

## Article

# A Proposed Method of Automating Data Processing for Analysing Data Produced from Eye Tracking and Galvanic Skin Response

Javier Sáez-García <sup>1,\*</sup>, María Consuelo Sáiz-Manzanares <sup>2,\*</sup>  and Raúl Marticorena-Sánchez <sup>3</sup> 

<sup>1</sup> Higher Polytechnic School (Campus Milanera), Universidad de Burgos, 09001 Burgos, Spain

<sup>2</sup> Department of Health Sciences, GIR DATAHES, UIC JCYL No. 348, Universidad de Burgos, 09001 Burgos, Spain

<sup>3</sup> ADMIRABLE Research Group, Department of Computer Engineering, Higher Polytechnic School (Campus Vena), Universidad de Burgos, 09006 Burgos, Spain; rmartico@ubu.es

\* Correspondence: jsg1013@alu.ubu.es (J.S.-G.); mcsmanzanares@ubu.es (M.C.S.-M.)

**Abstract:** The use of eye tracking technology, together with other physiological measurements such as psychogalvanic skin response (GSR) and electroencephalographic (EEG) recordings, provides researchers with information about users' physiological behavioural responses during their learning process in different types of tasks. These devices produce a large volume of data. However, in order to analyse these records, researchers have to process and analyse them using complex statistical and/or machine learning techniques (supervised or unsupervised) that are usually not incorporated into the devices. The objectives of this study were (1) to propose a procedure for processing the extracted data; (2) to address the potential technical challenges and difficulties in processing logs in integrated multichannel technology; and (3) to offer solutions for automating data processing and analysis. A Notebook in Jupyter is proposed with the steps for importing and processing data, as well as for using supervised and unsupervised machine learning algorithms.

**Keywords:** eye tracking; galvanic skin response; behavioural monitoring; learning process; data processing



**Citation:** Sáez-García, J.; Sáiz-Manzanares, M.C.; Marticorena-Sánchez, R. A Proposed Method of Automating Data Processing for Analysing Data Produced from Eye Tracking and Galvanic Skin Response. *Computers* **2024**, *13*, 289. <https://doi.org/10.3390/computers13110289>

Academic Editor: Ananda Maiti

Received: 19 September 2024

Revised: 29 October 2024

Accepted: 5 November 2024

Published: 8 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The use of eye tracking technology, together with other physiological biomarker measurements such as psychogalvanic skin response (GSR) and electroencephalographic recordings (EEG), provides researchers with information about users' physiological behavioural responses during their learning processes when performing various types of tasks. These records are obtained through sensors and provide non-subjective information that is put together with other data from the user (learning outcomes, perceived satisfaction with the learning process, perception of the use of metacognitive strategies, etc.). This technology has been applied in a range of fields, each with different objectives. For example, in marketing it has been used to understand client behaviour in order to achieve customer loyalty and increase sales [1–11]. In healthcare sciences, this technology can be applied for diagnosis (e.g., mental illness [12], autism spectrum disorder [13–15], examining medical imaging [16], mental fatigue [17], application in various clinical needs [18], Alzheimer's [19], dysfunctional cerebral networks [20], various disorders [21], Rett's Syndrome [22], and activity with prosthetic limbs [23]). Eye tracking technology has also been used in therapeutic intervention (for example, in Autism Spectrum Disorder [15], cerebral palsy [24], in patients with depression [25,26], in Parkinson's [27], with various addictions [28], and for anorexia nervosa [29]). It has also been used in education in order to understand how learners process information when doing various tasks, with the ultimate aim being to improve students' learning processes. Eye tracking technology can be used in educational settings to

analyse different learning styles [30], for educational rehabilitation from problems related to dyslexia [31], and to examine the effectiveness of educational methodologies such as evidenced-based learning [32].

More specifically in educational settings, this technology is commonly used in higher education, especially with health sciences students [33–39]. One of the specific functions the technology offers is that it can give the teacher or researcher data that, once processed and analysed, will help in the design of explanatory models of how learning occurs in different types of learners [30].

Eye tracking technology is also being used in driving to understand the characteristics of different drivers in pursuit of better driving and fewer accidents [40,41]. It is also being used to test different effects on driving of things such as advertising [42].

Recent studies on human perception indicate that higher levels of attention are related to variables such as gender, age, and educational level. Studies on vigilance indicate that higher levels of self-attention [35] are related to shorter total fixation durations and fewer fixation counts while the user looks at picture stimuli [40]. Also, the use of animated images seems to improve learning outcomes [41]. Similarly, recent studies call for increased research into the effects of self-regulated learning strategy (SRL) training on student motivation, engagement and performance [43]. Specifically, using think-aloud protocols involving self-regulated strategies appears to improve learning outcomes [44]. Similarly, feedback on user performance appears to enhance deep learning in learners [45]. Along these lines, using materials that include self-regulated simulation activities has been shown to be very effective in increasing information acquisition and mitigating the effects of prior knowledge about the task on learning outcomes [46].

In other studies we have addressed the meanings of metrics from applying eye tracking technology integrated with GSR devices and their relationship to the use of cognitive and metacognitive strategies as well as learning profiles [37,47,48]. One reference that may help researchers to prepare metrics is Holmqvist et al. [49], a summary of which can be found in Table 1.

**Table 1.** Some representative metrics and interpretation in integrated multichannel eye tracking.

Metric	Unit of Measurement	Meaning	Interpretation in the Context of the Human Learning
Fixation count	Count	Number of fixations on a stimulus or part of a stimulus.	A higher number of fixations may be related to a difficulty in processing that information because it is novel or difficult for the learner to process.
Fixation duration	Milliseconds	Duration of fixation.	Refers to the reaction times of the learner. A longer duration may be related to a higher cognitive load in the processing of the stimulus. It may also be related to the use of metacognitive orientation strategies, i.e., searching for information or relating it to previous knowledge.
Saccade count	Count	Refers to the shift of gaze from one part of the stimulus to another.	A higher number of saccades implies the use of metacognitive orientation strategies, i.e., searching for information or relating it to prior knowledge. Likewise, the greater the amplitude of the saccade, the lower the cognitive effort, although this may also be related to information processing problems. Younger learners apply shorter saccades.
Pupil diameter	Millimetres	The mean pupil diameter is collected for all fixations within an AOI during a time interval.	It refers to the interest that a stimulus or part of it has for the learner. A larger pupil diameter may be related to increased cognitive load and/or difficulty in processing a task.
Number of visits		Number of bindings within a stimulus or part of a stimulus (of an area of interest or AOI).	Refers to attention to a stimulus or part of a stimulus.
Scan Path Length or Gaze Point	X and Y position coordinates	Refers to the chain of fixations in order of succession.	It involves a pattern of visual tracking on a stimulus. It gives information about how each learner processes information.

Table 1. Cont.

Metric	Unit of Measurement	Meaning	Interpretation in the Context of the Human Learning
SCR count	Count	The number of skin conductance responses (SCRs), for each Interval in Time of Interest.	It provides information about the emotional state of a learner. The SCR count can be used to identify which specific moments of a dynamic stimulus and specific information within a stimulus elicit an emotional response in a learner. A higher SCR count indicates a higher level of emotional arousal.
ER SCR Amplitude	Microsiemens	The amplitude of each event related skin conductance response (ER-SCR), for each Interval in Time of Interest. Time of Interest intervals that do not have an ER-SCR are calculated using filtered GSR data.	When an SCR occurs between 1 and 5 s after an event (ER-SCR), it is considered whether the stimulus elicits an emotional reaction in the learner. This measure can provide information about different emotional states such as: anxiety, stress, frustration, or relaxation.
GSR average	Microsiemens	The mean of the average galvanic skin response (GSR) signal, after filtering for each time of interest.	When environmental conditions are held constant, slow fluctuations in the GSR signal (between seconds and minutes) reflect changes in the participant’s emotional arousal. The researcher can use the average GSR metric in different sections of the session to determine whether a learner might be stressed, frustrated, or relaxed during the course of an experiment.

*How to Extract Records from Eye Tracking Devices and How to Process and Analyse Records*

The devices that apply eye tracking technology include records of static metrics (fixations, saccades, pupil diameter, blinking, etc.). They also apply various measuring parameters to those metrics such as velocity, duration, etc. These metrics are found by applying statistical resources such as frequencies, means, standard deviations, etc. (examples of these measures according to parameter may be found in Table 1).

Eye tracking devices also include measurement of dynamic metrics (these involve the points where the participant is looking on the screen’s cartesian co-ordinates), called gaze point or scan path. An example is shown in Figure 1.

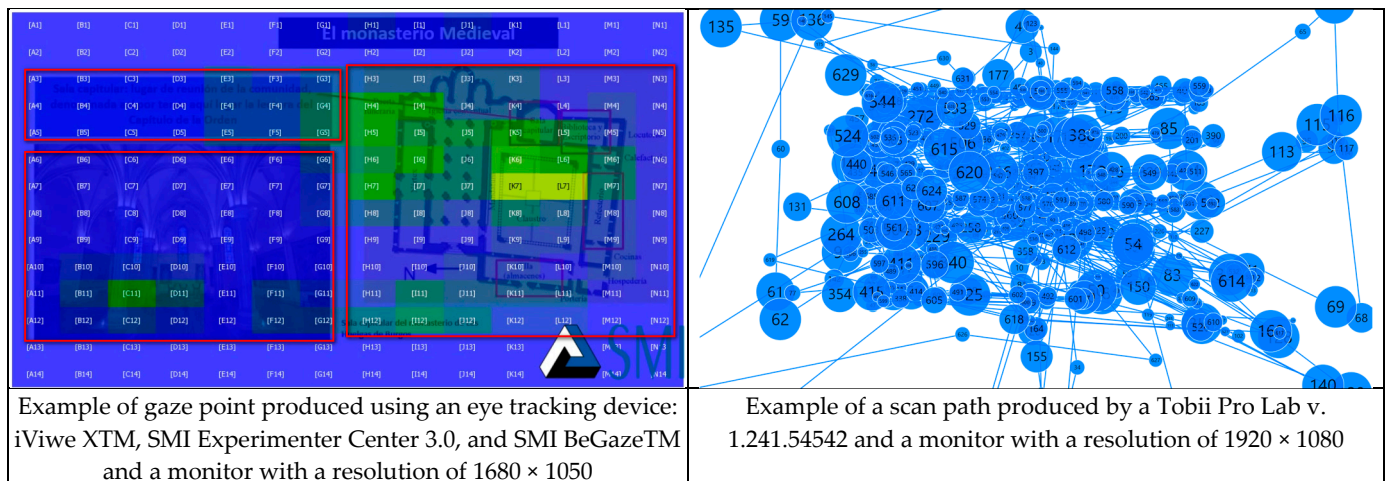


Figure 1. Examples of gaze point and scan path.

Nonetheless, more thorough analysis of dynamic metrics requires the application of machine learning techniques. These may be supervised techniques, which include techniques such as regression (about predicting numerical values), including linear regression, Support Vector Machine (SVM), Decision Tree, and Neural Networks. The other supervised techniques are classification (about predicting categories). These include SVM, discriminant analysis, nearest neighbour, and naïve Bayes. Machine learning techniques may also be unsupervised (which are applied without using labels in order to discover hidden patterns).

Unsupervised techniques include clustering algorithms (which help to group data based on similarity, such as: *k*-means, *k*-means++, etc.), and are applied to reduce dimensionality (such as in Principal Component Analysis—PCA) [50,51].

That said, it is important to remember that the data that can be extracted from eye tracking devices are unprocessed. This means that, in order to be able to apply statistical techniques or machine learning algorithms (supervised or unsupervised), they must first be processed. According to García et al. [52], pp. 10–13, the steps in the procedure are:

1. Produce the database.
2. Define the problem.
3. Understand the problem.
4. Process the data, with the following sub-steps:
  - 4.1. Prepare the data: clean the data, which includes operations to correct erroneous data, filter incorrect data from the dataset, and reduce unnecessary data. Other data cleaning tasks involve detecting discrepancies and dirty data (fragments of original data that make no sense)—these tasks are more closely related to understanding the original data and usually need human involvement in the process—and transformation of the data (in this step, the data are converted or consolidated so that the result of the data-mining process can be applied or be more effective). Sub-tasks within transformation include smoothing, construction of characteristics, and data aggregation or data summary; these require human supervision.
  - 4.2. Integrate the data (covers the fusion of data from various data stores). This process is applied to avoid redundancies and inconsistencies in the resulting dataset. Typical operations in data integration are identification and unification of variables and domains, analysis of the correlation of attributes, duplication of tuples, and detection of conflicts in data values from different sources.
  - 4.3. Normalise the data (this refers to the unity of the measure used, as this may affect the data analysis). This means that all of the attributes must be expressed in the same units of measurement and must use a common scale or range. This step is particularly useful in statistical learning methods.
  - 4.4. Impute missing data (this is a form of cleaning data, the aim of which is to fill in missing values using a reasonable estimation via different methods, such as applying a mean value).
  - 4.5. Identify noise (this refers to the act of smoothing in data transformation, the main objective of which is to detect random errors or variance in a measured variable).
5. Apply machine learning techniques.
6. Evaluation.
7. Exploit the results.

In summary, using integrated multichannel eye tracking technology can be of great help to professionals in various fields (marketing, health, education, driving, etc.). However, the processed information these devices provide is very basic. Therefore, we believe that ordinary users (researchers, teachers, professionals, etc.) should be able to get information about the results simply and easily [53]. Current research indicates that machine learning techniques can help make data analysis more effective, as they help people to interpret the data [54–57]. Many concepts in machine learning are related to ideas that were previously addressed in psychology and neuroscience [58].

Therefore, our research field could benefit from proposals for automating the preprocessing and processing [52] of data recorded by devices applying integrated eye tracking techniques (i.e., data from eye tracking and other data such from integrated GSR devices or electroencephalographic response (EEG)). There is a gap between the output of such software and the data analysis that researchers would perform. The present paper seeks to

bridge that gap, aiming to explore this issue in greater depth in order to help researchers in this field.

In view of the above, the objective of the current study was to suggest a procedure for processing data extracted from these types of devices by applying integrated multichannel, eye tracking technology.

## 2. Materials and Methods

### 2.1. Participants

A convenience sample was used to illustrate the proposed procedure to readers by example. More specifically, the study used a sample of 17 students (nine women and eight men) in the fourth year of a biomedical engineering degree at the University of Burgos. They were between 21 and 22 years old, with a standard deviation for both genders between 0.7 and 0.9 years. Participation was voluntary and without any financial reward. The students were told about the study objectives and signed their informed consent before taking part. In addition, the study was approved beforehand by the University of Burgos Bioethics Committee (report no. IO 03/2022).

### 2.2. Instruments

- (a) Tobii Pro Lab v. 1.241.54542 [59]. This is a hardware and software platform that is used by researchers to analyse human behaviour. It has a high degree of flexibility and can be used to perform advanced studies on attention and provide deep analysis of cognitive processes. This software was chosen due to it currently being one of the most widely used in research [60], as it makes it easy to integrate other measuring instruments, such as GSR.
- (b) Shimmer 3 GSR+ [61]. The Shimmer3 GSR+ (Galvanic Skin Response) unit provides connections and pre-amplification for a Galvanic Skin Response (Electrodermal Resistance Measurement—EDR/Electrodermal Activity (EDA) data acquisition channel). The GSR+ unit is suitable for measuring the electrical characteristics or conductance of the skin. The device is compatible for use in conjunction with Tobii Pro Lab v. 1.241.54542 and the metrics are integrated with those obtained with the Tobii Pro Lab v. 1.241.54542 device.
- (c) Virtual Laboratory 8—Eye tracking technology applied to Early Intervention II. A virtual classroom (also called a lab) in the eEarlyCare-T Research Project. This classroom is free and open access after logging into the platform [https://www2.ubu.es/earlycare\\_t/en/project](https://www2.ubu.es/earlycare_t/en/project) (accessed on 8 August 2024) [62]. Images from the virtual classroom are provided in Figure A1.
- (d) Python libraries: Pandas, sklearn, matplotlib, seaborn.

### 2.3. Procedure

The objective of this study was to suggest a procedure for processing data extracted from an eye tracking device that incorporated GSR recording. More specifically, the study used a Tobii Pro Lab v. 1.241.54542 [59] with Shimmer 3 GSR+ [61]. It addressed the technical difficulties and record processing challenges that users may have to deal with when using this integrated multichannel technology. Lastly, the study also aimed to propose potential solutions for automating data analysis via the design of a Jupyter Notebook in which machine learning algorithms were applied.

Researchers do not normally work with very large samples of participants because the experimental process for these types of studies is laborious and needs to be tailored. This means spending a lot of time on each user in each experimental phase. Despite the small numbers of participants, each one produces a large amount of diverse records, meaning a huge volume of data. Integrated multichannel eye tracking devices facilitate the processing and analysis of data through frequency analysis—for example, through data extracted using the heat map model, which visualises the most common gaze points recorded for the object of study, be that an image, text, video, or, web page. Figure 2 shows an example.



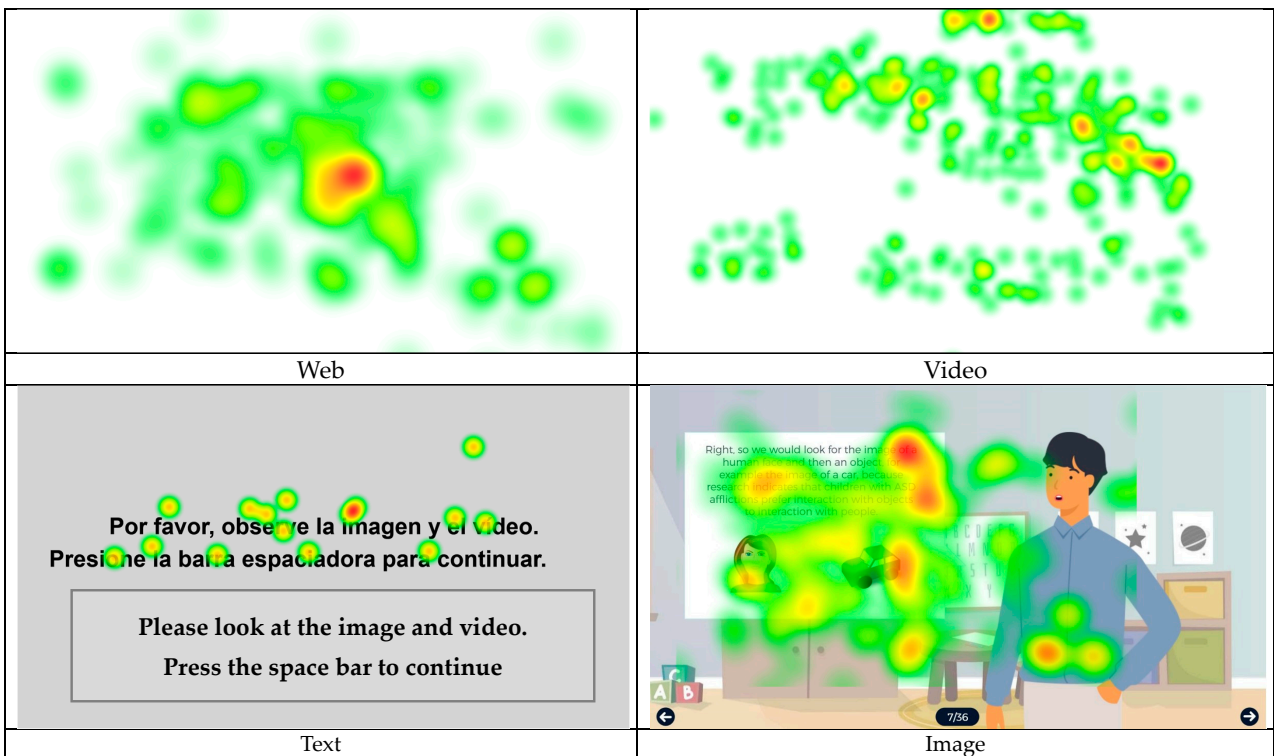


Figure 2. Heat Map for different stimuli (web, video, text, and image).

In addition, integrated multi-channel eye tracking technology provides information about the gaze positioning chain in hierarchical order of appearance. This is a dynamic measure called scan path or gaze point, an example of which is shown in Figure 3.

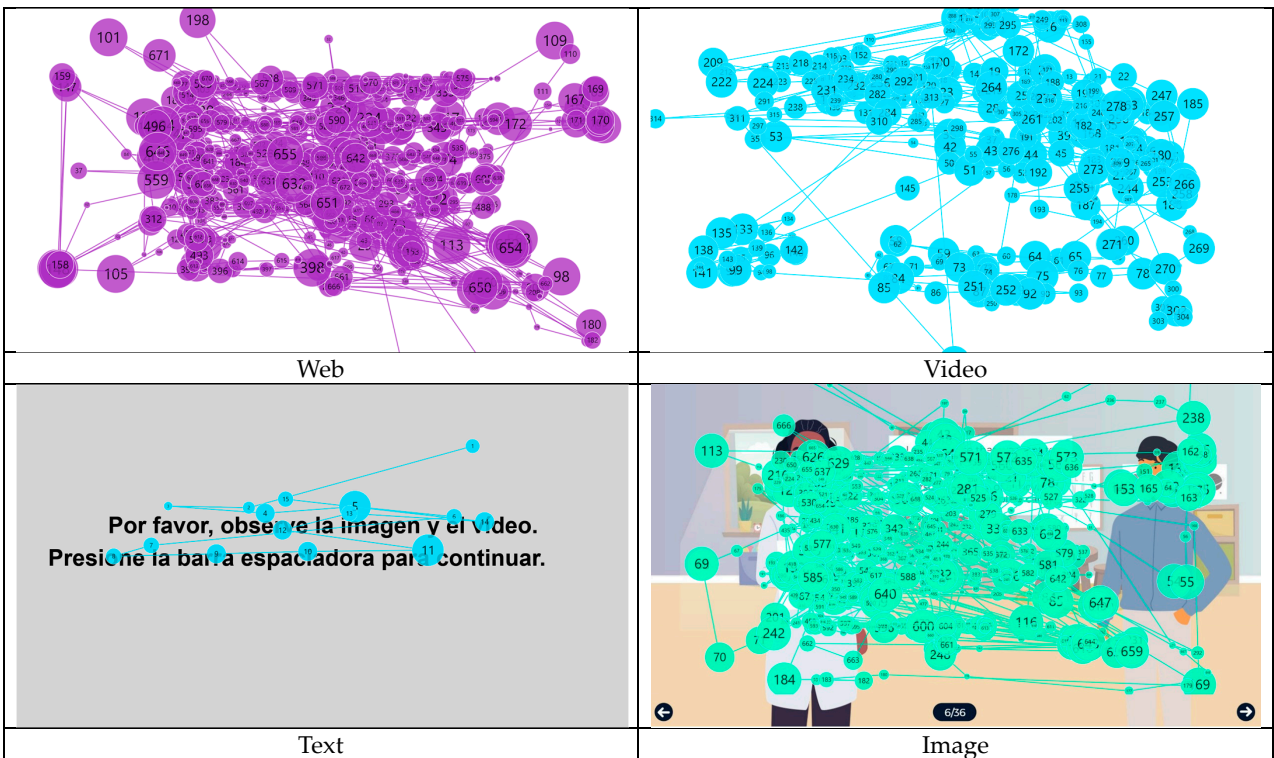


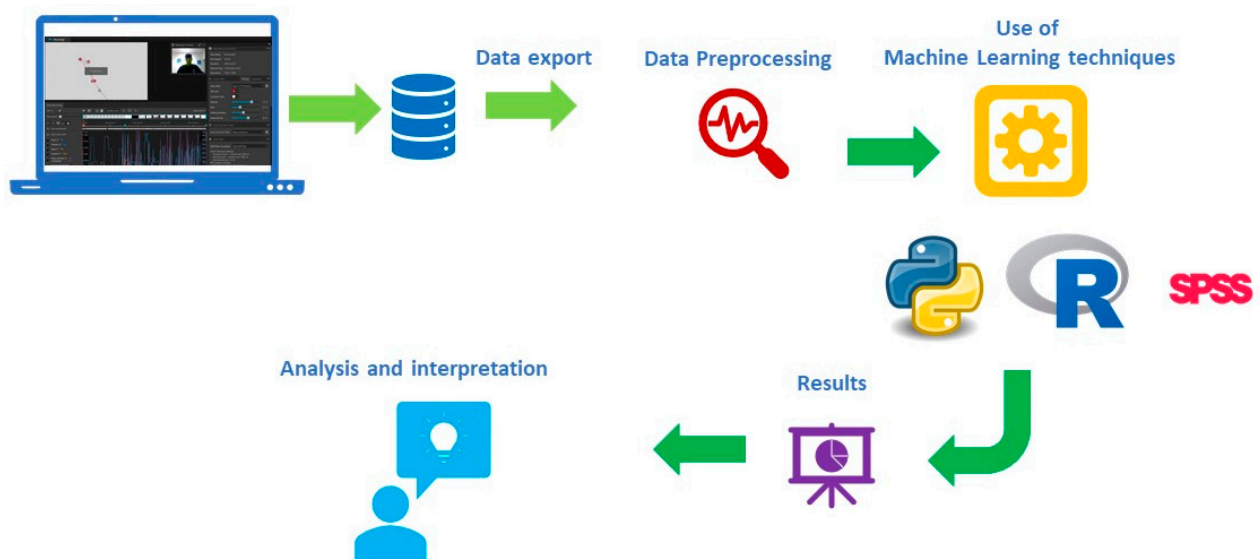
Figure 3. Gaze Point in different stimuli (web, video, text, and image).

Furthermore, a variety of metrics can be extracted for each participant once Areas of Interest (AOIs) have been established. AOIs can be divided into relevant AOIs and non-relevant AOIs. Relevant AOIs refer to the essential information in the learning process. The users should pay more attention to them. Non-relevant AOIs refer to information that the user should not pay attention to. Table 2 gives an example of some of these metrics, the meaning of the record, and the unit of measurement. These metrics refer specifically to the Tobii Pro Lab technology v. 1.241.54542 [59]. In addition, researchers can extract metrics for saccades and fixations about parameters of duration, start, end, mapping, direction, etc. These records can be produced in various formats, including .xlsx, .tsv, .plog, etc.

**Table 2.** Most representative metrics, meaning, and unit of measurement.

Metric	Meaning	Unit of Measurement
Fixation point	The normalized horizontal and vertical coordinate of the fixation point.	Normalized coordinates (DACS)
Average pupil diameter	The average diameter of the pupil of the fixation. Calculated using the resulting pupil diameter after applying pupil diameter.	Millimeters
Saccade direction	The angle of the straight line between the preceding fixation and succeeding fixation. This can only be applied to whole saccades.	Degrees
Average velocity	The average velocity across all samples belonging to the saccade, even outside the interval.	Degrees/second
Peak velocity	The maximum velocity across all samples belonging to the saccade, even outside the interval.	Degrees/second
Saccade amplitude	The amplitude for whole saccades.	Degrees
Mouse position	The position of the mouse.	Pixels (DACS)
Galvanic skin response (GSR)	The raw galvanic skin response signal of the participant.	Microsiemens
Average GSR	The average galvanic skin response (GSR) signal, after filtering, for an interval.	Microsiemens
Number of SCR	The number of skin conductance responses (SCRs) for an interval.	Count
Amplitude of event related SCR	The amplitude of each event-related skin conductance response (ER-SCR), for an interval. ER-SCRs are calculated using filtered GSR data.	Microsiemens

However, in order to be able to analyse the recorded data, researchers must process and analyse most of these records. Furthermore, once processed, statistical or machine learning techniques (supervised or unsupervised) must be applied in order to analyse and subsequently interpret the data [52]. Figure 4 shows an outline of the process that should be followed. It applies a more complex log analysis procedure than the default analysis offered by the devices.



**Figure 4.** Procedure for analysing records produced with integrated multichannel eye tracking technology.

## 2.4. Data Analysis

The objective of this study was not to test any research hypotheses, but rather to suggest how to create a Notebook in Jupyter as a way of providing an example of how one might process data recorded via integrated multichannel eye tracking technology (in this case incorporating a GSR device). More specifically, we applied supervised machine learning algorithms (regression—linear regression in this case) and unsupervised machine learning algorithms (clustering: using  $k$ -means) using the Python libraries: pandas, sklearn, matplotlib, and seaborn.

## 3. Results

A Notebook was made in Jupyter with the steps for importing and processing the database extracted from Tobii Pro Lab v. 1.241.54542 [59], after a pre-processing phase. In this phase, all detected noise was cleaned. The steps to be followed are described below.

**Step 1.** Clean and filter the files (one file per participant was exported), in this case, they were extracted in .xlsx format.

**Step 2.** Import libraries to read the extracted files.

```
Import
import pandas as pd
```

**Step 3.** Create a function to concatenate the files

### **Input:**

- data\_folder: path to the folder containing data files
- extension: file extension (e.g., .xlsx)

### **Output:**

- concatenated\_data: A DataFrame formed by concatenating all the files.

### **Process:**

1. **Initialize** files as the list of files in data\_folder.
2. **Initialize** an empty list list\_dataframes.
3. **For** each file in files:
  - Set full\_path as the complete path of the file.
  - **If** full\_path is a valid file and its extension matches extension:
    - **If** the extension is .xlsx:
      - Display message: "Reading file i + 1/len(files): file".
      - Read the file into a DataFrame df.
      - Append df to list\_dataframes.
    - **Else:** Continue to the next file.
  - **If** an error occurs while reading the file:
    - Display message: "Error reading file file: error\_message".
4. **If** list\_dataframes is not empty:
  - Concatenate all DataFrames in list\_dataframes into concatenated\_data.
  - Display message: "All files have been processed!".
  - Return concatenated\_data.
5. **Else:**
  - Display message: "No files were found for processing".
  - Return None.

**Step 4.** Data Preprocessing.

### **Input:**

- data: DataFrame containing the data



- `delete_rows`: dictionary where keys are column names and values are rows to be deleted
- `filter_rows`: dictionary where keys are column names and values are rows of interest
- `select_columns`: list containing the columns to be kept

**Output:**

- `processed_data`: A DataFrame with the processed data.

**Process:**

1. **For each column, values in `delete_rows`:**
  - Remove rows from data where the column contains any of the values.
2. **For each column, values in `filter_rows`:**
  - Filter data to keep only rows where the column contains any of the values.
3. **Select columns of interest in data using `select_columns`.**
4. **Return `processed_data`, which is now the processed DataFrame.**

**Step 5. Data Processing with Filters and Selection.****Process:**

1. **Create a dictionary `delete_rows`:**
  - Keys: 'Participant name'
  - Values: ['Test XX 2', 'Participant\_X3', 'NP\_01', 'NP\_01\_2', 'NP\_02']
  - Purpose: To remove rows where the 'Participant name' matches any value in the list.
2. **Create a dictionary `filter_rows`:**
  - Keys: 'Sensor'
  - Values: ['Mouse', 'GSR']
  - Purpose: To filter rows where the 'Sensor' column contains only 'Mouse' and 'GSR'.
3. **Create a list `select_columns`:**
  - ['Participant name', 'Gender', 'Audio', 'Recording duration', 'Eye movement type', 'Galvanic skin response (GSR)']
  - Purpose: To select the columns of interest from the DataFrame.
4. **Apply the processing:**
  - Call `processing_Data` function with the parameters: `data`, `delete_rows`, `filter_rows`, and `select_columns`.
5. **Return the processed data as `processed_data`.**

**Step 6. Data grouping and aggregation****Input:**

- `data`: DataFrame containing participant data

**Output:**

- `grouped`: DataFrame with data grouped by 'Participant name' and aggregated metrics

**Process:**

1. **Group the data by 'Participant name', applying the following aggregation functions:**
  - **Gender:** Select the first value (assuming gender is consistent across all rows).
  - **Audio:** Select the first value (assuming the audio characteristic is consistent across all rows).
  - **Recording duration:** Select the first value (assuming test duration is the same across all rows).
  - **Eye movement type:**
    - If both 'Fixation' and 'Saccade' exist in the column:

- Compute the ratio of 'Fixation' events to 'Saccade' events.
  - **Else:** Return None.
  - **Galvanic skin response (GSR):** Compute the mean value for GSR.
2. **Reset the index** of the grouped DataFrame.
  3. **Rename** the 'Eye movement type' column to 'Fixation to Saccade Ratio' for a more descriptive name.
  4. **Return** the grouped DataFrame.

**Step 7. Load data****Input :**

- location: File path of the Excel file

**Output :**

- data: Loaded DataFrame from the Excel file

**Process:**

1. **Load** the Excel file from location into a DataFrame data
2. **Return** the data

**Step 8. Combine data****Input :**

- data1: First DataFrame
- data2: Second DataFrame
- column: Column name on which to merge both DataFrames

**Output :**

- data: Merged DataFrame based on the common column

**Process:**

1. **Merge** data1 and data2 on the column
2. **Return** the merged data

An example is presented in [Figure 5](#).

Participant name	Gender	Visualisation	Recording duration	Fixation to Saccade Ratio	Galvanic skin response (GSR)	
0	NP_01	Female	no audio	237413	7.327740	1.827789
1	NP_02	Male	no audio	134613	8.908907	2.833627
2	NP_03	Female	no audio	276957	6.652087	2.613371
3	NP_04	Female	with audio	264543	8.836283	0.582247
4	NP_05	Female	no audio	321678	5.249753	0.858351
5	NP_06	Female	no audio	296533	7.607504	0.575296
6	NP_07	Male	no audio	323631	9.199501	2.142786
7	NP_08	Female	no audio	241836	7.555053	6.657798
8	NP_09	Male	no audio	237535	7.835564	0.044381
9	NP_10	Male	no audio	250432	6.875283	6.913355
10	NP_11	Male	with audio	255166	7.862782	14.114906
11	NP_12	Male	with audio	220215	8.055693	0.013019
12	NP_13	Female	no audio	246682	7.146648	0.181956
13	NP_14	Male	no audio	399585	7.582560	0.184202
14	NP_15	Female	no audio	249501	10.839836	0.082633
15	NP_16	Male	no audio	267439	9.540541	10.012614
16	NP_17	Female	no audio	241825	5.634638	0.197920

**Figure 5.** DataFrame of the data grouped by participants.

An example of the final integration of the data is presented in Figure 6.

Participant name	Gender	Recording duration	Fixation Saccade Ratio	Glavanic skin response (GSR)	Visualisation Lab	Visualisation Type	Test success
NP_01	Female	237413	7.327740	1.827789	1	0	2
NP_02	Male	134613	8.908907	2.833627	1	0	5
NP_03	Female	276957	6.652087	2.613371	1	0	4
NP_04	Female	264543	8.836283	0.582247	1	1	0
NP_05	Female	321678	5.249753	0.858351	0	1	0
NP_06	Female	296533	7.607504	0.575296	1	0	3
NP_07	Male	323631	9.199501	2.142786	1	0	2
NP_08	Female	241836	7.555053	6.657798	1	0	1
NP_09	Male	237535	7.835564	0.044381	1	0	3
NP_10	Male	250432	6.875283	6.913355	1	0	4
NP_11	Male	255166	7.862782	14.114906	0	1	0
NP_12	Masculino	220215	8.055693	0.013019	0	1	0
NP_13	Female	246682	7.146648	0.181956	1	1	2
NP_14	Male	399585	7.582560	0.184202	1	0	0
NP_15	Female	249501	10.839836	0.082633	0	0	1
NP_16	Male	267439	9.540541	10.012614	1	0	0
NP_17	Female	241825	5.634638	0.197920	1	0	0

Figure 6. Final data integration.

### Step 9. Clustering preparation

#### Input:

- data: DataFrame containing the original data

#### Output:

- data\_normalized: DataFrame with encoded, imputed, and normalized data ready for clustering

#### Process:

1. **Detect Boolean variables:**
  - Identify Boolean columns.
  - Store these column names in `booleans`.
2. **Detect non-Boolean categorical variables:**
  - Identify categorical columns.
  - Store these column names in `categorical`.
3. **Encode categorical variables:**
  - Convert categorical variables into Boolean (binary) variables.
  - Store the result in `data_encoded`.
4. **Convert Boolean columns to binary values:**
  - For each column in `booleans`:
    - Convert the column values to binary (0 or 1).
5. **Handle missing values:**
  - Apply a **mean imputation** strategy to fill missing values:
    - Impute missing values using mean in `data_encoded`.
    - Store the imputed data in `data_filled`.
6. **Normalize the data:**
  - Apply **StandardScaler** to normalize the data:

- Scale the imputed data so that all features have the same scale.
  - Store the normalized data in `data_normalized`.
- 7. **Return** the `data_normalized`.

#### Final Step: Apply Clustering Preparation to Combined Data

1. **Call** `clustering_preparation(data_combined)` to normalize and prepare the data for clustering.
2. **Return** `data_normalized`.

#### Algorithm: K-Means Clustering (Elbow Method)

##### Input:

- `data`: DataFrame containing the preprocessed data for clustering
- `k_range`: List of values representing the range of potential cluster numbers

##### Output:

- `inertia`: List of inertia values for different cluster sizes

##### Process:

1. **Initialize** an empty list `inertia` to store the inertia values for each K.
2. **For each value of k** in `k_range`:
  - **Create** a KMeans object with the following parameters:
    - `n_clusters = k`: Specifies the number of clusters.
    - `random_state = 42`: Ensures reproducibility.
    - `n_init = 10`: Number of times the KMeans algorithm will be run with different initializations.
  - **Fit** the KMeans model to `data` using the `fit` method.
  - **Append** the `inertia` value of the current KMeans model to the `inertia` list.

#### Algorithm: K-Means Clustering Visualization

##### Input:

- `data`: Original DataFrame with all data (including the cluster assignment)
- `normalised_data`: Data that has been preprocessed and normalized for clustering
- `delete_columns`: List of columns to be removed (if necessary) before visualization
- `clusters`: The number of clusters from the KMeans model

##### Output:

- A plot showing relationships between variables involved in clustering.

##### Process:

1. Initialize a figure for the plot with size (12, 12).
2. Create a pairplot using Seaborn (`sns.pairplot`):
  - Set `data` as the source DataFrame.
  - Specify the variables to plot using `numerical_cols` (the numerical columns involved in the clustering).
  - Set `hue` to 'Cluster' to color the points by their assigned cluster.
  - Use the `palette='tab10'` to specify a color palette for the clusters.
  - Set `plot_kws` with additional plot options:
    - `alpha = 0.6`: Transparency of the points.
    - `s = 100`: Size of the scatter plot points.
    - `edgecolor = 'w'`: White edge around the points.
3. Add a title to the plot:
  - Title format: "Relationships between variables involved in clustering with KMeans({clusters} clusters)".
  - Adjust the title position with `y=1.02`.
4. Save the plot as a PDF file:

- File name: Variables\_KMeans\_{clusters}Cluster.pdf.
5. Show the plot using `plt.show()`.  
In this case the graph points to the application of  $k = 2$  (see Figure 7).

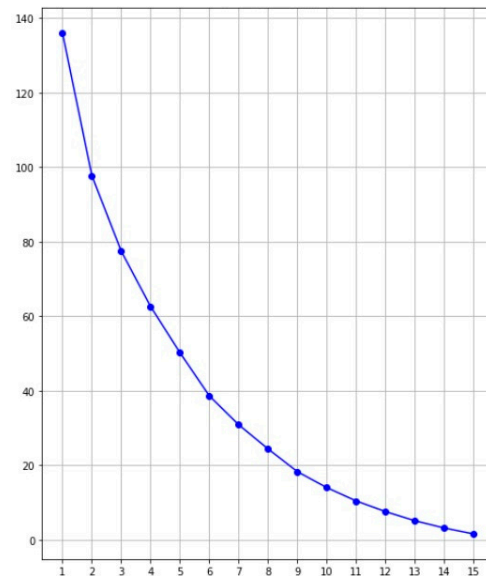


Figure 7. Graph of the elbow method.

An example of the graph obtained is shown in Figure 8.

#### Step 10. Linear Regression and Feature Impact Visualization

##### Input:

- `data`: DataFrame containing the feature variables and target variable
- `target_column`: The column name of the target variable to be predicted

##### Output:

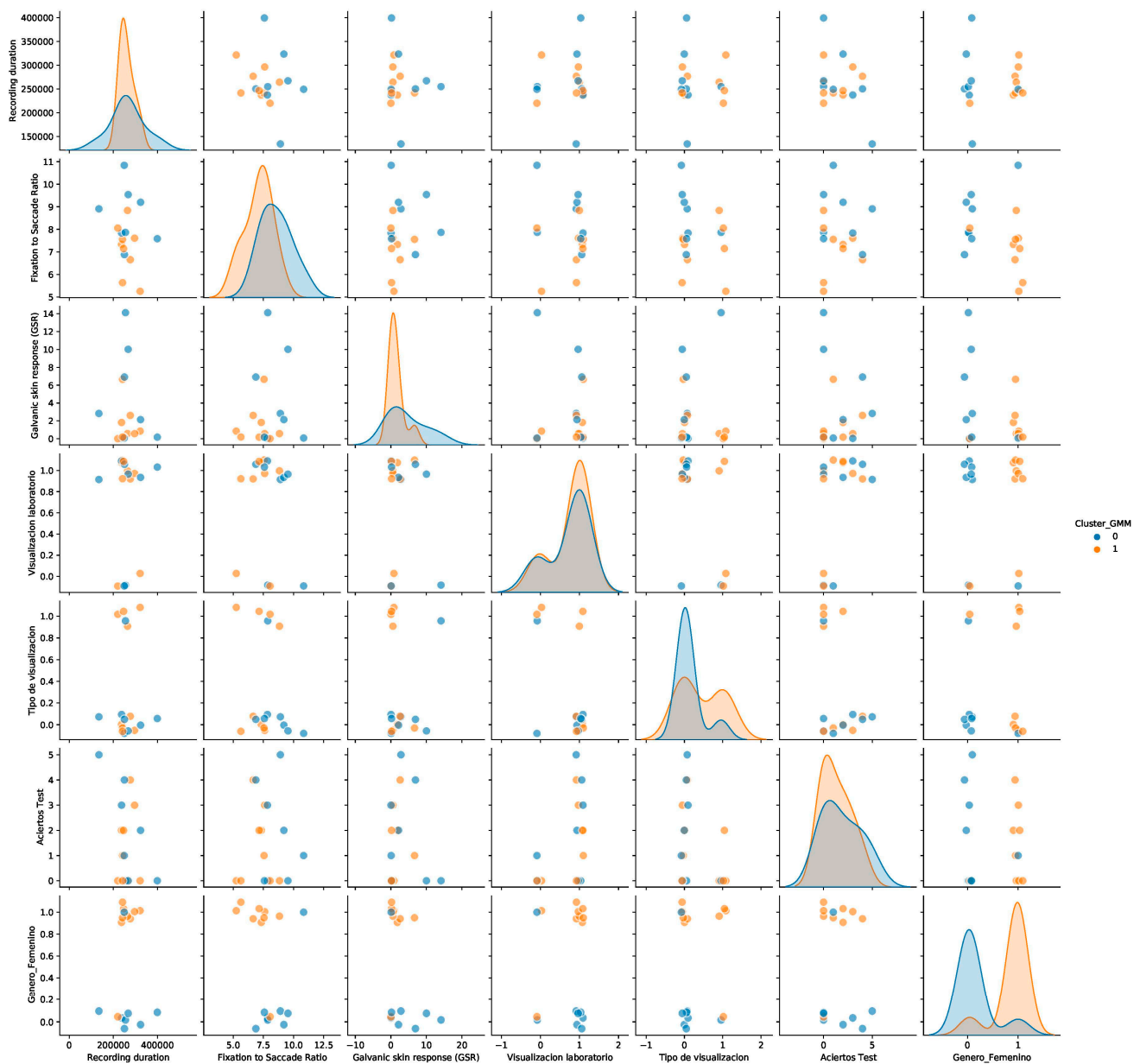
- `df_coefficients`: DataFrame containing the coefficients and  $p$ -values of each feature

##### Process:

1. **Split the DataFrame** into features ( $X$ ) and target variable ( $y$ ):
  - Set  $y$  as the target variable column.
  - Set  $X$  as the remaining features.
2. **Scale the features** using `StandardScaler`:
  - Apply `StandardScaler` to scale the feature values to the same scale.
  - Store the scaled data in `X_scaled`.
3. **Initialize the Linear Regression model**:
  - Create a `LinearRegression` object.
4. **Create empty lists** to store coefficients and  $p$ -values:
  - `coefficients`: Stores the regression coefficients (weights).
  - `p_values`: Stores the  $p$ -values for each coefficient, indicating their statistical significance.
5. **Fit the linear regression model for each feature**:
  - **For each feature column** in  $X$ :
    - Extract the feature column from `X_scaled`: `X_temporal = X_scaled[:, [i]]`.
    - Fit the model.
    - Append the coefficient (`model.coef_[0]`) to `coefficients`.
    - Add a constant column to `X_temporal` for calculating the intercept: `X_temporal = X_temporal.add_constant(X_temporal)`.



- Create and fit the OLS model.
  - Append the  $p$ -value (`model_sm.pvalues [1]`) to `p_values`.
6. **Create a DataFrame** with the coefficients and  $p$ -values:
    - Store the results in `df_coefficients`.
  7. **Sort the DataFrame** by the absolute value of the coefficient:
    - Sort `df_coefficients` by coefficient magnitude.
  8. **Visualize the regression coefficients:**
    - Create a figure for the bar plot.
    - Use Seaborn to create a bar plot.
    - Set plot titles and axis labels.
    - Save the plot as a PDF.
    - Show the plot.
  9. **Return the DataFrame** with the sorted coefficients and  $p$ -values (`df_coefficients`).



**Figure 8.** Scatter plot of the relationship between all variables.

Specifically, the regression coefficients and their significance at  $\alpha = 0.05$  were found. As Table 3 shows, none of them were significant. There was only a pattern of difference in the variables “Type of presentation”  $\alpha = 0.07$  and “Recording duration”.

**Table 3.** Lineal Regression for each of the variables.

Variables (Characteristics)	Standardized Beta Coefficients	Standard Errors of Coefficients	<i>p</i> -Value
Type of presentation	0.48	0.23	0.07
Recording duration	−0.45	0.23	0.07
Gender	−0.15	0.25	0.57
GSR	−0.16	0.24	0.73
Fixation saccade ratio	−0.19	0.24	0.52

#### 4. Discussion

The aim of this study was to highlight the need for integrated multichannel eye tracking devices to also offer—in addition to analysis of metrics based on using simple statistical tests (e.g., duration of an interval, number of events, time of each event, total duration of a fixation on an AOI, mean duration of fixations on an AOI, the number of fixations on an AOI, the duration of visits to a stimulus, the number of visits, the number of clicks on an AOI, gaze direction, eye position, mean duration of fixations, minimum duration of fixations, maximum duration of fixations, pupil diameter, etc.)—the possibility of processing records by applying other analytical techniques, such as more complex statistical or machine learning techniques (both supervised and unsupervised). Various studies have made effective use of supervised machine learning techniques (classification and regression) and unsupervised techniques (clustering) to interpret eye tracking data [50,51,54–57]. The objective is for the processing of all of this data to be done as smoothly as possible, automating the phases that researchers currently have to do by hand [52]. This paper provides an example of how data may be extracted and how machine learning algorithms may be automated, specifically in this case with the Python programming language—although the procedure may be used with other languages, such as R and Matlab.

#### 5. Conclusions

Currently, devices applying eye tracking technology give researchers and users simple information for data analysis (frequencies, means, simple analysis of gaze paths for each participant). Extracting this data is relatively painless. However, the devices record much more information than that. To make the best analysis of all of this data, these records need to be pre-processed and processed. This requires an understanding of computation or IT. This is a major limitation on making use of the potential offered by multichannel integrated eye tracking technology. This is a pressing issue, as we believe that this technology should be habitually applied in advertising, health, education, and driving (among other areas).

##### 5.1. Study Limitations

The limitations of this study are related to the small sample of user data. Because of that, future studies will increase the size of the samples in order to produce more versatile Jupyter Notebooks that can be used in different research fields. In addition, this study only applied two machine learning algorithms (supervised: linear regression; unsupervised: *k*-means clustering). Because of that, other studies will expand the use of supervised and unsupervised algorithms.

##### 5.2. Future Lines of Research

The use of integrated multichannel eye tracking technology can offer professionals in marketing, education, and health (among others) real-time analysis of the processing of all of the information the devices record. This functionality will be extremely useful in various fields of application (for example, in education it will be useful for teachers to

have prediction, classification, or clustering data to help them create tailored educational resources. Similarly, doctors and psychologists may find help in individualised differential diagnosis and treatment). In other words, it opens up a wide range of possibilities for practical application. However, the obstacle of data processing means that this technology is not being used to its fullest potential. In this article, we have presented an example of how the phases of data pre-processing and processing may be automated, including the use of supervised and unsupervised machine learning algorithms. We would encourage researchers to continue working along these lines, and those responsible for the technology to incorporate these analytical resources into their multichannel eye tracking devices. Current researchers who avail themselves of machine learning techniques are more effective at analysing eye tracking data. All of this will contribute to this technology being used more habitually, and to improving the conclusions studies can produce, helping to improve research in the fields in which it is applied.

**Author Contributions:** Conceptualisation, all authors; methodology, all authors; software, all authors; validation, all authors; formal analysis, all authors; investigation, all authors; resources, all authors; data curation, all authors; writing—original draft preparation, M.C.S.-M. and R.M.-S.; writing—review and editing, all authors; visualisation, all authors; supervision, all authors; project administration, all authors; funding acquisition, M.C.S.-M. and R.M.-S. All authors have read and agreed to the published version of the manuscript.

**Funding:** Project “Voice assistants and artificial intelligence in Moodle: a path towards a smart university” SmartLearnUni. Call 2020 R&D&I Projects—RTI Type B. MINISTRY OF SCIENCE AND INNOVATION AND UNIVERSITIES. STATE RESEARCH AGENCY. Government of Spain, grant number PID2020-117111RB-I00”. Specifically, in the part concerning the application of multichannel eye tracking technology with university students and Project “Specialized and updated training on supporting advance technologies for early childhood education and care professionals and graduates” (eEarlyCare-T) grant number 2021-1-ES01-KA220-SCH-000032661 funded by the EUROPEAN COMMISSION. In particular, the funding has enabled the development of the e-learning classroom and educational materials.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of UNIVERSITY OF BURGOS (protocol code IO 03/2022 and date of approval 7 April 2022).

**Data Availability Statement:** We confirm that the data is not available for ethical reasons, but that it can be requested as long as the person, the institution to which it belongs and the purposes for which the data is used are credited.

**Conflicts of Interest:** The authors declare no conflicts of interest.

Appendix A

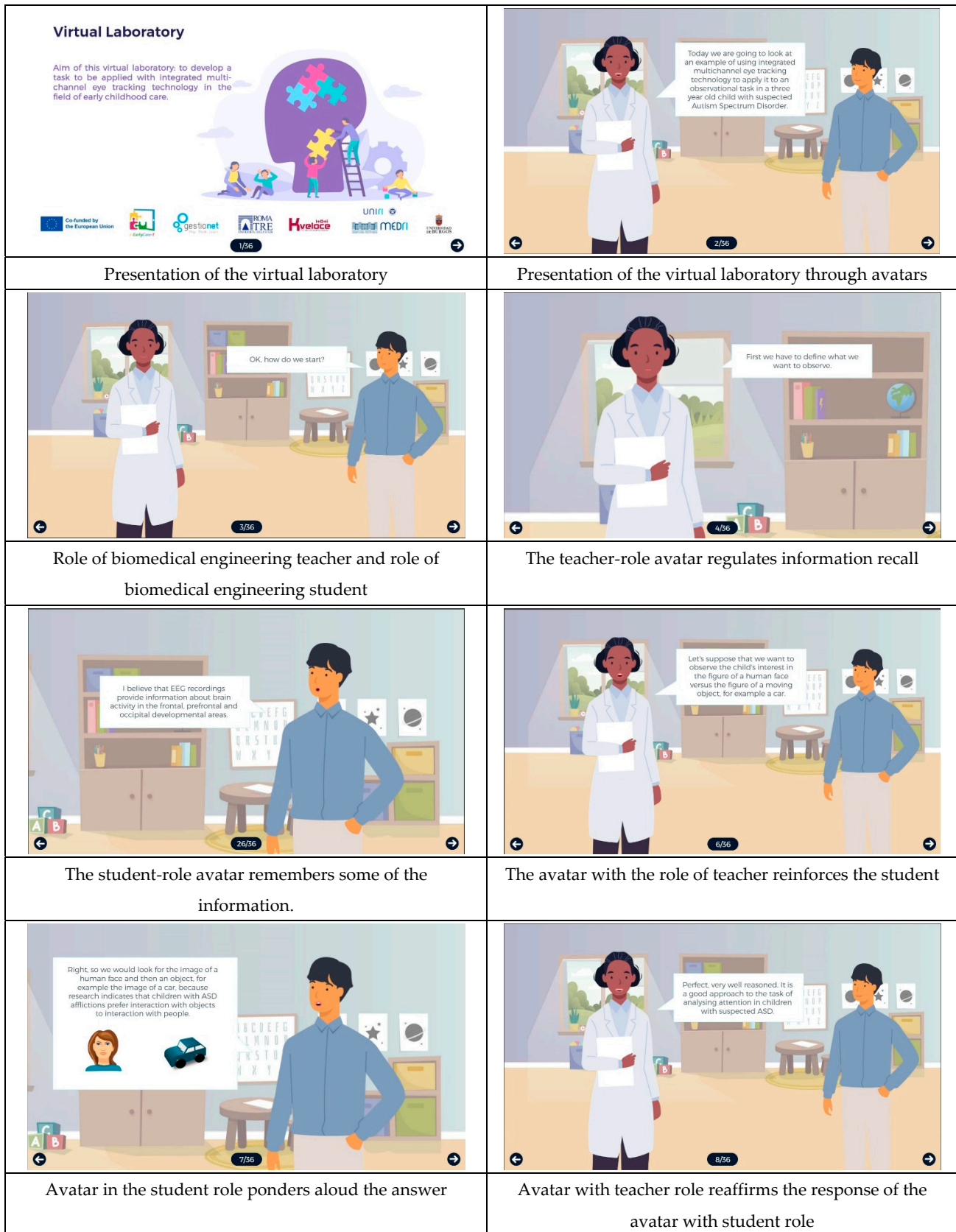


Figure A1. Cont.

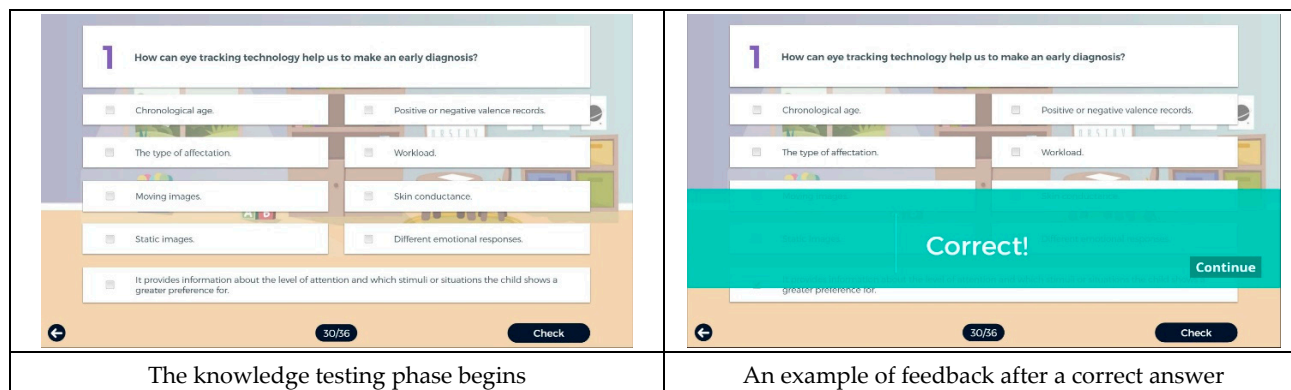


Figure A1. Description of the virtual laboratory.

## References

- Retamosa, M.; Aliagas, I.; Millán, A. Displaying ingredients on healthy snack packaging: A study on visual attention, choice, and purchase intention. *J. Sens. Stud.* **2024**, *39*, e12944. [\[CrossRef\]](#)
- Bajaj, R.; Ali Syed, A.; Singh, S. Analysing applications of neuromarketing in efficacy of programmatic advertising. *J. Consum. Behav.* **2024**, *23*, 939–958. [\[CrossRef\]](#)
- Bhardwaj, S.; Thapa, S.B.; Gandhi, A. Advances in neuromarketing and improved understanding of consumer behaviour: Analysing tool possibilities and research trends. *Cogent Bus. Manag.* **2024**, *11*, 2376773. [\[CrossRef\]](#)
- Calderón-Fajardo, V.; Anaya-Sánchez, R.; Rejón-Guardia, F.; Molinillo, S. Neurotourism Insights: Eye Tracking and Galvanic Analysis of Tourism Destination Brand Logos and AI Visuals. *Tour. Manag. Stud.* **2024**, *20*, 53–78. [\[CrossRef\]](#)
- Modi, N.; Singh, J. An analysis of perfume packaging designs on consumer's cognitive and emotional behavior using eye gaze tracking. *Multimed. Tools Appl.* **2024**, *83*, 82563–82588. [\[CrossRef\]](#)
- Thiebaut, R.; Elshourbagi, A. The Effect of Neuromarketing and Subconscious Branding on Business Profitability and Brand Image: A New Business Model Innovation for Startups. In *Fostering Global Entrepreneurship Through Business Model Innovation*; Gupta, V., Ed.; IGI Global: Hershey, PA, USA, 2024; pp. 217–252. [\[CrossRef\]](#)
- Cong, L.; Luan, S.; Young, E.; Miroso, M.; Bremer, P.; Torricco, D.D. The Application of Biometric Approaches in Agri-Food Marketing: A Systematic Literature Review. *Foods* **2023**, *12*, 2982. [\[CrossRef\]](#) [\[PubMed\]](#)
- Al-Nafjan, A.; Aldayel, M.; Kharrat, A. Systematic Review and Future Direction of Neuro-Tourism Research. *Brain Sci.* **2023**, *13*, 682. [\[CrossRef\]](#) [\[PubMed\]](#)
- Khondakar, F.K.; Sarowar, H.; Chowdhury, M.H.; Majumder, S.; Hossain, A.; Dewan, M.A.A.; Hossain, Q.D. A systematic review on EEG-based neuromarketing: Recent trends and analyzing techniques. *Brain Inform.* **2024**, *11*, 17. [\[CrossRef\]](#)
- Tavares-Filho, E.R.; Hidalgo, L.G.S.; Lima, L.M.; Spers, E.E.; Pimentel, T.C.; Esmerino, E.A.; Cruz, A.G. Impact of animal origin of milk, processing technology, type of product, and price on the Boursin cheese choice process: Insights of a discrete choice experiment and eye tracking. *J. Food Sci.* **2024**, *89*, 640–655. [\[CrossRef\]](#)
- Madlenak, R.; Chinoracky, R.; Stalmasekova, N.; Madlenakova, L. Investigating the Effect of Outdoor Advertising on Consumer Decisions: An Eye-Tracking and A/B Testing Study of Car Drivers' Perception. *Appl. Sci.* **2023**, *13*, 6808. [\[CrossRef\]](#)
- Kim, M.; Lee, J.; Lee, S.Y.; Ha, M.; Park, I.; Jang, J.; Jang, M.; Park, S.; Kwon, J.S. Development of an eye-tracking system based on a deep learning model to assess executive function in patients with mental illnesses. *Sci. Rep.* **2024**, *14*, 18186. [\[CrossRef\]](#) [\[PubMed\]](#)
- Perkovich, E.; Laakman, A.; Mire, S.; Yoshida, H. Conducting head-mounted eye-tracking research with young children with autism and children with increased likelihood of later autism diagnosis. *J. Neurodev. Disord.* **2024**, *16*, 7. [\[CrossRef\]](#) [\[PubMed\]](#)
- Amirbay, A.; Mukhanova, A.; Baigabylov, N.; Kudabekov, M.; Mukhambetova, K.; Baigusheva, K.; Baibulova, M.; Ospanova, T. Development of an algorithm for identifying the autism spectrum based on features using deep learning methods. *Int. J. Electr. Comput. Eng. (IJECE)* **2024**, *14*, 5513–5523. [\[CrossRef\]](#)
- Bent, C.; Glencross, S.; McKinnon, K.; Hudry, K.; Dissanayake, C.; The Victorian ASELCC Team; Vivanti, G. Predictors of Developmental and Adaptive Behaviour Outcomes in Response to Early Intensive Behavioural Intervention and the Early Start Denver Model. *J. Autism Dev. Disord.* **2024**, *54*, 2668–2681. [\[CrossRef\]](#)
- Ibragimov, B.; Mello-Thoms, C. The Use of Machine Learning in Eye Tracking Studies in Medical Imaging: A Review. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 3597–3612. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mehmood, I.; Li, H.; Umer, W.; Tariq, S.; Wu, H. Non-invasive detection of mental fatigue in construction equipment operators through geometric measurements of facial features. *J. Saf. Res.* **2024**, *89*, 234–250. [\[CrossRef\]](#)
- Schmidt, A.; Mohareri, O.; DiMaio, S.; Yip, M.C.; Salcudean, S.E. Tracking and mapping in medical computer vision: A review. *Med. Image Anal.* **2024**, *94*, 103131. [\[CrossRef\]](#) [\[PubMed\]](#)
- Boujelbane, M.A.; Trabelsi, K.; Salem, A.; Ammar, A.; Glenn, J.M.; Boukhris, O.; AlRashid, M.M.; Jahrami, H.; Chtourou, H. Eye Tracking During Visual Paired-Comparison Tasks: A Systematic Review and Meta-Analysis of the Diagnostic Test Accuracy for Detecting Cognitive Decline. *J. Alzheimer's Dis.* **2024**, *99*, 207–221. [\[CrossRef\]](#)



20. Klotzek, A.; Jemni, M.; Groves, S.J.; Carrick, F.R. Effects of Cervical Spinal Manipulation on Saccadic Eye Movements. *Brain Sci.* **2024**, *14*, 292. [[CrossRef](#)]
21. Pauszek, J.R. An introduction to eye tracking in human factors healthcare research and medical device testing. *Hum. Factors Healthc.* **2023**, *3*, 100031. [[CrossRef](#)]
22. Passaro, A.; Zullo, A.; Di Gioia, M.; Curcio, E.; Stasolla, F. A Narrative Review on the Use of Eye-Tracking in Rett Syndrome: Implications for Diagnosis and Treatment. *OBM Genet.* **2024**, *8*, 250. [[CrossRef](#)]
23. Hill, W.; Lindner, H. Using eye tracking to assess learning of a multifunction prosthetic hand: An exploratory study from a rehabilitation perspective. *J. Neuroeng. Rehabil.* **2024**, *21*, 148. [[CrossRef](#)] [[PubMed](#)]
24. Pulay, M.Á.; Szabó, É.F. Developing Visual Perceptual Skills with Assistive Technology Supported Application for Children with Cerebral Palsy. *Acta Polytech. Hung.* **2024**, *21*, 25–38. [[CrossRef](#)]
25. Feldmann, L.; Zsigo, C.; Mörtl, I.; Bartling, J.; Wachinger, C.; Oort, F.; Schulte-Körne, G.; Greimel, E. Emotion regulation in adolescents with major depression—Evidence from a combined EEG and eye-tracking study. *J. Affect. Disord.* **2023**, *340*, 899–906. [[CrossRef](#)] [[PubMed](#)]
26. Tao, Z.; Sun, N.; Yuan, Z.; Chen, Z.; Liu, J.; Wang, C.; Li, S.; Ma, X.; Ji, B.; Li, K. Research on a New Intelligent and Rapid Screening Method for Depression Risk in Young People Based on Eye Tracking Technology. *Brain Sci.* **2023**, *13*, 1415. [[CrossRef](#)]
27. Brien, D.C.; Riek, H.C.; Yep, R.; Huang, J.; Coe, B.; Areshenkoff, C.; Grimes, D.; Jog, M.; Lang, A.; Marras, C.; et al. Classification and staging of Parkinson’s disease using video-based eye tracking. *Park. Relat. Disord.* **2023**, *110*, 105316. [[CrossRef](#)] [[PubMed](#)]
28. Ghiță, A.; Hernández-Serrano, O.; Moreno, M.; Monràs, M.; Gual, A.; Maurage, P.; Gacto-Sánchez, M.; Ferrer-García, M.; Porrás-García, B.; Gutiérrez-Maldonado, J. Exploring Attentional Bias toward Alcohol Content: Insights from Eye-Movement Activity. *Eur. Addict. Res.* **2024**, *30*, 65–79. [[CrossRef](#)]
29. Puttevils, L.; De Bruecker, M.; Allaert, J.; Sanchez-Lopez, A.; De Schryver, N.; Vervaet, M.; Baeken, C.; Vanderhasselt, M.-A. Attentional bias to food during free and instructed viewing in anorexia nervosa: An eye tracking study. *J. Psychiatr. Res.* **2023**, *164*, 468–476. [[CrossRef](#)]
30. Guo, X.; Liu, Y.; Tan, Y.; Xia, Z.; Fu, H. Hazard identification performance comparison between virtual reality and traditional construction safety training modes for different learning style individuals. *Saf. Sci.* **2024**, *180*, 106644. [[CrossRef](#)]
31. Virlet, L.; Sparrow, L.; Barela, J.; Berquin, P.; Bonnet, C. Proprioceptive intervention improves reading performance in developmental dyslexia: An eye-tracking study. *Res. Dev. Disabil.* **2024**, *153*, 104813. [[CrossRef](#)]
32. Liang, Z.; Ga, R.; Bai, H.; Zhao, Q.; Wang, G.; Lai, Q.; Chen, S.; Yu, Q.; Zhou, Z. Teaching expectancy improves video-based learning: Evidence from eye-movement synchronization. *Br. J. Educ. Technol.* **2024**. [[CrossRef](#)]
33. Kok, E.M.; Niehorster, D.C.; van der Gijp, A.; Rutgers, D.R.; Auffermann, W.F.; van der Schaaf, M.; Kester, L.; van Gog, T. The effects of gaze-display feedback on medical students’ self-monitoring and learning in radiology. *Adv. Health Sci. Educ.* **2024**. *online ahead of print*. [[CrossRef](#)] [[PubMed](#)]
34. Sáiz-Manzanares, M.C.; Marticorena-Sánchez, R.; Escolar-Llamazares, M.C.; González-Díez, I.; Martín Antón, L.J. Using integrated multimodal technology: A way to personalised learning in Health Sciences and Biomedical engineering Students. *Appl. Sci.* **2024**, *14*, 7017. [[CrossRef](#)]
35. Mullen, B. Self-Attention Theory: The Effects of Group Composition on the Individual. In *Theories of Group Behavior*; Springer Series in Social, Psychology; Mullen, B., Goethals, G.R., Eds.; Springer: New York, NY, USA, 1987. [[CrossRef](#)]
36. Korteland, R.J.; Kok, E.; Hulshof, C.; van Gog, T. Teaching through their eyes: Effects on optometry teachers’ adaptivity and students’ learning when teachers see students’ gaze. *Adv. Health Sci. Educ.* **2024**. [[CrossRef](#)]
37. Sáiz-Manzanares, M.C.; Marticorena-Sánchez, R.; Martín-Antón, L.J.; González-Díez, I.; Carbonero-Martín, I. Using eye tracking technology to analyse cognitive load in multichannel activities in university students. *Int. J. Hum. Comput. Interact.* **2024**, *40*, 3263–3281. [[CrossRef](#)]
38. Wronski, P.; Wensing, M.; Ghosh, S.; Gärtner, L.; Müller, W.; Koetsenruijter, J. Use of a quantitative data report in a hypothetical decision scenario for health policymaking: A computer-assisted laboratory study. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 32. [[CrossRef](#)]
39. Lee, Y.; de Jong, N.; Donkers, J.; Jarodzka, H.; van Merriënboer, J.G. Measuring Cognitive Load in Virtual Reality Training via Pupillometry. *IEEE Trans. Learn. Technol.* **2024**, *17*, 704–710. [[CrossRef](#)]
40. Wang, Y.; Lu, Y.; Shen, C.-Y.; Luo, S.-J.; Zhang, L.-Y. Exploring product style perception: A comparative eye-tracking analysis of users across varying levels of self-monitoring. *Displays* **2024**, *84*, 102790. [[CrossRef](#)]
41. Cazes, M.; Noël, A.; Jamet, E. Cognitive effects of humorous drawings on learning: An eye-tracking study. *Appl. Cogn. Psychol.* **2024**, *38*, e4178. [[CrossRef](#)]
42. Tarkowski, S.; Caban, J.; Dzieńkowski, M.; Nieoczym, A.; Zarajczyk, J. Driver’s distraction and its potential influence on the extension of reaction time. *Arch. Automot. Eng.* **2022**, *98*, 65–78. [[CrossRef](#)]
43. Cheng, G.; Di, Z.; Xie, H.; Wang, F.L. Exploring differences in self-regulated learning strategy use between high- and low-performing students in introductory programming: An analysis of eye-tracking and retrospective think-aloud data from program comprehension. *Comput. Educ.* **2024**, *208*, 104948. [[CrossRef](#)]
44. Omobolanle, O.; Abiola, A.; Nihar, G.; Mohammad, K.; Abiola, A. Detecting Learning Stages within a Sensor-Based Mixed Reality Learning Environment Using Deep Learning. *J. Comput. Civ. Eng.* **2024**, *37*, 04023011. [[CrossRef](#)]
45. Bouwer, R.; Dirx, K. The eye-mind of processing written feedback for revision. *Learn. Instr.* **2024**, *85*, 101745. [[CrossRef](#)]

46. Ferreira, C.P.; Soledad Gonzalez-Gonzalez, C.; Francisca Adamatti, D.; Moreira, F. Analysis Learning Model with Biometric Devices for Business Simulation Games: Brazilian Case Study. *IEEE Access* **2024**, *12*, 95548–95564. [[CrossRef](#)]
47. Sáiz-Manzanares, M.C.; Marticorena-Sánchez, R.; Martín-Antón, L.J.; Almeida, L.; Carbonero-Martín, I. Application and challenges of eye tracking technology in Higher Education. *Comunicar* **2023**, *76*, 1–12. [[CrossRef](#)]
48. Sáiz-Manzanares, M.C.; Ramos Pérez, I.; Arnaiz-Rodríguez, Á.; Rodríguez-Arribas, S.; Almeida, L.; Martín, C.F. Analysis of the learning process through eye tracking technology and feature selection techniques. *Appl. Sci.* **2021**, *11*, 6157. [[CrossRef](#)]
49. Holmqvist, K.L.U.; Nyström, M.L.U.; Andersson, R.L.U.; Dewhurst, R.L.U.; Halszka, J.; van de Weijer, J.L. *Eye Tracking: A Comprehensive Guide to Methods and Measures*; Oxford University Press: Oxford, UK, 2011.
50. Vortmann, L.-M.; Putze, F. Combining Implicit and Explicit Feature Extraction for Eye Tracking: Attention Classification Using a Heterogeneous Input. *Sensors* **2021**, *21*, 8205. [[CrossRef](#)]
51. Sáiz-Manzanares, M.C.; Rodríguez Diez, J.J.; Marticorena Sánchez, R.; Zaparaín Yáñez, M.J.; Cerezo Menéndez, R. Lifelong Learning from Sustainable Education: An Analysis with Eye Tracking and Data Mining Techniques. *Sustainability* **2020**, *12*, 1970. [[CrossRef](#)]
52. García, S.; Luengo, J.; Herrera, F. *Data Preprocessing in Data Mining*; Springer: London, UK, 2015. [[CrossRef](#)]
53. Thilderkvist, E.; Dobsław, F. On current limitations of online eye-tracking to study the visual processing of source code. *Inf. Softw. Technol.* **2024**, *174*, 107502. [[CrossRef](#)]
54. Cho, S.-W.; Lim, Y.-H.; Seo, K.-M.; Kim, J. Integration of eye-tracking and object detection in a deep learning system for quality inspection analysis. *J. Comput. Des. Eng.* **2024**, *11*, 158–173. [[CrossRef](#)]
55. Liu, Z.; Yeh, W.-C.; Lin, K.-Y.; Lin, C.-S.H.; Chang, C.-Y. Machine learning based approach for exploring online shopping behavior and preferences with eye tracking. *Comput. Sci. Inf. Syst.* **2024**, *21*, 593–623. [[CrossRef](#)]
56. Zhang, H.; Wang, C. Integrated neural network-based pupil tracking technology for wearable gaze tracking devices in flight training. *IEEE Access* **2024**, *12*, 133234–133244. [[CrossRef](#)]
57. Born, J.; Ram, B.; Ramachandran, N.; Romero Pinto, S.A.; Winkler, S.; Ratnam, R. Multimodal Study of the Effects of Varying Task Load Utilizing EEG, GSR and Eye-Tracking. *bioRxiv* **2019**, 798496. [[CrossRef](#)]
58. Lindsay, G.W. Attention in Psychology, Neuroscience, and Machine Learning. *Front. Comput. Neurosci.* **2024**, *14*. [[CrossRef](#)]
59. Tobii AB Corp. *Tobii Pro Lab [Computer Software]*, version 1.241.54542; Tobii Corp: Danderyd, Sweden, 2024.
60. Grabinger, L.; Hauser, F.; Wolff, C.; Mottok, J. On Eye Tracking in Software Engineering. *SN Comput. Sci.* **2024**, *5*, 729. [[CrossRef](#)]
61. Shimmer3 GSR. *Shimmer [Computer Software]*, version 3; Shimmer Corp.: Dublin, Ireland; Boston, MA, USA, 2024.
62. Sáiz-Manzanares, M.C.; Marticorena-Sánchez, R. *Manual for the Development of Self-Regulated Virtual Laboratories*; Servicio de Publicaciones de la Universidad de Burgos: Burgos, Spain, 2024. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.