What Is Hidden in Clear Sight and How to Find It—A Survey of the Integration of Artificial Intelligence and Eye Tracking

Maja Kędras and Janusz Sobecki *

Department of Informatics, Faculty of Computer Systems and Management, Wrocław University of Science and Technology, 50-370 Wrocław, Poland; maja.kedras@pwr.edu.pl
* Correspondence: janusz.sobecki@pwr.edu.pl

Abstract: This paper presents an overview of the uses of the combination of eye tracking and artificial intelligence. In the paper, several aspects of both eye tracking and applied AI methods have been analyzed. It analyzes the eye tracking hardware used along with the sampling frequency, the number of test participants, additional parameters, the extraction of features, the artificial intelligence methods used and the methods of verification of the results. Finally, it includes a comparison of the results obtained in the analyzed literature and a discussion about them.

Keywords: eye tracking; artificial intelligence; machine learning

1. Introduction

In the era of system personalization and with growing emphasis on the user experience, eye tracking technologies are becoming increasingly in demand. Since eye tracking generates an immense amount of data, it is very challenging, or even quite impossible in some cases, to process it by hand. One of the solutions to this problem is using artificial intelligence which is able to automatically identify problem areas or places of a particular user’s attention. The abundance of data is not the only issue which may be solved by artificial intelligence (AI).

Eye tracking in general is used to monitor the places of human sight concentration. Since the beginning of eye tracking studies in the nineteenth century, and over the years, many different technologies and methods have been established and introduced into various disciplines [1]. Currently, the most common technology used in eye tracking is video recording of the eyes using natural or infrared light. For over thirty years eye trackers were widely used in different types of UX and psychology studies [2], such as the usefulness of web or desktop applications, perception of information in the form of graphics or texts, comparative studies of the effectiveness of system interfaces, correlation of eye tracking data with the strategy of searching for information in web systems, etc.

Most of the eye tracking research is conducted using specialized sensors or devices, which are often very expensive and may be need specialized knowledge to operate. Despite the device used, they generate a lot of rough data, growing with the sampling rate. These data should be processed to find the pattern of the eye’s focus, which is usually carried out with the help of different AI tools, especially machine learning (ML) algorithms.

AI is a field of computer science that aims to create intelligent machines. The main issues with AI include programming computers for certain issues such as knowledge, reasoning, problem-solving, perception, learning, planning and the ability to manipulate and move objects [3]. Machine Learning (ML) is an application of AI based around the idea that we let machines have access to data and we let them learn for themselves [4]. ML algorithms use computational methods to “learn” information directly from data without relying on a predetermined equation as a model in order to become more accurate in predicting outcomes without being explicitly programmed to do so.
Currently, the application of AI methods allows researchers to conduct similar experiments using just a consumer-grade camera, which makes this kind of research more accessible. The combination of these two technologies can also be used in various eye movement recognition systems, for example, speech generation systems for paralyzed people, fatigue detection systems, or even in virtual reality games.

Combining eye tracking with artificial intelligence can bring many benefits to science. However, the number of publications on this subject remains relatively small. In this work, the existing applications of these technologies will be reviewed and their quality and usefulness assessed. Further possible directions for the development of this field will also be proposed.

The content of the paper is as follows. Section 2 introduces the research methodology of the review. In Sections 2.1 and 2.2 the applications of the combination of eye tracking and artificial intelligence are described and categorized. The following chapter describes the eye trackers and the sampling frequencies used in the analyzed literature. Section 2.3 describes the number of people participating in the re-search and provides available information about them. Section 2.4 contains information about the additional parameters used with eye tracking data. Section 2.5 categorizes the feature extraction types, whereas the eighth chapter analyzes the methods of artificial intelligence used in the research and their number per study. Section 2.7 analyzes the methods of the results’ verification and their number per study. Section 3 shows comparable results obtained in the analyzed literature. The last chapter provides a summary and discussion of the collected data.

2. Materials and Methods

This survey is based on Systematic Literature Review (SLR) methodology [5]. This methodology allows the work of other researchers in the field to be summarized in an orderly and reproducible manner. It was used here for the purpose of investigation the use of artificial intelligence in the field of eye tracking. For this purpose, the following research questions were asked:

1. Which eye trackers were used when collecting data for AI?
2. Which sampling frequencies were used when collecting data for AI?
3. What kind of non-eye tracking parameters were used when AI was used?
4. How many people participated in the experiments collecting eye tracking data to be used with AI?
5. What is the gender distribution of the participants and what age range are they in?
6. How were the features extracted?
7. How many artificial intelligence methods were used in one eye tracking study?
8. Which methods of artificial intelligence were used with eye tracking data?
9. How were the results of using AI with eye tracking data verified?

For this survey papers were collected using the Scopus database. Papers were collected from the time period from 2015 to 2020. Only papers available in English have been taken into account. There were 5 queries used:

1. eye AND tracking AND artificial AND intelligence;
2. eye AND movement AND artificial AND intelligence;
3. gaze AND estimation AND artificial AND intelligence;
4. smartphone AND eye AND tracking;
5. webcam AND eye AND tracking.

The above queries relate to the words that are present in the article keywords, titles and abstracts. The search results overlapped so duplicates were removed before proceeding to abstract analysis. Then, the criteria for rejection of the publication were selected. They are as follows:

1. The research analyzed only static images.
2. The research detected only the eye position.
3. The eye tracking data were collected using Electroencephalography (EEG).
4. Artificial intelligence was not used on eye tracking data nor to calculate eye tracking data.

5. The paper is not accessible.

After application of the procedure described above, 93 papers were selected.

The methodology employed in this systematic literature review is a robust approach for summarizing the work of other researchers in the field of eye tracking and artificial intelligence. However, in conjunction with the established selection criteria, there are several potential limitations to consider. To start with, this review relies on papers available in the Scopus database, which may not include all relevant research in the field. Additionally, the use of works exclusively in English leads to language bias which may have caused the omission of valuable studies. An important limitation is also the period from which the works for analysis were selected. The field of artificial intelligence and eye tracking is rapidly evolving, and this time frame may omit recent developments and studies that were not yet published at the time of conducting the survey. Lastly, while the search queries used are comprehensive, it is possible that some relevant studies are missed due to the limitations of the keyword search. Different terminology or less common keywords might not have been included in the search.

2.1. Applications of Artificial Intelligence Enhanced Eye Tracking

Artificial intelligence-enhanced eye tracking has its applications in many different areas. Teaching and learning applications are the largest group observed. Three of those studies use eye tracking data to predict student performance [6–9]. Another interest is reading, in terms of predicting reading ability [10], recognizing reading behavior [11,12], detecting readability [13] and detecting words which are difficult for the readers [14]. In the case of words’ analysis, their understanding was also predicted [15]. Another use of AI and eye tracking was in identifying levels of comprehension [16]. It was also used to predict SAT scores [17], cognitive abilities [18], learning curves [19] and detect the speed of learning [20]. The next usage of AI and eye tracking was concentrated on predicting the type of prior disclosure [21]. The last application from this group was in predicting the social plane of interaction of a teacher conducting their classes [22].

The second group of applications of eye tracking enhanced with artificial intelligence is in emotion recognition. First of the considered studies focused on predicting which of several emotions was being felt by the participants [23] and the second predicted the aesthetic impression of a website [24]. Others were predicting reactions to advertising [25], predicting perceived face attractiveness [26], recognizing affect [27] and recommending paintings which the participants would like [28]. The rest of the papers focused on a single emotion: excitement [29], enjoyment [30], interest [31], confidence [32], confusion [33], stress [34] and satisfaction [35]. All of the listed publications focused on participants’ emotions but two additional studies were considered which were predicting the emotions of other people using the eye tracking data of their observers [36,37].

The third group distinguished in this paper is that of medical applications. AI-enhanced eye tracking is mostly used to detect neurological disorders such as autism spectrum disorder [38,39], schizophrenia [40], Parkinson’s disease [41] or dyslexia [42]. Three of the considered publications used eye tracking data to predict whether the patient has any neurological disease [43–45] and one detected organs in fetal ultrasound images [46].

Another group which could be identified is that of human behavior. One example of that group would be detecting the type of the participant’s activity both, when using a computer [47,48] and in everyday life [49–52]. Another usage is in predicting decision strategy [53] as well as taking in consideration ethical decision making [54]. Eye tracking data were also used to teach AI to play games [55], predict dwell time in museums [56], automatically assess surgery skills [57] and detect eye contact [58]. In terms of human behavior, research about intention was also found [59–62].
The fourth group distinguished in this survey is research on predicting tiredness, attention \[63,64\] and distraction \[65–68\]. In contrast to attention and distraction, tiredness was not detected directly \[69\]. It was indicated by task demand \[70\], take-over time \[71\], mental workload \[72\], operator overload \[68\] and reduced alertness \[73\].

The fifth identified group is research using eye tracking data as a way to interact with a software. It was used as to authenticate user, both by entering the password \[74–77\] and using sclera biometrics \[78\], to detect defined gestures \[79–82\], the desired direction of movement \[83\] or choosing an answer in a questionnaire form \[84\].

A separate category is that of gaze estimation, which was carried out in 13 studies \[85–97\]. The two last studies which do not fit any of the described categories used eye tracking data to recognize and classify objects \[98\] and distinguish Chinese ethnic groups \[99\].

2.2. Eye Trackers

Researchers used the following eye trackers (Figure 1). It is worth noting that in two studies \[12,24\] two different eye tracking devices were used. One used SMI and Tobii and the other used Tobii and EyeTribe eye trackers \[24\]. In both cases they were used in parallel and were not compared.

![Figure 1. Percentage distribution of the eye trackers used in the analyzed works.](image-url)

Almost a quarter of the studies were conducted using Tobii hardware, and more than one tenth used a simple web camera. Sadly over 8% of papers did not specify the type of eye tracker used, but even without that data we can say that there are a lot of options for eye tracking. In total, 12.9% of studies used hardware which was used only by them in the
scope of the analyzed research. On the one hand, it prevents monopolization of the market by a single manufacturer, which also means that eye tracking research is more accessible to perform, but on the other hand it may make studies harder to compare.

Since Tobii eye trackers have been used the most times, it is possible to observe the use of many different models. The most common was the Tobii T120 (5 papers), next was the Tobii EyeX (4 papers) and lastly the Tobii T60, Tobii X1 and Tobii X2-30 (two papers each). The Tobii hardware used in only one paper were the Tobii 175, Tobii 4C, Tobii Steelseries Sentry, Tobii TX300, Tobii X2-60, Tobii X3-120 and Tobii X300. One paper did not specify the exact model used. Regarding the SMI hardware, the most popular model was the SMI RED 250 (4 papers). One paper used the SMI RED 4 and two did not specify the model they used.

In terms of sampling rate, 4.3% of studies used external data and 38.7% of studies did not specify it. The remaining 57% papers used the sampling rates shown in the graph below. When an interval was given, the minimum value was included (Figure 2).

![Percentage distribution of the frequencies of the eye trackers used in the analyzed papers.](image)

The frequencies used only by one paper were (in Hz) 4.5, 5, 10, 15, 17, 28, 50, 150, 176, 240, 256, 3000 and 8000.

There is no clear consensus between the researchers about the proper eye tracking sampling frequency, but there is a tendency to use higher frequencies (above 200 Hz) when using velocity-based event detection algorithms [100]. In terms of detecting saccades and fixations’ sampling frequency, a change from 60 Hz to 120 Hz does not seem to provide significant improvement in the fixations’ detection rate [101], but it is important when evaluating saccades [96]. For this exact purpose frequencies lower than 200 Hz are discouraged in the case of saccades’ speed studies [102]. Overall, since fixations take less time, they require smaller frequencies than saccades and microsaccades [103]. That is why low-level research connected with visual cognition usually requires frequencies of 1000 Hz to 2000 Hz [104].

In terms of this survey, we can observe a tendency to use frequencies of 30 Hz and 60 Hz. The 30 Hz frequency gained its popularity as it was used as an the American television standard NTSC, whereas 60 Hz was commonly used in cameras. The third sampling rate, in terms of the number of studies which used it, is 120 Hz, and apart from that we can clearly see that, similar to the case of eye trackers, there is no tendency to use one particular frequency. There are also no justification for the selected frequency, at most researchers include the justification that it is a frequency sufficient for the conducted research. This indicates that scientists are using the highest sampling rate of the eye tracker...
at the disposal of the researchers. The use of lower frequencies occurs mainly when there is a need to synchronize an eye tracker with another sensor which has a lower sampling rate. There is no clear reason for the highest sampling frequencies. They were used for detecting the speed of reading (8000 Hz), detecting people with dyslexia (3000 Hz), predicting web user click intention (1000 Hz), using eye-tracking data as an input for teaching AI to play computer games in a similar way to humans (1000 Hz), detecting cognitive health (500 Hz), predicting intention (500 Hz) and detecting reading abilities (500 Hz). We can say that they were used for research connected with cognition and mental health, so, psychological studies.

In terms of the lowest sampling frequencies, most of those under 30 Hz were used when a web or mobile camera was chosen as an eye tracker (4), some used Tobii hardware (3) and one used HTC Vive. They were usually used for tasks related to detecting predefined types of behavior: predicting targets (17 Hz), detecting eye contact (25 Hz), task recognition (25 Hz) and behavior identification (28 Hz). They were also used for attention estimation (5 Hz) and identifying levels of user comprehension (15 Hz). The study with the lowest sampling frequency (4.5 Hz) used higher sampling frequencies as well, but since it was conducted by the participants themselves, using their own hardware, such frequencies were the lowest used but not the only ones. The aim of that study was gaze estimation.

In terms of illumination during an eye tracking experiment, 25.8% of the papers included some information about lighting conditions but almost all of them would not be sufficient to conduct an experiment in similar conditions. They only mention that they kept the illumination constant or similar throughout the experiment. Only one paper included results for different illuminations. When considering only experiments conducted using cameras and not eye trackers, the percentage of papers including information on lighting conditions is 57.89% which is considerably higher than the overall percentage. This is understandable, since cameras are more sensitive to changes in lighting than eye trackers, however, it would be advisable to make these data more accurate and appear more frequently in papers.

2.3. Participants

Overall, 49.25% of the participants of the described studies were men and 50.75% were women. However, when we look at the average proportions per study there are usually 55.91% men and 44.09% women. Sadly, not all researchers specified the gender ratio of their participants, so the figures given are based only on the studies that have done so.

The size of the research group varies greatly, with the smallest consisting of only one person and the largest having 2334 participants. As with eye trackers and sampling rates, there is no clear trend here. Researchers usually chose a group of 17 to 33 people, and it can be theorized that, again, this is simply the smallest group which allows for obtaining statistically significant results.

Since gathering participants may be the one of the biggest challenges of eye tracking studies, it may be surprising that only 5.38% of the analyzed papers used databases created by other researchers, but this may be explained by the fact that such databases are few and they may not be sufficient for very specific applications (Figure 3).

As much as 31.18% of the papers have not given additional information about the participants. A total of 22.58% clearly defined the participants of their study as students and 8.6% have stated an age range which strongly suggests that their participants were also students. Finally, 4.3% of the papers described their participants as children. Clearly, the biggest issue is these papers not giving proper information about their participants. However, based on the available data, we can infer that most of the research is carried out on young people, in particular students, and adults and the elderly are not adequately represented in eye tracking research.

Other information which the papers were usually lacking was the participants’ vision. Only 22.58% included such information, but in most cases (42.86% of the papers with information about the participant’s vision) it stated that participants had normal or
corrected-to-normal vision without giving accurate data on the proportion and method of correction. A total of 23.81% of papers included information about the number of participants wearing glasses, and the exact same amount of papers included only people with correct vision. Only one paper conducted an experiment with participants both wearing glasses and not wearing them, and one stated that participants were not asked to remove their glasses during the experiment which suggests that there were participants wearing glasses in that study.

![Figure 3. Number of participants in the studies.](image)

Additionally, it is worth noting that only 17.2% of papers contained information about the approval of the research by an ethics committee. One paper took Google’s AI Principles into consideration when designing its experiment but has not included any information about the approval of an ethics committee.

2.4. Additional Data for Artificial Intelligence

Additional parameters were used for artificial intelligence by 23.65% of the analyzed studies. The most commonly used data were obtained with electroencephalography (EEG), which is a non-invasive method of recording electrical activity on the scalp which is used to determine the activity of the brain. Since many experiments which have used eye tracking data are connected with cognition this is understandable. Equally popular is movement and position data. Position data might be especially important since it may influence eye tracking data. Another parameter which can be used is a video of the face, which allows researchers to estimate the emotions of study participants. An equally common parameter is the time which the participant needed to perform the task under consideration. Lastly, data about the study subjects were used. Four studies used their age and three used their gender.

Since the area of eye tracking research is quite wide, the parameters that may be considered are also quite diverse and sometimes really study-specific, like, for example, data describing studied texts, images or videos. All of the additional parameters used in the remaining papers are included in Table 1.
### Table 1. Additional data used for teaching in artificial intelligence algorithms.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Task</th>
<th>Additional Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>[64]</td>
<td>attention estimation</td>
<td>EEG, head movement</td>
</tr>
<tr>
<td>[38]</td>
<td>identifying children with ASD</td>
<td>questionnaire, age, gender</td>
</tr>
<tr>
<td>[56]</td>
<td>predicting dwell time in a museum</td>
<td>face expression, body movement, interaction trace logs</td>
</tr>
<tr>
<td>[27]</td>
<td>affect recognition</td>
<td>EEG, ECG</td>
</tr>
<tr>
<td>[8]</td>
<td>predicting students’ performance and effort</td>
<td>EEG, face videos, arousal data from wristband</td>
</tr>
<tr>
<td>[71]</td>
<td>predicting take-over time</td>
<td>head position, body posture, simulation data</td>
</tr>
<tr>
<td>[30]</td>
<td>predicting liking a video</td>
<td>infrared thermal image, heart rate, face expression</td>
</tr>
<tr>
<td>[32]</td>
<td>predicting user confidence</td>
<td>Time</td>
</tr>
<tr>
<td>[25]</td>
<td>predicting reaction to ads</td>
<td>gender, age, survey, time, ad parameters, behavior connected with an ad (e.g., sharing)</td>
</tr>
<tr>
<td>[13]</td>
<td>predicting readability</td>
<td>text features</td>
</tr>
<tr>
<td>[17]</td>
<td>predicting SAT score</td>
<td>Time</td>
</tr>
<tr>
<td>[36]</td>
<td>predicting the emotion of an observed person</td>
<td>EEG, empatica bracelet</td>
</tr>
<tr>
<td>[32]</td>
<td>predicting social plane of interaction</td>
<td>EEG, accelerometer, audio, video</td>
</tr>
<tr>
<td>[33]</td>
<td>detecting user confusion</td>
<td>mouse actions, distance of the user’s head from the screen</td>
</tr>
<tr>
<td>[72]</td>
<td>predicting mental workload</td>
<td>Reaction time</td>
</tr>
<tr>
<td>[42]</td>
<td>detecting people with dyslexia</td>
<td>age, text characteristics</td>
</tr>
<tr>
<td>[74]</td>
<td>predicting reduced driver alertness</td>
<td>EEG</td>
</tr>
<tr>
<td>[19]</td>
<td>predicting learning curve</td>
<td>perceptual speed, verbal working memory, visual working memory, locus of control</td>
</tr>
<tr>
<td>[37]</td>
<td>classifying emotions in pictures</td>
<td>image</td>
</tr>
<tr>
<td>[89]</td>
<td>predicting eye movement</td>
<td>distance between the object and the distractor</td>
</tr>
<tr>
<td>[41]</td>
<td>predicting Parkinson symptoms’ development</td>
<td>age, sex, duration of the disease</td>
</tr>
<tr>
<td>[23]</td>
<td>emotion estimation</td>
<td>head movement, body movement, audio, video of the face</td>
</tr>
</tbody>
</table>

#### 2.5. Features Extraction

Feature selection usually begins every application of AI methods. The application of a proper feature selection method has a very large impact on the obtained results of the AI algorithms, no matter the area of application. This is of even greater importance when the AI algorithms have to deal with large amounts of data, as is usually the case in image processing applications. This is also the case in eye tracking. One of the main features of all eye trackers is the frequency of gathering data on the participant eye focus coordinates as well as eye blinks and pupil size, which can vary based on changes in lighting and the mental state of the participant [105]. The eye tracker frequency may vary from 15 Hz to 1000 Hz. With increasing frequency, the amount of data increases, so in some cases we need to select some features which aggregate the raw data, such as dwell times on AOI or heat maps [12].

Feature extraction is also crucial when working with visual imagery. Simplification may include scaling down images, converting them to grey-scale and using Principal Component Analysis. Such an approach can transform sets with thousands of features to
several dozen components ready for further analysis [106]. This method may be further improved by the autoencoding technique, which efficiently reduces dimensionality and extracts meaningful features from eye-tracking data [107].

In this paragraph we have presented the categorization of the feature selection types which are present in the analyzed literature. We propose the following categories:

A. Typical eye tracking data which are gathered via typical eye tracker software, such as Tobii Studio or Tobii Pro Lab [104];
B. Eye tracking data after some additional processing, which are not present in typical eye tracker software, such as more sophisticated statistics or transformations (i.e., DWT);
C. The application of some basic ML algorithms to eye tracker data such as k-means or decision trees;
D. The application of neural networks or deep learning;
N. Not specified in the article.

The figure below presents the percentages of each of the feature extraction types used in the analyzed literature (Figure 4).

2.6. Artificial Intelligence Methods Used with Eye Tracking

Most of the surveyed studies used only one artificial intelligence but as many as 37.6 percent decided to compare at least two different AIs. It is worth noting, however, that in many works using only one AI, the researchers compared results using different parameters (Figure 5).

Testing different methods of artificial intelligence is very beneficial because often there is no clear rationale for using one particular solution instead of another.

This is especially true when we look at the type of methods used. As many as 40.9 percent of the works used AIs that were not used in any of the other considered works, which clearly shows the variety of available solutions. It is also difficult to observe tendencies to use specific algorithms in specific areas of research, apart from the most popular AI seeming to be the support-vector machine (SVM), which is commonly used for data classification, especially in the field of image recognition. The second choice for researchers was Random Forest, which creates a multitude of decision trees and calculates the result based on the predictions of the individual trees. The third method was a convolutional neural network (CNN), which is commonly used in image recognition. It is worth noting...
that the multilayer perceptron (MLP) and CNN are both types of artificial neural networks (ANN), which, combined, were used in 34.5% of the considered papers, but since the researchers decided to list them separately this separation was kept in this survey.

![Figure 5. Number of artificial intelligence methods used in each paper.](image)

By analyzing the use of SVMs and neural networks (all types), changes in their popularity over the years can be noticed. In 2015, the SVM was used in 65% of surveyed publications, whereas 29% has chosen neural networks. From that moment on, SVMs began to be used less and less, and neural networks more and more often. In 2020, SVMs would be used in 22% of the analyzed studies, and neural networks in 56%. Neural networks are becoming an increasingly popular method of artificial intelligence in many fields, and it is not surprising that this is the case for eye tracking applications. There are more and more ready-made solutions that allow scientists to use this method, even without detailed knowledge of it, and, what is more, they allow the use of incomplete data. Much smaller changes in use occurred in the case of Random Forest, which was used in 30% of publications in 2016, and then decreased its share to 22% in 2020. The continuing popularity of Random Forest is probably due to the ease of using this method while it provides fairly good results (Figure 6).

The artificial intelligence methods which were used by only one paper include a bag of visual words, Bayes net, Bayesian classifier, Bayesian lasso regression, boosted logistic regression, canopy, CNN + long short-term memory, decomposition tree, Deep Bayesian Network, discriminant analysis, DNN, double q-learning, extremely randomized trees, farthest first, generalized additive models, generative model base method, gradient boost, hidden-state conditional random fields, hierarchical clustering, lasso regression, least-squares regression, low-rank constraint, Mahalanobis distance-based classifier, mixed group ranks, multi-layer combinatorial fusion, multinomial logistic regression, radial basis function, random sample consensus, recurrent neural network, recurrent neural networks with long short-term memory, repeated incremental pruning to produce error reduction, semi-supervised extreme learning machine, sequential minimal optimization, Static Bayesian Network with supervised clustering, a strengthened deep belief network, Tabu search, transfer learning and a Viola-Jones algorithm with haar cascade classifiers.
In terms of the best results, we can see a similar tendency. Mostly the best results were achieved by AIs which were used in only one paper, then SVMs, followed by Random Forest and convolutional neural networks. If a study used only one AI it was considered to be the best (Figure 7).
The artificial intelligence methods which were chosen by only one paper include a Gaussian process regression, Bayesian lasso regression, long short-term memory network, generative model base method, DNN, transfer learning, Deep Bayesian Network, Tabu search, decision tree, low-rank constraint, linear discriminant analysis, ensemble, lasso regression, naive Bayes, a strengthened deep belief network, semi-supervised extreme learning machine, Support vector regression, Decomposition tree, recurrent neural network, CNN + long short-term memory, Viola-Jones algorithm with haar cascade classifiers, recurrent neural networks with long short-term memory, random sample consensus and linear regression.

2.7. Methods for Verification of the Results

Almost three quarters of the researchers used only one method of verification for their results, while the rest used from two to five methods (Figure 8).

Most of the studies used accuracy as a result of their study, but, secondarily, there are methods used by only one paper which are often study-specific. It makes it extremely difficult to compare different results so a comparison will be made between only the results which were verified using the accuracy value. Apart from that, the next three most popular indicators were precision, recall and f-score (also called f1-score). Including accuracy, they are all based on components of the confusion matrix: the true positive (TP), false positive (FP), false negative (FN) and true negative (TN). Their formulas are as follows [108]:

\[
\text{accuracy} = \frac{TP + TN}{TP + FP + FN + TN}
\]

\[
\text{precision} = \frac{TP}{TP + FN}
\]

\[
\text{recall} = \frac{TP}{TP + FN}
\]

\[
\text{f-score} = \frac{2}{\text{precision} + \text{recall}}
\]

All papers used those formulas or did not specify which ones they had used, possibly considering them universally accepted (Figure 9).
Most of the studies used accuracy as a result of their study, but, secondarily, there are methods used by only one paper which are often study-specific. It makes it extremely difficult to compare different results so a comparison will be made between only the results which were verified using the accuracy value. Apart from that, the next three most popular indicators were precision, recall and f-score (also called f1-score). Including accuracy, they are all based on components of the confusion matrix: the true positive (TP), false positive (FP), false negative (FN) and true negative (TN). Their formulas are as follows:

- Accuracy: \( \frac{TP + TN}{TP + FP + FN + TN} \)
- Precision: \( \frac{TP}{TP + FN} \)
- Recall: \( \frac{TP}{TP + FN} \)
- F-score: \( \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \)

All papers used those formulas or did not specify which ones they had used, possibly considering them universally accepted (Figure 9).

The verification methods used by only one paper include accuracy in degrees, angular error, average distance error, average error, average hit ratio, average success rate, average visual angle error, confidence, cross-validation error, D error, discounted cumulative gain, equal error rate, error, error in centimeters, false positive rate, G-mean, gaze estimation bias (degrees), improvement in game score, Mann–Whitney u-value, mean absolute residual, mean and standard deviation of fp and fn, mean angular error, mean squared error, overall average error, R-squared, relative difference to baseline, reliability, root mean squared error, percent of screen size error, sensitivity index, specificity, support and visual angle.

3. Results

In the event that no single result was given, the best result obtained in a given publication is given. Some of the papers have not specified the number of participants (ns) and some used external data sources (ext.) (Table 2).

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Task</th>
<th>AI</th>
<th>N</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[47]</td>
<td>detecting the type of behavior when using laptop</td>
<td>SVM</td>
<td>ns</td>
<td>99.77%</td>
</tr>
<tr>
<td>[67]</td>
<td>detecting driver distraction</td>
<td>Semi-Supervised Extreme Learning Machine</td>
<td>34</td>
<td>97.2%</td>
</tr>
<tr>
<td>[95]</td>
<td>gaze estimation</td>
<td>Random Forest</td>
<td>10</td>
<td>97.2%</td>
</tr>
<tr>
<td>[37]</td>
<td>classifying emotions in pictures</td>
<td>Strengthened Deep Belief Network</td>
<td>40</td>
<td>97.1%</td>
</tr>
<tr>
<td>[45]</td>
<td>predicting neurological diseases</td>
<td>Random Forest</td>
<td>96</td>
<td>96.88%</td>
</tr>
<tr>
<td>[12]</td>
<td>predicting type of reading</td>
<td>SVM</td>
<td>30</td>
<td>96.69%</td>
</tr>
<tr>
<td>[81]</td>
<td>detecting eye gestures</td>
<td>naive Bayes</td>
<td>ext</td>
<td>95.0%</td>
</tr>
<tr>
<td>[11]</td>
<td>reading behavior recognition</td>
<td>Hidden Markov Model</td>
<td>4</td>
<td>95.0%</td>
</tr>
<tr>
<td>[92]</td>
<td>gaze estimation</td>
<td>ANN</td>
<td>29</td>
<td>94.1%</td>
</tr>
<tr>
<td>[61]</td>
<td>predicting targets</td>
<td>MLP</td>
<td>5</td>
<td>94.0%</td>
</tr>
<tr>
<td>[48]</td>
<td>detecting computer activity type</td>
<td>CNN</td>
<td>150</td>
<td>93.15%</td>
</tr>
<tr>
<td>Ref.</td>
<td>Task</td>
<td>AI</td>
<td>N</td>
<td>Accuracy</td>
</tr>
<tr>
<td>------</td>
<td>-------------------------------------</td>
<td>-----------------------------------------</td>
<td>----</td>
<td>------------</td>
</tr>
<tr>
<td>[76]</td>
<td>authentication</td>
<td>CNN</td>
<td>26</td>
<td>93.14%</td>
</tr>
<tr>
<td>[63]</td>
<td>detecting attention</td>
<td>SVM</td>
<td>10</td>
<td>93.1%</td>
</tr>
<tr>
<td>[80]</td>
<td>detecting eye gestures</td>
<td>SVM</td>
<td>5</td>
<td>93.0%</td>
</tr>
<tr>
<td>[46]</td>
<td>detecting organs</td>
<td>AdaBoost</td>
<td>10</td>
<td>92.5%</td>
</tr>
<tr>
<td>[29]</td>
<td>predicting excitement</td>
<td>DNN</td>
<td>20</td>
<td>92.0%</td>
</tr>
<tr>
<td>[75]</td>
<td>smartphone authentication</td>
<td>Random Sample Consensus</td>
<td>21</td>
<td>91.6%</td>
</tr>
<tr>
<td>[82]</td>
<td>eye gestures for patients</td>
<td>Recurrent Neural Network</td>
<td>270</td>
<td>91.4%</td>
</tr>
<tr>
<td>[69]</td>
<td>detecting mental fatigue</td>
<td>SVM</td>
<td>18</td>
<td>91.0%</td>
</tr>
<tr>
<td>[54]</td>
<td>predicting ethical decision-making</td>
<td>MLP</td>
<td>75</td>
<td>90.7%</td>
</tr>
<tr>
<td>[22]</td>
<td>predicting social plane of interaction</td>
<td>Gradient Boosted Decision Tree</td>
<td>1</td>
<td>90.6%</td>
</tr>
<tr>
<td>[59]</td>
<td>predicting movement intention</td>
<td>CNN</td>
<td>24</td>
<td>88.37%</td>
</tr>
<tr>
<td>[32]</td>
<td>predicting user confidence</td>
<td>Random Forest</td>
<td>23</td>
<td>88.0%</td>
</tr>
<tr>
<td>[73]</td>
<td>detecting operator overload</td>
<td>Linear Discriminant Analysis</td>
<td>20</td>
<td>87.91%</td>
</tr>
<tr>
<td>[39]</td>
<td>detecting people with ASD</td>
<td>SVM</td>
<td>130</td>
<td>86.89%</td>
</tr>
<tr>
<td>[13]</td>
<td>predicting readability</td>
<td>MLP</td>
<td>ext</td>
<td>86.62%</td>
</tr>
<tr>
<td>[50]</td>
<td>identifying children’s behavior</td>
<td>Random forest</td>
<td>32</td>
<td>84.0%</td>
</tr>
<tr>
<td>[38]</td>
<td>distinguishing children with ASD</td>
<td>Logistic Regression</td>
<td>33</td>
<td>83.9%</td>
</tr>
<tr>
<td>[62]</td>
<td>predicting web user click intention</td>
<td>ANN</td>
<td>25</td>
<td>82.0%</td>
</tr>
<tr>
<td>[83]</td>
<td>choosing direction</td>
<td>Viola-Jones Algorithm with HAAR Cascade Classifiers</td>
<td>ns</td>
<td>82.0%</td>
</tr>
<tr>
<td>[30]</td>
<td>predicting liking a video</td>
<td>ANN</td>
<td>30</td>
<td>81.8%</td>
</tr>
<tr>
<td>[35]</td>
<td>detecting satisfaction</td>
<td>SVM</td>
<td>30</td>
<td>80.53%</td>
</tr>
<tr>
<td>[42]</td>
<td>detecting people with dyslexia</td>
<td>SVM</td>
<td>97</td>
<td>80.18%</td>
</tr>
<tr>
<td>[20]</td>
<td>detecting speed of learning</td>
<td>Random Forest</td>
<td>161</td>
<td>80.0%</td>
</tr>
<tr>
<td>[99]</td>
<td>distinguishing Chinese ethnic groups</td>
<td>SVM</td>
<td>35</td>
<td>80.0%</td>
</tr>
<tr>
<td>[58]</td>
<td>detecting eye contact</td>
<td>SVM</td>
<td>ns</td>
<td>80.0%</td>
</tr>
<tr>
<td>[41]</td>
<td>predicting Parkinson symptoms'</td>
<td>Decomposition Tree</td>
<td>10</td>
<td>79.5%</td>
</tr>
<tr>
<td></td>
<td>development</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[40]</td>
<td>detection of Schizophrenia</td>
<td>Generative Model Base Method</td>
<td>44</td>
<td>79.2%</td>
</tr>
<tr>
<td>[70]</td>
<td>detecting task demand</td>
<td>Random Forest</td>
<td>48</td>
<td>79.0%</td>
</tr>
<tr>
<td>[72]</td>
<td>predicting mental workload</td>
<td>Ensemble</td>
<td>20</td>
<td>78.0%</td>
</tr>
<tr>
<td>[19]</td>
<td>predicting learning curve</td>
<td>Random Forest</td>
<td>95</td>
<td>77.0%</td>
</tr>
<tr>
<td>[17]</td>
<td>predicting SAT score</td>
<td>Decision Tree</td>
<td>30</td>
<td>76.67%</td>
</tr>
<tr>
<td>[15]</td>
<td>predicting word understanding</td>
<td>SVM</td>
<td>16</td>
<td>75.6%</td>
</tr>
<tr>
<td>[68]</td>
<td>detecting mind wandering</td>
<td>SVM</td>
<td>178</td>
<td>74.0%</td>
</tr>
<tr>
<td>[65]</td>
<td>detecting cognitive distraction</td>
<td>SVM</td>
<td>18</td>
<td>73.0%</td>
</tr>
<tr>
<td>[27]</td>
<td>affect recognition</td>
<td>Long Short-Term Memory Network</td>
<td>130</td>
<td>72.8%</td>
</tr>
<tr>
<td>[57]</td>
<td>automatic surgery skills assessment</td>
<td>Random Forest</td>
<td>9</td>
<td>69.0%</td>
</tr>
<tr>
<td>[96]</td>
<td>gaze estimation</td>
<td>ANN</td>
<td>10</td>
<td>68.31%</td>
</tr>
</tbody>
</table>
Table 2. Cont.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Task</th>
<th>AI</th>
<th>N</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[18]</td>
<td>predicting user’s cognitive abilities</td>
<td>Random forest</td>
<td>166</td>
<td>66.1%</td>
</tr>
<tr>
<td>[60]</td>
<td>predicting intention</td>
<td>SVM</td>
<td>20</td>
<td>64.0%</td>
</tr>
<tr>
<td>[9]</td>
<td>predicting student’s performance</td>
<td>Logistic Regression</td>
<td>95</td>
<td>63.0%</td>
</tr>
<tr>
<td>[32]</td>
<td>detecting user confusion</td>
<td>Random Forest</td>
<td>136</td>
<td>61.0%</td>
</tr>
<tr>
<td>[49]</td>
<td>predicting type of task</td>
<td>Hidden Markov Model</td>
<td>8</td>
<td>57.0%</td>
</tr>
<tr>
<td>[21]</td>
<td>predicting the prior disclosure type from eye data</td>
<td>Gradient Boosted Decision Tree</td>
<td>20</td>
<td>53.5%</td>
</tr>
<tr>
<td>[88]</td>
<td>predicting the duration of gaze fixation</td>
<td>SVM Ext.</td>
<td>49.1%</td>
<td></td>
</tr>
<tr>
<td>[36]</td>
<td>predicting the emotion of an observed person</td>
<td>MLP</td>
<td>44</td>
<td>42.7%</td>
</tr>
</tbody>
</table>

When looking at the results we can see that the SVM gives better results than Random Forest, but it is important to note that it is possible that this was caused by the fact that SVM was used more often. However, neural networks seem to give consistently higher results than Random Forest, even though Random Forest gave two results higher than neural networks were capable of giving in this survey.

There is no clear correlation between the number of participants and the resulting accuracy of the AI. A larger number of subjects might lead to having too diverse a dataset, which may make the prediction more challenging, but on the other hand, having a smaller number of participants may lead to a dataset not diverse enough to properly predict the desired parameters.

The papers which used more than one artificial intelligence method produced slightly better results. It might be also worth noticing that the half of the papers with the higher accuracies did not use sampling frequencies higher than 256 or lower than 30. Sadly, some of the frequencies were not specified. Furthermore, of the top 20 results, only one used an additional parameter, which may suggest that studies with information clearly associated with eyeball movement give the best results, and that combining eye tracking data with other kinds of data is not always the best choice.

There seems to be no correlation between the result and the type of eye tracker, especially since the researchers used many different devices.

4. Discussion

Eye tracking and artificial intelligence appeared in a variety of applications connected to measuring academic performance, emotion recognition, medical studies, human behavior and tiredness detection. This combination also made it possible to use eye movement as an input and to track it using digital, web and mobile cameras.

The choice of the eye tracker, sampling frequency, artificial intelligence algorithm and verification method is characterized by a huge variety, which may result from the various fields of application. Many researchers decide to use unique solutions that do not appear in other works, which may indicate that this remains a new field of research, which is just developing research standards.

There are some noticeable trends though. The most popular eye trackers were made by Tobii and most common sampling frequencies were 30 and 60 Hz. The most popular artificial intelligence methods used for eye tracking data analysis were SVM and Random Forest, while the results were most often judged on the basis of their accuracy. Unfortunately, information about the illumination during the experiments is usually lacking.

The research was carried out on groups of various sizes, most often from 17 to 33 people, with a relatively equal gender division. On the other hand, almost one quarter of experiments were conducted on students, while adults and the elderly were usually
underrepresented. Sadly, information about the participant’s eye-sight was usually not included and even if it was it was not detailed enough to replicate the study. Another issue was the lack of clear information about the approval of an ethics committee.

One clear parameter allowing for the higher accuracy of artificial intelligence using eye tracking data has not been found. Recommendations could include the use of a sampling rate between 30 and 265 Hz and the use of more than one artificial intelligence method.

Clearly, eye tracking data analysis, especially with the support of artificial intelligence, can teach us a lot about human nature, and there is still a lot to discover. The fields in which this technology can be used are very diverse, which on the one hand makes it difficult to compare the results, but on the other hand shows the great possibilities of its use. The accuracy of some solutions still leaves room for improvement, but they still show various correlations between human behavior and emotions, due to which they can act as a clue for researchers in the fields of psychology, medicine, didactics, etc., when choosing the subjects of their research.

Future research directions offer a wealth of opportunities, including expanding this technology’s applications across diverse demographic groups, enhancing multimodal integrations, and exploring novel clinical, educational, and cross-cultural domains. Exploring education and the detection of psychological disorders holds great promise as a starting point for future research. In education, the integration of AI-enhanced eye tracking can transform teaching and learning methods, while in the realm of psychology, it offers potential for early diagnosis and interventions, including interventions in real time during tasks like driving. These research directions are poised to yield practical, impactful solutions. To further advance the field, interdisciplinary collaborations could foster the development of holistic solutions that draw from various domains and have broader societal impacts. Additionally, cross-cultural validation is essential, particularly in emotion recognition and behavior prediction, to ensure the cultural sensitivity and accuracy of AI models. Real-time interventions based on gaze behavior in educational settings hold potential for enhancing learning outcomes.

There are, however, some areas which should be improved in future studies. In some papers, vital information about the study participants is often missing, leaving a significant gap in our understanding. This lack of information spans from the basic number of participants to more detailed aspects like gender distribution, age ranges, and eye sight parameters (such as the use of glasses). Such information should be included by future researchers to ensure that their studies are more comprehensive and ultimately more impactful. Perhaps the most crucial improvement in terms of participants’ information is the need for the inclusion of documented approval from ethics committees.

A promising avenue for collaborative progress could revolve around the establishment of comprehensive eye tracking databases, allowing researchers from diverse backgrounds to access and analyze this valuable resource. It is noteworthy that a mere 5.33% of papers in our study drew upon external data sources, signifying a missed opportunity for the broader scientific community to benefit from shared datasets and foster collective advancements in eye tracking research. Incorporating additional parameters, such as electroencephalography (EEG), participant position, and movement data, as well as experiment-related factors like duration and stimulus parameters, holds the potential to enrich the depth and breadth of eye tracking studies. These additional dimensions offer a holistic perspective on the cognitive and contextual aspects influencing visual attention, paving the way for more comprehensive and nuanced findings in the field of eye tracking research. Researchers should consider these multifaceted variables as valuable assets in their quest to unravel the complexities of visual perception and cognition.

Furthermore, fostering the use and comparison of multiple AI methods within research endeavors is poised to substantially elevate the quality and rigor of eye tracking studies. The absence of a clear rationale for choosing one AI solution over another underscores the need for comprehensive comparisons. By exploring and evaluating various AI techniques,
researchers can identify the most effective solutions tailored to the specific challenges they aim to address.

Lastly, a critical advancement required in the field of eye tracking research lies in the inclusion of standardized and diverse methods for analyzing result verifications. While accuracy emerges as the most frequently employed parameter, it is crucial to consider its future integration with additional performance metrics like precision, recall, and F-score. These established metrics provide a more comprehensive understanding of the true accuracy of AI models. Furthermore, the research community may find it advantageous to explore or develop novel verification methods tailored specifically to the nuances of AI and eye tracking. This approach not only enhances the quality and precision of eye tracking research but also facilitates the comparability of results across different papers. By establishing standardized verification methods and broadening the spectrum of the metrics employed, researchers can effectively compare and benchmark their findings with those from other studies.

5. Conclusions

Eye tracking data analysis, particularly when combined with artificial intelligence, offers valuable insights into human behavior and emotions. Its versatile applications make result comparison challenging but highlight its immense potential. While their accuracy can still be improved, these solutions provide valuable insights for researchers in psychology, medicine, education, and other fields, when selecting their research subjects.


Funding: This research received no external funding.

Data Availability Statement: The data is contained in the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References
6. Sharma, K.; Giannakos, M.; Dillenbourg, P. Eye-tracking and artificial intelligence to enhance motivation and learning. Smart Learn Environ. 2020, 7, 13. [CrossRef]
58. Pettersson, J.; Falkman, P. Human Movement Direction Classification using Virtual Reality and Eye Tracking. Procedia Manuf. 2020, 51, 95–102. [CrossRef]
73. Larue, G.S.; Rakotonirainy, A.; Petitt, A.N. Predicting Reduced Driver Alertness on Monotonous Highways. IEEE Pervasive Comput. 2015, 14, 78–85. [CrossRef]
75. Tiwari, A.; Pal, R. Gaze-Based Graphical Password Using Webcam. In Information Systems Security; Ganapathy, V., Jaeger, T., Shyamasundar, R., Eds.; ICIS 2018. Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 11281. [CrossRef]


100. Leube, A.; Rifai, K.; Rifai, K. Sampling rate influences saccade detection in mobile eye tracking of a reading task. *J. Eye Mov. Res. 2017*, 10. [CrossRef] [PubMed]


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.