
- [10] R. Krueger, S. Koch, and T. Ertl, “Saccadelenses: Interactive exploratory filtering of eye tracking trajectories,” 10 2016, pp. 31–34.

- [11] A. F. Klaib, N. O. Alsrehin, W. Y. Melhem, H. O. Bashtawi, and A. A. Magableh, “Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and internet of things technologies,” *Expert Systems with Applications*, vol. 166, p. 114037, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417420308071>
- [12] F. Ungureanu, R. G. Lupu, A. Cadar, and A. Prodan, “Neuromarketing and visual attention study using eye tracking techniques,” in *2017 21st International Conference on System Theory, Control and Computing (ICSTCC)*, 2017, pp. 553–557.
- [13] S. Özdel, E. Bozkir, and E. Kasneci, “Privacy-preserving scanpath comparison for pervasive eye tracking,” *Proc. ACM Hum.-Comput. Interact.*, vol. 8, no. ETRA, May 2024. [Online]. Available: <https://doi.org/10.1145/3655605>
- [14] D. C. Mézière, L. Yu, E. D. Reichle, T. von der Malsburg, and G. McArthur, “Using eye-tracking measures to predict reading comprehension,” *Reading Research Quarterly*, vol. 58, no. 3, pp. 425–449, 2023. [Online]. Available: <https://ila.onlinelibrary.wiley.com/doi/abs/10.1002/rrq.498>
- [15] J. Ulloa, S. Rojas, P. Hernández, and C. Vargas, “Review of eye tracking metrics involved in emotional and cognitive processes,” *Frontiers in Psychology*, vol. 14, p. 1172635, 2023.
- [16] O. González B.C., Riffo, “Lectura digital: Navegación y comprensión lectora en dos estructuras de hipertexto,” *Paideia*, 2018.
- [17] A. Poole and L. J. Ball, “Eye tracking in human-computer interaction and usability research : Current status and future prospects,” 2004. [Online]. Available: <https://api.semanticscholar.org/CorpusID:972615>
- [18] K. González and R. Riffo, “Lectura digital: navegación y comprensión lectora en dos estructuras de hipertexto,” *Estudios pedagógicos (Valdivia)*, vol. 44, no. 3, pp. 113–126, 2018.
- [19] L. Salmerón, J. J. Cañas, W. Kintsch, and I. Fajardo, “Reading strategies and hypertext comprehension,” *Discourse Processes*, vol. 40, no. 3, pp. 171–191, 2005.
- [20] J. Naumann and L. Salmerón, “Does navigation always predict performance? effects of navigation on digital reading are moderated by comprehension skills,” *Reading Research Quarterly*, vol. 51, no. 4, pp. 399–416, 2016.
- [21] S. R. Ltd., *EyeLink Portable Duo User Manual*, 2023, portable Duo Eye-Tracking System Manual. [Online]. Available: <https://www.sr-research.com>

- [22] M. N. Mikhailenko M and K. M, “Eye-tracking in immersive virtual reality for education: A review of the current progress and applications,” *Frontiers in Education*, vol. 7, 2022.
- [23] B. J. Hou, Y. Abdrabou, F. Weidner, and H. Gellersen, “Unveiling variations: A comparative study of vr headsets regarding eye tracking volume, gaze accuracy, and precision,” in *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2024, pp. 650–655.
- [24] E. D. Systems, “Eyeon elite,” <https://eyetechds.com/eyeon-elite/>, 2025, accedido: 2025-05-19.
- [25] C. Andrà and et al., “The role of saccadic movements in arithmetic problem-solving: Insights from eye-tracking studies,” *South African Journal of Childhood Education*, vol. 4, no. 2, pp. 122–131, 2009. [Online]. Available: <http://www.sajce.co.za>
- [26] G. M. M. N. Tahri Sqalli M, Aslonov B and S. H. Y, “Eye tracking technology in medical practice: a perspective on its diverse applications,” *Frontiers in Medicine*, 2023.
- [27] J. Thevenot, M. B. López, and A. Hadid, “A survey on computer vision for assistive medical diagnosis from faces,” *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1497–1511, 2018.
- [28] M. O. Al-Moteri, M. Symmons, V. Plummer, and S. Cooper, “Eye tracking to investigate cue processing in medical decision-making: A scoping review,” *Computers in Human Behavior*, vol. 66, pp. 52–66, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0747563216306495>
- [29] V.-W. J. Rosch, J.L., “A review of eye-tracking applications as tools for training,” *Cogn Tech Work*, 2013.
- [30] M.-K. O. A. e. a. Strohmaier, A.R., “Eye-tracking methodology in mathematics education research: A systematic literature review,” *Educ Stud Math*, 2020.
- [31] J. Novák, J. Masner, P. Benda, P. Šimek, and V. Merunka, “Eye tracking, usability, and user experience: A systematic review,” *International Journal of Human–Computer Interaction*, vol. 40, no. 17, pp. 4484–4500, 2024. [Online]. Available: <https://doi.org/10.1080/10447318.2023.2221600>
- [32] L. Huang, J. Ouyang, and J. Jiang, “The relationship of word processing with l2 reading comprehension and working memory: Insights from eye-tracking,” *Learning and Individual Differences*, vol. 95, p. 102143, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1041608022000309>

- [33] M. Palmović, L. A. Jäger, and N. Hollenstein, “Editorial: Eye-tracking while reading for psycholinguistic and computational models of language comprehension,” *Frontiers in Psychology*, vol. 14, 2023. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1326408>
- [34] Y. Zhang, Q. Li, S. Nahata, T. Jamal, S.-K. Cheng, G. Cauwenberghs, and T.-P. Jung, “Integrating large language model, eeg, and eye-tracking for word-level neural state classification in reading comprehension,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.
- [35] M. I. GmbH, *Mangold Vision Manuals*, 2023, accessed: 2025-05-19. [Online]. Available: <https://www.videoanalysis.net/resources/mangoldvision/manuals/en/index.html?3212.html>
- [36] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 4th ed. SAGE Publications, 2013.
- [37] G. Cumming, *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. Routledge, 2012.
- [38] B. Winter, *Linear Models and Linear Mixed Effects Models in R with Linguistic Applications*, 2013, accessed July 2025.
- [39] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. O’Reilly Media, 2019.
- [40] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [41] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A next-generation hyperparameter optimization framework,” *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2623–2631, 2019.
- [42] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

UNIVERSIDAD DE CONCEPCIÓN – FACULTAD DE INGENIERÍA
RESUMEN DE MEMORIA DE TÍTULO

Departamento	: Departamento de Ingeniería Civil Eléctrica
Carrera	: Ingeniería Civil Biomédica
Nombre del memorista	: Ignacio Andrés Fritis Zacconi
Título de la memoria	: Análisis de Datos de Eye-Tracker durante textos digitales
Fecha de la presentación oral	: 28 de Agosto de 2025
Profesor(es) Guía	: Esteban Pino, Mabel Urrutia
Profesor(es) Revisor(es)	: Pamela Guevara
Concepto	:
Calificación	:

Resumen

Esta memoria estudia los patrones de lectura digital mediante tecnología de seguimiento ocular (Eye-Tracking). Se evaluó el comportamiento visual de 41 participantes, clasificados por su nivel de comprensión lectora, mientras leían textos con diferentes estructuras y niveles de dificultad.

Se analizaron métricas como duración de fijaciones, longitud de sacadas, fijaciones por minuto, regresiones y tiempo total de lectura. Se aplicaron pruebas estadísticas que mostraron diferencias significativas entre tipos de lectores, especialmente en fijaciones más cortas y sacadas más largas en lectores hábiles.

Además, se entrenaron modelos de Machine Learning y Redes Neuronales (incluyendo XGBoost y MLP) para clasificar a los participantes según su desempeño lector, logrando un F1-score de hasta 0.753. Los resultados confirman la utilidad del Eye-Tracking para evaluar habilidades lectoras en entornos digitales.
