



UNIVERSITY OF PAVIA  
FACULTY OF ENGINEERING  
DEPARTMENT OF INDUSTRIAL ENGINEERING AND INFORMATION

MASTER'S DEGREE IN COMPUTER ENGINEERING

MASTER THESIS

# Gaze-Based Biometric Authentication Through Dynamic Text Reading

**Candidate:** Davide Mascheroni

**Supervisor:** Prof. Marco Porta

**Co-supervisor:** Piercarlo Dondi and Hoàng Nam Lê

**Academic Year:** 2024/2025

# Table of contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Behavioral Biometric</b>	<b>5</b>
2.1	Identification vs Verification . . . . .	5
2.2	Biometric based on identification . . . . .	5
2.3	Behavioral biometric . . . . .	6
2.4	Some applications of behavioral biometric . . . . .	7
2.5	Soft Biometric . . . . .	9
<b>3</b>	<b>Eye tracker</b>	<b>10</b>
3.1	Visual attention . . . . .	10
3.1.1	Von Helmholtz’s “Where” . . . . .	10
3.1.2	James’ “What” . . . . .	10
3.1.3	Gibson’s “How” . . . . .	10
3.1.4	Broadbent’s “Selective Filter” . . . . .	10
3.1.5	Deutsch and Deutsch’s “Importance Weightings” . . . . .	11
3.1.6	Yarbus and Noton and Stark’s “Scanpaths” . . . . .	11
3.1.7	Treisman’s “Glue” . . . . .	11
3.1.8	Kosslyn’s “Window” . . . . .	11
3.1.9	Bottom-up model . . . . .	11
3.2	The human eye . . . . .	12
3.2.1	Anatomy of human the eye . . . . .	12
3.2.2	The retina and photoreceptor . . . . .	12
3.2.3	Color perception . . . . .	13
3.2.4	Type of eye movements . . . . .	14
3.2.5	Fixation . . . . .	14
3.3	Eye tracking techniques . . . . .	15
3.3.1	Electro-OculoGraphy (EOG) . . . . .	15
3.3.2	Scleral Contact Lens/Search Coil . . . . .	15
3.3.3	Photo-OculoGraphy (POG) or Video-OculoGraphy (VOG) . . . . .	16
3.3.4	Video-Based Combined Pupil/Corneal Reflection . . . . .	16
<b>4</b>	<b>Related work</b>	<b>18</b>
4.1	ATM-like solutions . . . . .	18
4.2	Methods based on fixation and scanpath analysis . . . . .	20
4.3	Methods based on eye/gaze velocity . . . . .	22
4.4	Methods based on pupil size . . . . .	24
4.5	Methods based on oculomotor features . . . . .	24
4.6	Methods based on head orientation . . . . .	26
<b>5</b>	<b>Experiment on Gaze-based Biometric with dynamic text</b>	<b>28</b>
5.1	Set up . . . . .	28
5.2	Stimuli . . . . .	29
5.2.1	Preliminary experiments . . . . .	30
5.2.2	Animations . . . . .	33
5.3	Captured Eye Tracking Features . . . . .	35
5.4	Dataset creation . . . . .	37
5.4.1	Fixation-Based Features (f0–f10) . . . . .	37
5.4.2	Fixation Distance Features (f11–f20) . . . . .	37
5.4.3	Saccade Speed Features (f21–f30) . . . . .	38
5.4.4	Scanpath Feature (f31) . . . . .	39
5.4.5	Pupil Diameter Features (f32–f51) . . . . .	39
5.4.6	Pupil Diameter Ratio Features (f52–f61) . . . . .	40
5.4.7	Pupil Diameter Difference Features (f62–f70) . . . . .	41
5.4.8	Blink Features (f71–f74) . . . . .	41

5.4.9	Directional Movement Features (f75–f82)	42
5.5	Machine Learning Analysis	43
5.5.1	Identification without feature selection	44
5.5.2	Identification with feature selection	46
5.5.3	Identification with single animation - single train(ST) type	47
5.6	Static vs Dinamic text	49
<b>6</b>	<b>Appendix</b>	<b>49</b>
6.1	Parameter grid	49
6.2	Identification with feature selection - Best k features	50
<b>7</b>	<b>Conclusion</b>	<b>63</b>
<b>8</b>	<b>Acknowledgment</b>	<b>63</b>
<b>9</b>	<b>References</b>	<b>64</b>

## Abstract

# 1 Introduction

In recent years, new information technologies have become increasingly widespread in everyday life. Initially, the most prominent example was the personal computer, which revolutionized the way individuals work, communicate, and access information. However, technological development has progressed rapidly, and today there is a wide variety of smart devices available on the market. Such technologies encompass tablets, wearable devices, and home automation systems, with mobile phones representing the most prominent example. According to Statista [1], in 2024 more than 7.2 billion people worldwide own a smartphone, making it one of the most ubiquitous devices in human history. The growing reliance on such devices has brought tremendous convenience but also introduced significant security challenges. Mobile devices store and process sensitive personal information, such as banking details, private communications, and location history. This makes them prime targets for cyberattacks, and without adequate protection, they could be exploited for identity theft, financial fraud, and unauthorized access to personal data. To address these risks, multiple layers of security have been developed for mobile devices, including passwords, PINs, fingerprint scans, and facial recognition. However, these traditional methods often have limitations, such as being susceptible to theft, spoofing, or user inconvenience. This thesis focuses on a more advanced and promising approach: behavioral biometrics. Behavioral biometrics examine the distinctive ways in which users interact with a device, enabling continuous identity verification. This approach provides user authentication in a less intrusive way, but unfortunately it doesn't reach the accuracy of the traditional biometric approach. For this reason behavioral biometric is rarely used as the main approach for user identification, but instead represents an additional layer of security to build on top of the traditional approaches.

## 2 Behavioral Biometric

As per Jim Holdsworth et al. [88] Behavioral biometrics is a form of authentication that analyzes the unique patterns in a user's activity to verify their identity. In particular this discipline is able to collect, in a non intrusive way, a distinctive way in which a user achieve a task. As we said earlier this technique achieve low performance to be used as the only authentication method in a system, so it is used as a complement of more secure solutions. This additional level of security will focus only on a verification purpose to detect possible attacks or fraud. A major advantage over traditional PIN or password-based methods is that mimicking the genuine user's behavioral patterns to gain system access is considerably more difficult. These advantages have contributed to the growing prominence of the field, which has attracted considerable research attention in recent years.

### 2.1 Identification vs Verification

An important thing to mention is surely the difference between identification and verification. In fact, this difference represents the basis through which various biometric technologies have been developed.

Identification is the process by which a subject is identified, based on some characteristics that it possesses, from a large number of other subjects. Identification establishes a user's identity based on the information they provide. This can include a username, email address, or any other information that uniquely identifies them.

Verification is the process by which a machine can verify the correct identity of a subject. In particular a user can only be accepted or rejected, but not recognized over a large number of users. Verification confirms that the user is who they claim to be by comparing the information provided by the user against a trusted source, such as a government-issued ID, a passport, or a driver's license.

### 2.2 Biometric based on identification

The most well-known biometric technologies are those based on iris recognition and fingerprint scanning. Unlike traditional passwords, which identify a user through information they know, these technologies rely on intrinsic physical characteristics of the individual. They were specifically developed with the goal of user identification.

Fingerprints are a unique feature that every person has, and can be used to distinguish a person between every other subjects. Biometrics uses the fingerprint to identify and authenticate a person. First, the subject places their finger on a scanner which will be able to save the fingerprint as a biometric profile. This, unless there are particular accidents in the user's life, will remain the same until death. After this first step we will move on to the authentication phase. In this phase the subject will place their finger on the scanner again, which will take the biometric data and compare them with those saved in the previous step. The outcome of this procedure will be the authentication and identification of a user by the system if the data are matching and a reject if the data differs between them.

Biometric iris recognition uses the intrinsic composition of the human eye to authenticate and identify someone. The process is similar to that of a fingerprint, with some minor differences in the scanner used to create the biometric profile of a subject.

The figure 1 shows a real world example of scanners used for fingerprint and iris authentication and identification.

As per Nist [2] the best fingerprint systems can reach an accuracy of more than 99%. In particular, their test used operational fingerprints from a variety of U.S. and state government sources. A total of 48,105 fingerprint sets from 25,309 individuals, comprising 393,370 distinct fingerprint images, were used in the test. The performance was based on how many fingerprint images of a person are matched. The best system was accurate 98.6 percent of the time on single-finger tests, 99.6 percent of the time on two-finger



(a) Scanner for fingerprint



(b) Scanner for iris recognition

Figure 1: Different type of scanner

tests, and 99.9 percent of the time for tests involving four or more fingers. These results were obtained for a false positive rate of 0.01%, which indicates the people who should have been rejected but were instead accepted by the system.

For what concern the accuracy of iris recognition a Nec [3] study set a benchmark of 99.58% accuracy. This study was performed for a 1:N identification problem and it included over 500.000 participants, using both of their eyes for the trial.

As we expected these technology are very strong and reliable because are built over some intrinsic characteristic of a subject. These characteristics can be used to uniquely identify a person and remain unchanged throughout life. Another advantage is also that, unlike password-based systems, the user does not have to remember secret information. Furthermore, it is more difficult to try to authenticate using another person's fingerprint or iris compared to a system based on what a subject knows.

## 2.3 Behavioral biometric

Behavioral biometrics is a discipline primarily focused on the verification problem. Unlike physical biometrics, it does not rely on inherent physical characteristics, but rather on a person's behavior. Each individual behaves differently, using specific skills or strategies to accomplish tasks in everyday life. So the machine will be able to collect information about how a subject behave and it will be able to tell us if the user is really who is claiming to be. As we can expected the behavior is a soft biometric and less robust compared to the fingerprint or iris recognition. In fact this technology is not used in substitution of the former biometric system, but will be used as another security layer to verify the identity of a subject. In comparison to the former biometry the system can collect the data in a non intrusive way, without the user even realizing it, and it doesn't need an expensive hardware to do so.

As per Behavioral Biometrics for Human Identification: Intelligent Applications [4] behavioral biometric can be classified in five different categories based on the information about the user being collected:

- Authorship based biometric: Is used to collect something that a user produce like a drawing or a written text. Based on this the system will be able to recognize some particular characteristics about the style and verify a subject.
- Human Computer Interaction: Every user has a distinctive behavior when interact directly with a machine. The most famous method to do so is certainly the keyboard and the mouse. The way in which the user perform these actions is used as a method for verification.
- Indirect HCI-based biometrics: This method is related to the second point, so the interaction of a user with a machine. The goal is to use what the users leaves during the interaction, as an indirect

information, to verify it. Some examples of indirect information are audit log, program execution traces and memory usage.

- **Motor skills:** This is probably the best research group because it is based on some characteristic that the user have and learn during his life. The motor skill indicate the ability of the user to move muscles in a particular way to perform a task. These skills are distinctive to each user and is difficult even to replicate them in the same exact way, so it can be exploited from a verification application.
- **Purely behavioral biometrics:** The last category is based on the pure ability and the behavior used from a user to perform a task. These doesn't measure any muscle movement or something in particular, but it collect all the distinctive action that a user does. These actions vary from user to user and are based on the knowledge of a subject or the particular way in which he performs a task.

Behavioral biometrics is characterized by several key characteristics. Their universality is limited in the general population because behavioral traits vary in strength or may not exist for everyone. However, when applied in a specific domain, they can achieve full universality since all users are expected to demonstrate the relevant behavior.

In terms of uniqueness, behavioral patterns such as writing styles, gaming strategies, or preferences are not highly distinctive. This makes them more effective for verifying a claimed identity than for identifying individuals in large groups, unless the user base is very small.

Permanence is also a challenge since behavior changes over time as people learn new skills or adapt their techniques. This issue, known as concept drift, is addressed in research on intrusion detection, where adaptive systems are designed to accommodate evolving behaviors.

On the positive side, collectability is straightforward: data can be gathered unobtrusively, often without user awareness, and at low cost. Performance, however, tends to be weaker for identification in large databases, though verification accuracy is generally strong for certain behavioral biometrics.

With respect to acceptability, these methods are often well received due to their non-intrusive nature, though privacy and ethical concerns may arise.

Finally, circumvention is relatively difficult since it requires deep knowledge of another person's behavior, but once obtained, imitation can be simple. This highlights the importance of securely encrypting behavioral profiles.

## 2.4 Some applications of behavioral biometric

This section is used to present different real world application of behavioral biometric. These and other applications are described in Chapter 1, Taxonomy of Behavioral Biometrics [4], where they are subdivided into the five categories discussed earlier. In the thesis only three applications per typology will be represented to have a general overview. Is important to notice that some applications can be characterized with more than one class.

Starting from Authorship based biometric we have a biometric based on sketch recognition. With this method, proposed by Al-Zubi et al. [5], a user is invited to draw a simple sketch of a simple figure, for example some circles, and he can do it freely in the way he wants. This method captures the distinctive drawing style of the subject, allowing the system to build a model for authentication, achieving an Equal Error Rate (EER) of 7.2%. Another example of this category is email behavior used to identify anomalies into mail messages proposed by Stolfo et al. [6]. In fact the users follow some habits when they write or read an email. The time in which the subject reads the mails and empty the mail list, the structure of the mail, the structure of the greetings and the recipient address can be used to construct feature vectors and develop predictive models based on them. The application achieved a detection rate of 90.5% in identifying suspicious messages. The programming style is also a method used for authorship and was suggested by Spafford et al. [7]. With this method is possible to recognize who is the author of a program and this is useful to prevent malware or virus, with a detection rate of 73%. To create a model



the indentation, the comment style, the programming language used, variable names, type editor and so on are used as features.

The second group of models is the Human and computer interaction class and the email behavior described before is also classified under this name. Another example is the game strategy approach, in which a player model is build based on his characteristic when he plays. Yampolskiy et al. [8] proposed a system for verification of online poker players based on a behavioral profile which represents a statistical model of player's strategy. Some factors to consider include the range of cards a player holds and the level of aggression in their playstyle. This method can be used not only for verification, reaching an EER of 7%, but also to give other players hints about an opponent. Keystroke method is a method that can also be used for identification. The latter check the pattern used and the style a user uses when typing some characters on the keyboard. For example some users are really fast while others use only two fingers to type. We need only to type a short word, like a password, to be able to use this for verification while we need a lot more text if we want to take this model into an identification process. The keystroke approach for verification was proposed by Bergadano et al. [9], achieving a False Acceptance Rate (FAR) of 0.01% and a False Rejection Rate (FRR) of 4%.

The main applications of indirect HCI are surely the audit log, GUI interaction and network traffic. The audit log is a track that the operating system write automatically about a user activity. CPU and I/O usage, file creation and number of usage of a folder are data collected in a non-intrusive manner and it can be used to verify that the person using the PC is the legitimate user, with a detection rate of 93% and a false positive rate of 8%. Given the overwhelming amount of information available, using a random sample is a reasonable approach. Since the high potential of this approach there are a lot of researches in this field and one of them was performed by Denning, 1987 [10]. While GUI interaction, proposed by Garg et al. [11], collects data about the interaction of a user with a graphic interface. The collected data enables the creation of advanced behavioral profiles of the system's users. This comprehensive information can provide insights that are not obtainable from commonly analyzed command-line data. In fact it reaches a detection rate of 96.15% and a false positive rate of 3.85%. Network traffic analysis has grown significantly in recent years due to the large volume of data arriving from outside the organization. The main focus is to analyse various packet attributes in network traffic, including IP protocol types, packet size, server port numbers, source and destination IP prefixes, Time-To-Live values, IP/TCP header lengths and so on. The goal is to build a user profile based on these data and easily detect a deviation from the normal behavior and mark it as suspect. Due to the importance of this topic, extensive research has been conducted, including the work of Novikov [12]. The combined studies achieved a detection rate of 96.2% with a false positive rate of only 0.0393%.

One common approach of verification based on motor skill quality was proposed by Westeyn et al. [14] and is based on blinking. During the creation of the model the user listen a song and blink following the rhythm. Different factor are taken in consideration as the moment in which the user blink, the distance between each of them and the time in which the eye remain closed during that action. After the storage of a subject data is possible to being authenticated blinking again with the same stored song with a detection rate of 82.02%. Another famous application is the gait verification method proposed by BenAbdelkader et al. [16]. The profile of a user is built on the way a user walk taking in consideration the length of the steps, the bounce, distance between the two feet and distance between a foot and the head. This characteristic can changed due to some external factor like pregnancy, body weight or injuries. The advantage of this technique is that can be recorder in a non-intrusive manner using a camera and reaching a detection rate of 90%. The keystroke approach, described earlier, is another method that falls into both this category and the HCI category.

The last category is purely behavioral and shares with the authorship category the application of sketch verification. Another important application of this category is car driving style. People operate vehicles in very different ways—some drive cautiously and slowly, while others are more aggressive, often speeding or tailgating. Consequently, driving behavior can be considered a form of behavioral biometric. Erdogan et al. [18] and Erzin et al. [19] demonstrated that by analyzing variables such as accelerator and brake pedal pressure (in kilogram force per square centimeter), vehicle speed (in revolutions per minute), and steering angle (ranging from  $-720$  to  $+720$  degrees), it is possible to achieve effective genuine versus impostor driver authentication with a detection rate of 88.25%. The last example of this class is based on credit card usage. Data mining techniques are widely applied in credit card fraud detection, often by

identifying statistical outliers such as unusual transactions, distant locations, or simultaneous card use. Outlier detection methods rely on distance, density, projection, or distribution analysis. Brause et al. [20] developed a rule-based model using symbolic and analog transaction data, showing that analog data alone is insufficient for reliably detecting fraud with a detection rate of 99.995% and a FRR of 20%.

## 2.5 Soft Biometric

Behavioral biometrics is considered a soft biometric approach because it is generally weaker and less stable than traditional biometric systems. Unlike traditional biometric applications, such as fingerprint or iris recognition, which are based on physiological traits that remain largely constant throughout a person's life, behavioral biometrics rely on dynamic patterns of human activity.

Traditional biometric systems, particularly fingerprint and iris recognition, are regarded as highly reliable and relatively strong approaches. They are built upon physical characteristics that an individual permanently possesses, making them highly distinctive and resistant to change. As noted earlier, these two modalities can reach nearly 100% accuracy in identification tasks, making them capable of reliably authenticating an individual even within large populations.

By contrast, behavioral biometric technologies can sometimes be applied for identification, but they are generally more suitable for verification tasks. Their typically lower accuracy in identification makes them less suitable as standalone replacements for traditional physiological biometrics. This limitation arises from the fact that behavior is inherently variable: the way a person types, writes, speaks, or drives may change depending on mood, fatigue, health, stress, or environmental conditions.

Nevertheless, behavioral biometrics hold a unique advantage. Unlike traditional biometrics, which often require active cooperation (e.g., placing a finger on a scanner or looking into an iris reader), behavioral systems can operate passively. They collect data continuously in the background, often without explicit effort or awareness on the part of the user. This makes them particularly attractive for continuous authentication scenarios. Instead of relying on a single login event, the system monitors user behavior throughout a session to ensure that the authenticated person remains the one interacting with it. Another strength of behavioral biometrics is their potential to complement traditional systems. For instance, combining fingerprint authentication with continuous keystroke analysis or touch dynamics provides a layered defense: even if a fingerprint is compromised, the behavioral system can still flag anomalies. This layered approach enhances both usability and security, reducing the risk of unauthorized access.

Behavioral biometrics also raise important privacy and ethical considerations. Because these systems can function unobtrusively in the background, they may collect user data without the individual's explicit awareness. While this capability enhances security, it also underscores the need for strict regulation, informed consent, and robust data protection measures to prevent misuse.

## 3 Eye tracker

Before describing an eye tracker we have to understand why the point in which a person is looking is so important. The reason is that when a person looks at something, they focus on a scene in fine detail and direct their attention to that point. Therefore, if we can detect where a person is looking during an activity, we can infer their path of attention. We can also know which part of the scene can be labeled as important for a subject.

### 3.1 Visual attention

Humans receive numerous stimuli during any activity, but they cannot process all of them due to limited cognitive capacity. As a result, they must select a subset of stimuli they find more interesting or important and focus their processing on those. The psychologist William James gives the definition of visual attention as: “Many filtered into few for perception”. This section, according to chapter 1: Visual attention [21], will contain the history of the main study about visual attention during the past years.

#### 3.1.1 Von Helmholtz’s “Where”

In the late 19th century, Von Helmholtz (1925) emphasized visual attention as a key mechanism of perception. He noted that while attention naturally shifts to new stimuli, it can also be consciously directed even toward peripheral objects without moving the eyes. However, he mainly linked attention to eye movements toward spatial locations, which serve as evidence of overt visual attention.

#### 3.1.2 James’ “What”

In contrast to Von Helmholtz’s spatial (‘where’) view of attention, James (1891) described attention as an internal, covert process concerned with the identity or meaning of stimuli the ‘what’. These perspectives are complementary, reflecting two aspects of visual processing. The foveal, or ‘what,’ component enables detailed recognition of central objects, while the parafoveal, or ‘where,’ component guides spatial awareness and directs attention toward peripheral stimuli. Together, they illustrate how visual attention integrates object identification with orientation in the surrounding environment. Although this dual view forms the basis for computational models of bottom-up attention, it is inherently simplistic. A comprehensive understanding of visual attention must also account for higher-level, voluntary, and cognitive factors that influence how attention is directed.

#### 3.1.3 Gibson’s “How”

In the 1940s, Gibson (1941) introduced intention as a third factor in visual attention, emphasizing the role of advance preparation in shaping reactions. His view highlighted the interplay between expectation and intention, showing how perception can shift depending on preconceptions, such as reading ‘sael’ as ‘seal’ or ‘sail’. This ‘what to do’ aspect underscores the importance of considering perceptual expectations when designing experiments.

#### 3.1.4 Broadbent’s “Selective Filter”

In the 1950s, Broadbent (1958) introduced the ‘selective filter’ model of attention, showing through auditory experiments that while information enters in parallel, it is subsequently filtered into limited-capacity sensory channels.

### 3.1.5 Deutsch and Deutsch’s “Importance Weightings”

Deutsch and Deutsch (1963) opposed Broadbent’s selective filter, proposing instead that all sensory input is processed, with importance weightings determining attention. Treisman (1971) later reconciled both models by introducing an attenuation filter that weakens, rather than blocks, irrelevant signals, followed by ‘dictionary units’ tuned to importance and context the ‘where’ and the ‘what’. However, the unresolved scene integration problem highlighted how, despite selective perception, we can still construct a coherent overall view of complex scenes, a challenge further explored through eye movement studies and Gestalt theories.

### 3.1.6 Yarbus and Noton and Stark’s “Scanpaths”

Early eye movement studies challenged the Gestalt idea of parallel, holistic visual recognition. Yarbus (1967) showed that viewers’ gaze followed sequential patterns when answering questions about an image, while Noton and Stark (1971) identified ‘scanpaths’, demonstrating that even without specific instructions, observers tend to fixate on informative regions in variable sequences. These findings suggest that a coherent visual scene is constructed piecemeal, supporting James’ ‘what’ aspect of attention, where foveal vision selectively processes regions of interest.

### 3.1.7 Treisman’s “Glue”

Posner et al. and Noton and Stark extended the theories of visual attention originally proposed by Von Helmholtz and James, later integrated by Broadbent and Deutsch & Deutsch. Treisman further unified these ideas through the Feature Integration Theory (FIT), which posits that attention acts as the ‘glue’ binding separate features at specific locations to perceive objects as unified wholes. FIT relies on a master map indicating feature locations such as color, orientation, size, and depth while eye tracking is often used experimentally to validate the theory.

### 3.1.8 Kosslyn’s “Window”

Kosslyn (1994) proposed a refined model of visual attention, introducing an adjustable ‘attentional window’ that selectively transmits information from the visual buffer, similar to Broadbent’s and Treisman’s filters. The model also posits a redundant, stimulus-based attention-shifting subsystem, relevant to mental imagery, which involves forming internal representations resembling actual perception. Notably, eye movements during REM sleep may reflect this internal attentional process, though this remains uncertain.

### 3.1.9 Bottom-up model

Visual attention can be conceptualized through two complementary components: the “what”, corresponding to foveal, object-focused processing, and the “where”, associated with parafoveal, spatial guidance. In a bottom-up model, peripheral vision first detects salient features at low resolution, which then guide eye movements to reposition the fovea on regions of interest for detailed inspection. This framework provides a powerful foundation for computational models of visual search, yet it remains incomplete because it does not account for higher-level, voluntary, or cognitive influences. Voluntary attention enables individuals to selectively focus on areas outside their direct foveal gaze, as demonstrated in scanpath studies and practical examples such as astronomers detecting faint stars through peripheral vision. However, while eye tracking offers valuable insights into attentional processes, it is limited to capturing overt eye movements. It cannot capture covert shifts of attention, in which focus is redirected without accompanying eye movements, which represents an important limitation for observational studies.”

## 3.2 The human eye

The human eye is a highly specialized organ responsible for capturing light from the environment and transforming it into neural signals that the brain interprets as visual perception. Its structure is finely adapted to focus incoming light, regulate brightness, and detect a wide spectrum of colors and details. Understanding the anatomy of the eye is crucial to explain how humans perceive the world, particularly the mechanisms of color vision, which rely on the retina and its photoreceptor cells: rods and cones.

### 3.2.1 Anatomy of human the eye

The human eye is approximately spherical, with a diameter of about 24 millimeters.

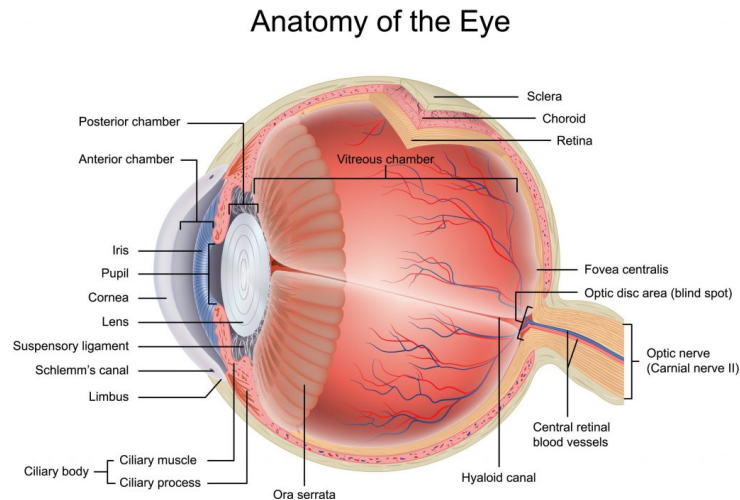


Figure 2: Anatomy of the human eye

Despite its relatively small size, it contains an intricate arrangement of structures, each serving a distinct role in the process of vision. At the very front lies the cornea, a transparent, curved surface that acts as the first point of contact for incoming light. Its main function is to bend, or refract, the light so that it can be properly directed toward the deeper structures of the eye. Just behind the cornea is the iris, the colored part of the eye, which regulates the size of the pupil, the central black opening. By widening or narrowing, the iris controls how much light enters the eye, functioning much like the aperture of a camera. Positioned behind the iris is the lens, a flexible and transparent structure that fine-tunes focus. Through the action of tiny ciliary muscles, the lens can change its shape, allowing the eye to adjust focus when shifting between near and distant objects, a process known as accommodation. Light then travels to the retina, a delicate layer of neural tissue lining the back of the eye. Far from being a passive screen, the retina actively processes the incoming light signals. It contains specialized photoreceptor cells, rods and cones, that convert light into electrical impulses. These signals are then transmitted through the optic nerve, a dense bundle of nerve fibers, to the brain's visual cortex, where they are interpreted and transformed into meaningful images.

### 3.2.2 The retina and photoreceptor

As we said earlier the retina plays a central role in vision, functioning as the site where light is transformed into neural signals that the brain can interpret. Within this thin layer of tissue reside two main types of photoreceptor cells, namely rods and cones, each of which is adapted for specific visual tasks.

Rods are highly sensitive to light and enable vision under low-light conditions. Although they do not mediate color perception, they are essential for detecting shapes, movement, and variations in brightness. Humans possess approximately 120 million rods in each eye, with the majority concentrated in the

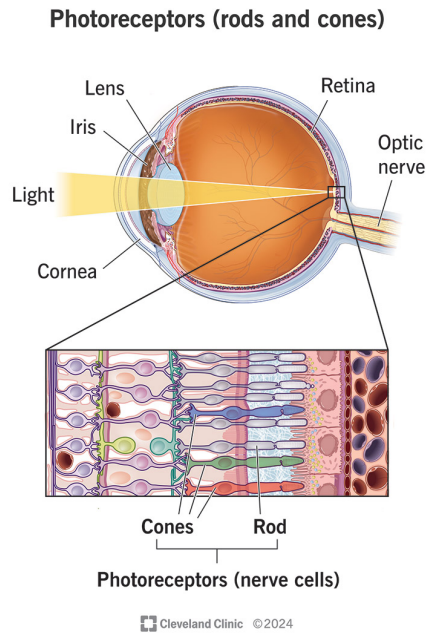


Figure 3: Rods and Cones

peripheral regions of the retina. This distribution explains why peripheral vision is far more effective in low-light environments, even though it lacks the ability to discriminate color. Cones, by contrast, are less sensitive in low-light conditions but are indispensable for perceiving color and fine detail. About 6 million cones are found in each eye, most of them concentrated in the central retina and especially within the fovea, a small depression that provides the sharpest visual acuity. Cones are divided into three subtypes according to the wavelengths of light to which they are most responsive. Some are sensitive to short wavelengths corresponding to blue light, others respond to medium wavelengths associated with green, and still others detect long wavelengths corresponding to red. The combined activity of these three types of cones forms the basis of human trichromatic vision, enabling us to perceive the full spectrum of visible colors. This characteristic explains why the RGB color model is among the most widely used frameworks in digital imaging and display technologies.

### 3.2.3 Color perception

The process of color perception begins when incoming light stimulates the cones. Each cone type has a broad response curve, meaning that their sensitivity ranges overlap. The brain does not simply interpret the activity of a single cone in isolation, but instead compares the relative responses of the three cone classes to determine the perceived color. For instance, strong activation of long-wavelength cones with weaker responses from the other two is interpreted as red. When medium- and long-wavelength cones are equally stimulated while short-wavelength cones contribute little, the result is perceived as yellow. If all three types respond at similar levels, the visual system interprets the signal as white. Thus, color vision is not a direct reading of individual wavelengths but an emergent property of neural comparison and integration across multiple cone pathways. The interaction between rods and cones further illustrates the complexity of human vision. Cones dominate during daylight conditions, but as light diminishes, rods become increasingly important. This shift, known as the Purkinje effect, reflects the higher sensitivity of rods to blue-green wavelengths. As a result, red objects appear to darken as twilight falls, while blue ones remain more easily visible. Such changes highlight the complementary nature of rods and cones: one system optimized for detail and color in bright environments, and the other specialized for sensitivity in low-light conditions.

### 3.2.4 Type of eye movements

Human vision is not a static process. The eyes are in constant motion, guided by a sophisticated neuromuscular system that allows the fovea, the central region of the retina responsible for high acuity, to be directed toward relevant visual targets. These movements are not random, but are precisely coordinated to stabilize images, explore the environment, and ensure that critical details are perceived accurately. Almost all normal primate eye movements can be classified into five basic types: saccadic, smooth pursuit, vergence, vestibular, and physiological nystagmus. Each serves a distinct function, yet together they form an integrated system that balances accuracy, stability, and adaptability in visual perception.

Saccades are rapid, ballistic shifts of the eye that redirect the fovea from one point of interest to another. They are the fastest of all eye movements, reaching velocities of up to 900 degrees per second, and typically last only a few tens of milliseconds. Despite their speed, saccades are precisely controlled, allowing humans to scan text when reading, explore complex scenes, or quickly detect changes in the environment. During saccades, visual sensitivity is temporarily reduced through saccadic suppression, preventing motion blur and maintaining visual stability.

In contrast to the abruptness of saccades, smooth pursuit movements allow the eyes to follow a moving object with remarkable fluidity. These movements are slower, generally limited to about 100 degrees per second, but are highly adaptive to the velocity of the target. Smooth pursuit depends on both visual feedback and predictive mechanisms, as the eyes often anticipate the trajectory of an object to maintain accurate tracking. This function is particularly critical for activities such as driving, sports, or any task where continuous monitoring of moving stimuli is required.

Vergence refers to the inward or outward rotation of the eyes that aligns them with targets at different depths. Unlike saccades or pursuit, vergence movements are disjunctive: the two eyes move in opposite directions. Convergence occurs when the eyes rotate inward to focus on a near object, while divergence refers to outward rotation for distant viewing. These adjustments are essential for depth perception and binocular vision, allowing humans to perceive the world in three dimensions. Vergence movements are slower than saccades but are tightly linked with accommodation of the lens to maintain clear and unified images across distances.

The vestibular system, located in the inner ear, interacts with eye movements to ensure visual stability during head motion. The vestibulo-ocular reflex (VOR) automatically generates compensatory eye movements that counteract head displacement, allowing the gaze to remain fixed on a target even when the body is in motion. For example, when walking or turning the head, the VOR maintains stable vision by producing equal and opposite eye rotations. This reflex is exceptionally fast, with response latencies of just a few milliseconds, underscoring its role in stabilizing vision in dynamic environments.

Although the eyes may appear still when fixating on an object, they are never completely motionless. Physiological nystagmus refers to small, involuntary movements—including tremors, drifts, and microsaccades—that occur during fixation. These tiny displacements prevent the image on the retina from fading due to neural adaptation. Without them, stationary objects would gradually disappear from perception. Physiological nystagmus, therefore, is a crucial mechanism for maintaining continuous visual awareness of the environment, even when the gaze is held steady.

Though described separately, these five types of movements rarely operate in isolation. In natural vision, they combine seamlessly to provide both stability and flexibility. For instance, during visual search, saccades bring the fovea to regions of interest, smooth pursuit maintains focus on moving targets, vergence adjusts for depth, vestibular reflexes stabilize the gaze against body motion, and physiological nystagmus ensures constant retinal stimulation. Together, they form a dynamic and adaptive system that underpins the human ability to explore, interact with, and interpret the visual world.

### 3.2.5 Fixation

Fixation is the process by which the eyes maintain their gaze on a single point, allowing the visual system to capture and process detailed information from the environment. Although the eyes may appear

motionless during fixation, they are in fact never completely still. Tiny involuntary movements such as microsaccades, tremors, and drifts occur constantly, preventing the image on the retina from fading due to neural adaptation and ensuring continuous visual awareness. These subtle motions are essential for maintaining clarity and precision, as they allow the fovea, the central region of the retina responsible for high-acuity vision, to remain accurately aligned with the target. Fixation is fundamental to how humans interact with the visual world. It provides a stable platform for extracting fine details, enabling activities that require careful attention, such as reading, inspecting objects, or monitoring a complex scene. During fixation, the visual system is able to prioritize relevant information, enhancing perception and supporting higher-level cognitive processes such as attention, memory, and decision-making. Even in dynamic environments, fixation works in concert with other eye movements: after rapid saccades redirect the fovea to a new location, fixation allows the visual system to stabilize and analyze the target. While during smooth pursuit of moving objects, brief fixations may occur to refine the image and maintain accuracy. Physiological nystagmus during fixation further ensures that retinal stimulation is continuous, allowing the perception of stationary objects to remain vivid and precise. In essence, fixation is not merely a passive pause of the eyes but a dynamic and intricate process. It balances stability and subtle motion, integrating seamlessly with other eye movements to support accurate, continuous, and detailed visual perception. Without fixation, the human ability to explore, interpret, and respond to the environment with precision would be significantly impaired, highlighting its critical role in the visual system

### 3.3 Eye tracking techniques

After recognizing the importance of eye movements, it becomes clear that we need a tool to track them, and this tool is called an eye tracker. In the past, eye trackers were highly expensive devices, mostly limited to research labs and specialized industries. Over the years, however, their price has steadily decreased, making the technology more accessible. Today, affordable solutions are available not only for scientific research but also for commercial applications such as gaming, marketing, user experience studies, and even healthcare. This reduction in cost has opened the door to new innovations, allowing eye tracking to move from a niche tool to a widely adopted solution across many fields. Eye trackers can be divided into two main categories: those that measure the position of the eye relative to the head and those that measure the orientation of the eye in space, also known as the Point of Regard (POR). Building on this distinction, Chapter 5 of *Eye Tracking Methodology: Theory and Practice* [21] further identifies four methodological approaches to eye movement measurement: Electro-Oculography (EOG), scleral contact lens or search coil methods, Photo-Oculography (POG) or Video-Oculography (VOG), and video-based techniques combining pupil and corneal reflection.

#### 3.3.1 Electro-OculoGraphy (EOG)

Electro-oculography (EOG) was the predominant method for recording eye movements around forty years ago and, although less common today, it still finds use in both clinical and research settings. The technique operates by detecting differences in the corneo-retinal standing potential, which are measured through electrodes placed on the skin surrounding the eye. The recorded signal typically falls within the range of 15 to 200  $\mu\text{V}$ , with a nominal sensitivity of approximately 20  $\mu\text{V}$  per degree of eye movement. While EOG is effective for capturing eye movements relative to head position, it lacks the precision required for determining the point of regard. Consequently, its application for gaze tracking is limited unless combined with complementary systems capable of monitoring head movements, such as head trackers.

#### 3.3.2 Scleral Contact Lens/Search Coil

One of the most accurate ways to measure eye movements uses a mechanical or optical reference attached to a contact lens that is worn directly on the eye. Early versions of this method, dating back to the late 19th century and later developed by Young and Sheena (1975) [22], used a plaster-of-Paris ring placed on the cornea, connected by mechanical linkages to recording devices. Modern implementations use a scleral contact lens, which covers both the cornea and part of the sclera (the white of the eye) to prevent the lens



from slipping. A small mounting stalk is attached to the lens, and various devices can be placed on this stalk, such as reflective markers, line diagrams, or wire coils. In magneto-optical eye-tracking systems, the wire coil is the most commonly used device. The coil is attached to the contact lens and moves together with the eye. As it rotates within a carefully controlled electromagnetic field, the system detects changes in the coil's orientation and generates precise measurements of eye movement. This setup allows extremely fine tracking of even very small rotations of the eye, making it one of the most accurate methods for recording eye position. Based on Young and Sheena researches [22] the scleral search coil is extremely precise, with an accuracy of about 5–10 arc-seconds over a small range of roughly 5°. However, it is also highly invasive. Inserting the lens requires practice and care, and wearing it can be uncomfortable. In addition, this method measures the eye's position relative to the head, so it is generally not suitable for directly determining the point of regard.

### 3.3.3 Photo-OculoGraphy (POG) or Video-OculoGraphy (VOG)

This category includes a wide range of eye movement recording techniques that rely on measuring distinct features of the eye during rotation or translation. Examples include the apparent shape of the pupil, the position of the limbus (the boundary between the iris and the sclera), and reflections from a nearby directed light source, often in the infrared spectrum. Although the specific methods differ, they are grouped together because they generally do not provide direct measurements of the point of regard. Measurements of ocular features in these techniques may be obtained either automatically or manually. Manual visual inspection, such as reviewing recorded eye movements frame by frame on videotape, is extremely time-consuming, prone to error, and limited by the temporal sampling rate of the video device. Automatic limbus tracking typically uses photodiodes mounted on spectacle frames and generally relies on invisible illumination, usually in the infrared range. Many of these methods also require the head to be held still, commonly using a head or chin rest, or a bite bar, to reduce movement (Young and Sheena, 1975)[22].

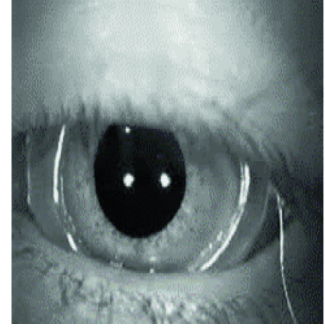
### 3.3.4 Video-Based Combined Pupil/Corneal Reflection

While the previously discussed techniques are generally suitable for measuring eye movements, they often do not directly provide point-of-regard measurements. To determine the point of regard, either the head must be stabilized so that eye position relative to the head corresponds to gaze direction, or multiple ocular features must be tracked to distinguish between head movements and eye rotations. Two key features commonly used for this purpose are the corneal reflection of an infrared light source and the center of the pupil. Video-based eye trackers employ relatively inexpensive cameras and image-processing systems to calculate the point of regard in real time. These systems may be table-mounted or head-mounted. The optical principles behind both configurations are essentially the same, with size and portability being the main differences. Modern video-based eye trackers are compact and often integrated into monitors, making them particularly suitable for interactive applications. In these systems, the corneal reflection, also known as a Purkinje image, is measured relative to the pupil center. The human eye generates four distinct Purkinje reflections, which are formed by light reflecting off different surfaces within the eye: the anterior and posterior surfaces of the cornea, and the anterior and posterior surfaces of the lens (Crane, 1994) [23]. Each reflection carries unique optical information, but most video-based eye tracking systems rely primarily on the first Purkinje reflection, which is produced by the anterior corneal surface. This reflection is the brightest and most stable, making it well suited as a reference point for calculating gaze direction when measured relative to the pupil center. With proper calibration, these trackers can reliably map a viewer's point of regard on a planar surface where calibration targets are displayed. Tracking both the pupil center and the corneal reflection allows the system to separate eye rotations from minor head movements. Specifically, the relative position between the pupil and the first Purkinje reflection changes with eye rotation but remains nearly constant with small translations of the head. More advanced systems, known as generation-V eye trackers, also measure the fourth Purkinje reflection (Crane & Steele, 1985) [24]. These dual-Purkinje image (DPI) eye trackers can distinguish between translational and rotational eye movements: both reflections shift equally during translation, but they move differently during rotation, allowing the separation of the two effects. Although highly accurate,

DPI systems often still require the head to be stabilized for optimal performance. This is because even small head translations or rotations can introduce noise into the measurement, making it difficult to distinguish subtle eye movements from head motion. Stabilization methods, such as chin rests, bite bars, or other mechanical supports, are therefore commonly used to minimize artifacts. While this limitation reduces the comfort and naturalness of the setup, it ensures the precision needed for research applications that demand extremely fine measurements of gaze direction.



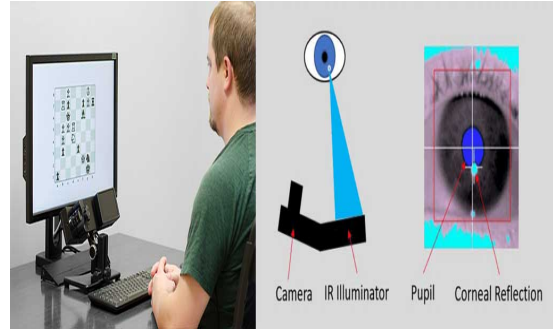
(a) Electro-OculoGraphy (EOG)



(b) Scleral Contact Lens/Search Coil



(c) Video-OculoGraphy (VOG)



(d) Corneal Reflection

Figure 4: Different types of eye tracker

## 4 Related work

This section presents an updated overview of the state of the art in gaze-based biometric solutions. As a starting point, we draw on the work presented in Chapter 9 – Biometric Authentication to Access Controlled Areas Through Eye Tracking [25], published in 2017. Building upon this foundation, the review is expanded to include significant experimental studies and technological developments that have emerged in the years since, with particular attention to contributions up to 2025. In this chapter, we deliberately exclude methods that rely on static characteristics of the human eye, such as iris-based recognition, which are already well established in biometric research. Instead, we focus on dynamic approaches, in which identity-related information is derived either from the manner in which an individual visually interacts with stimuli or from temporal changes in particular ocular characteristics. The way an individual explores a visual scene, whether in a free-viewing context or under controlled conditions, can reveal distinctive patterns that serve as biometric cues. In addition, time-dependent changes in ocular features, such as variations in pupil diameter, may provide further information useful for identity verification or recognition. Building on this perspective, five principal categories of eye-tracking-based biometric techniques will be considered. The first includes ATM-like solutions, which simulate practical authentication scenarios. The second focuses on fixation and scanpath analysis, where the sequence and duration of gaze points are examined. The third category emphasizes eye and gaze velocity, capturing the dynamics of eye movement. The fourth examines pupil size variations, which may be influenced by both physiological and cognitive factors. Finally, the fifth category explores oculomotor characteristics and head orientation, integrating motor control features of the visual system into biometric assessment.

### 4.1 ATM-like solutions

This section begins with a review of gaze-driven authentication systems, which have been developed as alternatives to traditional password or PIN entry methods. Such systems are particularly relevant in contexts resembling Automated Teller Machines (ATMs), where secure and efficient user authentication is essential. By replacing manual input on a keyboard with gaze-based interactions, these approaches aim not only to improve usability but also to reduce vulnerabilities such as shoulder surfing or keylogging, while at the same time opening new possibilities for integrating behavioral traits into the authentication process. While these solutions are not strictly biometric, since the underlying information often comes from sequences of alphanumeric characters or graphical symbols, the way such sequences are entered through eye movements introduces behavioral elements that may carry biometric value. For this reason, these approaches merit discussion as precursors to fully gaze-based biometric methods.

One of the earliest contributions in this area was presented by Maeder et al. [26], who explored gaze-directed authentication based on consciously fixating predefined regions of a visual stimulus for specified durations. In their study, participants were shown an image of Prague divided into a non-uniform  $3 \times 3$  grid. Each user selected six “points of interest,” with one in each grid cell, and authentication was achieved by reproducing the same sequence of fixations. In this design, the gaze path effectively served as a PIN, while the grid structure ensured both consistency and flexibility.

A further step toward practical implementations was proposed by Kumar et al. [27] in their system EyePassword, which aimed to reduce vulnerabilities such as shoulder surfing and eavesdropping. Four input modalities were compared: a traditional keyboard, gaze plus manual confirmation on an on-screen QWERTY keyboard, gaze with dwell time on a QWERTY layout, and gaze with dwell time on an alphabetic layout. Although average entry times across the dwell and trigger-based methods were comparable (between 9.2 and 12.1 seconds), error rates were notably higher when a manual trigger was used (15% versus 3–4%). These results highlighted both the feasibility of gaze input for secure authentication and the trade-offs between speed, accuracy, and usability.

In another line of research, De Luca et al. [28] proposed the use of eye gestures as authentication tokens. This method was inspired by the observation that visual patterns or geometric shapes are often easier to memorize than long alphanumeric strings. Here, authentication was performed by “drawing” gestures on the screen with the gaze, using eight basic directional strokes originating from or directed toward the screen’s center. Such gestures represented an alternative form of input, combining memorability with the

natural motor control of the eyes.

Similarly, Dunphy et al. [29] replaced numeric or textual credentials with a recognition-based scheme relying on faces. Users were asked to memorize a set of faces during an initial training phase and later authenticate by selecting them from sequences of grids containing distractor images. This approach emphasized recognition rather than recall, exploiting the human visual system’s efficiency in face perception.

Other researchers explored the use of gaze interaction with symbolic keyboards. Weaver et al. [30] developed a system where authentication was achieved by looking at symbols arranged on a virtual keyboard. Instead of relying solely on dwell time for character selection, gaze points were clustered and automatically analyzed to infer the intended input, thereby reducing delays and improving fluidity.

More recently, Cymek et al. [31] investigated authentication through smooth pursuit eye movements, which occur when the gaze follows a continuously moving target. Their method used a moving PIN pad where each digit followed a unique trajectory, defined by a sequence of up, down, left, or right movements. Authentication was achieved by tracking the trajectory of the correct digits with the eyes. Although this technique was slower than traditional dwell-time input, it offered distinct advantages: it reduced the need for calibration and made it more difficult for attackers to infer the entered PIN, since the input was encoded in movement patterns rather than in static gaze points.

Khamis et al. [37] conducted a rigorous comparison of cue-based authentication modalities for situated displays, focusing on touch, mid-air gestures, and calibration-free gaze. Through two in-depth user studies ( $N = 20$ ,  $N = 17$ ), they highlighted a clear tradeoff between usability and security. Gaze input was found to be more secure against shoulder-surfing attacks, but it was also more demanding and slower. Mid-air gestures offered slightly higher security than touch, though they were less favored in public contexts. Touch, on the other hand, remained the fastest but was less secure. Their findings underline important design implications for balancing efficiency, security, and user acceptance in public authentication scenarios.

Omar Shahzad Bhatti et al. [36] proposed a calibration-free authentication method for interactive public displays based on saccadic eye movements, addressing the limitations of slower gaze-based approaches. In a user study with ten participants, their method was compared against CueAuth, a smooth pursuit-based system by Khamis et al. [37]. The results showed a substantial improvement in accuracy (82.94% to 95.88%) and a significant reduction in entry time (from 18.28s to 5.12s), making the approach comparable to touch-based input in speed while maintaining strong security.

Authentication systems are now widely used to protect personal information, but common approaches such as PINs, biometrics, and smart cards remain vulnerable. PINs are easily exposed through shoulder-surfing, biometrics can be spoofed, and smart cards are at risk of theft or duplication. To address these limitations, Ail Hamid Malla [38] proposes a gaze-based authentication system designed to counter shoulder-surfing, video-based observation, and spoofing attempts. Experimental results demonstrate the effectiveness of the approach, achieving an accuracy of 97.5% (F-measure = 0.97) in authenticating legitimate users and 89.5% (F-measure = 0.89) in detecting intruder attempts.

Eye tracking has gained attention as a promising tool for biometric authentication, as eye movements are fast, distinctive, and require minimal effort from users. However, earlier gaze-based systems often struggled with high error rates or slow authentication. To address these issues, Ivo Služanović et al. [39] developed a method that uses reflexively triggered eye movements through interactive visual stimuli, enabling reliable feature extraction within seconds and eliminating the need for memorization. An additional advantage is that the stimulus can be uniquely generated for each login attempt, supporting a challenge–response mechanism that defends against replay attacks. A prototype tested with 30 participants demonstrated strong performance, achieving an equal error rate of 6.3% and reducing authentication time to about 5 seconds, outperforming many prior approaches.

Matin Fallahi et al. [40] proposed a multimodal biometric authentication system for Extended Reality (XR), addressing the usability and security limitations of password-based methods. Traditional password entry in XR is cumbersome and error-prone, often leading to weaker password choices and reduced security. To overcome this, their approach combines eye movement patterns with brainwave signals, using consumer-grade sensors that can be seamlessly integrated into XR devices. A prototype tested with 30 participants achieved an exceptionally low Equal Error Rate (EER) of 0.29%, significantly outperforming

single-modality systems based on eye movements (1.82%) or brainwaves (4.92%), as well as state-of-the-art biometric methods (EERs between 2.5% and 7%). Importantly, this solution supports seamless and immersive authentication through visual stimuli, eliminating the need for complex user interactions.

Traditional authentication methods such as passwords overlook personal behavioral traits of users. This study, suggested by Cantoni et al. [44], explores the use of eye-tracking data for biometric purposes during PIN entry via an on-screen virtual numeric keypad. Both identification and verification are investigated. Machine learning algorithms are applied to analyze gaze patterns across the full PIN sequence as well as at the level of individual digits. The results, while not achieving the precision of traditional biometrics, are satisfactory within the context of soft biometrics, where the goal is to complement existing authentication methods—here, the PIN itself—by providing an additional layer of security.

Haochen Wang [47], in his thesis, presents four case studies evaluating gaze-based biometrics, including doorbell name inspection for building access, static image observation, moving target observation, and eye-driven soft PIN input. Both identification and authentication were tested, including an analysis of user familiarity in the building access study. Results showed classification accuracies above 0.7 across most experiments, demonstrating the feasibility of gaze-based biometrics as a viable “soft biometric” method when combined with traditional authentication techniques.

## 4.2 Methods based on fixation and scanpath analysis

The use of fixations and scanpaths for biometric applications has been investigated for over a decade. One of the earliest contributions was by Silver and Biggs [53], who compared keystroke dynamics with eye-tracking data for identification purposes. Although keystroke-based models clearly outperformed gaze-based ones, this work was among the first to highlight the potential of eye tracking as an alternative technique for biometric verification.

Subsequent studies have sought to extract and refine gaze features for improved performance. Holland and Komogortsev [54], for example, examined reading scanpaths using a high-frequency eye tracker (1000 Hz) and a dataset of 32 participants. They employed a wide range of features, such as fixation count, average fixation duration, and saccade amplitudes, as well as aggregate scanpath descriptors including path length, area, and inflection count. By applying Gaussian cumulative distribution similarity measures and feature fusion, they achieved their best result with an Equal Error Rate (EER) of 27%.

In later work, Holland and Komogortsev [55] expanded this line of inquiry using both high-end (1000 Hz) and low-cost (75 Hz) eye trackers. From 32 and 173 participants, respectively, they extracted fixation and saccade features, which were analyzed with several statistical tests and machine learning classifiers. Their approach produced an EER of 16.5% and a rank-1 identification rate of 82.6%, demonstrating the potential of combining traditional statistics with machine learning for gaze-based biometrics.

Alternative methodologies have also been explored. Rigas et al. [56] introduced a graph-based approach in which fixations on face images were clustered using a minimum spanning tree (MST) technique. Comparing overlap in MST structures between subjects enabled user identification with accuracies ranging from 67.5% to 70.2% across 15 participants and eight sessions.

Galdi et al. [57] conducted an experiment involving 88 participants across three sessions. Sixteen grayscale facial images were used as stimuli, and each image was divided into 17 Areas of Interest (AOIs) corresponding to facial regions (e.g., eyes, mouth, nose). For each participant, feature vectors were constructed by calculating the average fixation duration per AOI across all images. Using Euclidean and Cosine distance measures for comparison, the system achieved its best performance with combined features, reaching an Equal Error Rate (EER) of 0.361.

Cantoni et al. [33] extended this approach with a larger participant pool (111 from the first study plus 24 additional subjects one year later) and introduced the Gaze ANalysis Technique (GANT). Instead of manually defined AOIs, an automatic face normalization algorithm was applied to divide each image into a uniform 7×6 grid (42 cells). Weights were calculated for each cell based on fixation density and duration, as well as for arcs connecting cells to capture gaze transitions. Similarity between models was

assessed by comparing difference matrices via the Frobenius norm. Trials with both single and combined features showed that the best results were achieved when all features (density, fixation duration, and weighted paths) were fused, yielding a minimum EER of 28%.

Further contributions expanded the scope of gaze-based biometrics beyond identity verification. Cantoni et al. [58] and Galdi et al. [59] applied features derived from GANT datasets to gender and age classification. Using Adaboost and Support Vector Machines, they reported moderate success, with gender classification rates near 60% and age classification rates around 55%.

Other research has focused on task dynamics as a differentiating factor. Biedert et al. [60], for instance, examined task learning effects to detect unauthorized system access. They hypothesized that genuine users would navigate familiar interfaces fluidly, while impostors would exhibit higher exploratory behavior. To quantify this, they introduced the Relative Conditional Gaze Entropy (RCGE) method, which analyzes gaze distribution patterns.

George and Routray [61], winners of the BioEye 2015 competition, developed a highly effective approach to biometric identification. Using recordings from 153 participants across three sessions (two separated by 30 minutes and one by one year), they extracted extensive fixation and saccade features and applied backward feature selection with a Radial Basis Function Network (RBFN). Their method achieved success rates of 98% across the shorter interval and 94% across the one-year interval, highlighting the robustness of gaze-based biometrics under temporal variability.

Yin et al. (2022) [34] proposed a gaze-based authentication method leveraging the spatiotemporal features of eye movements. To address the long recording times required by previous approaches, their method uses shorter eye movement recordings, recommending durations under 12 seconds based on deviations between gaze points and stimuli. Temporal motion of gaze points and spatial saccade patterns were used to represent individual identity. Experiments on two datasets (5 s and 12 s recordings) showed that open-set Equal Error Rates (EER) reached 10.62% for 12 s and 12.48% for 5 s, while closed-set EERs were 5.25% and 7.82%, respectively, demonstrating the method’s effectiveness for rapid, eye-movement-based authentication.

The COVID-19 pandemic underscored the need for contactless authentication methods, particularly those based on eye or gaze features, which remain effective even when users wear masks and can complement traditional password or PIN systems. In this context, Porta et al. [35] presented a study on gaze-based soft biometrics that employed simple animations as visual stimuli, in which small squares moved along different patterns and trajectories without the need for preliminary calibration of the eye tracker. Machine learning analysis of the collected data demonstrated strong performance, especially in verification tasks, achieving accuracies above 80% and Equal Error Rates (EER) often below 10%.

The study, suggested by Yasmeen Abdrabou et al. [48], explores using gaze behaviour to assess perceived password strength, aiming to help users select passwords resistant to guessing attacks. We propose incorporating users’ understanding of password strength into security mechanisms by analysing gaze patterns during password creation. In a proof-of-concept study ( $N = 15$ ), participants created both weak and strong passwords. Results show that password strength can be predicted from gaze behaviour with 86% accuracy using machine learning, paving the way for novel interfaces that encourage stronger password choices.

B. L. Tait [83] proposed a study in which 25 participants employed the EyeWriter eye-tracking system to assess whether individual eye movement behavior is sufficiently distinctive for secure biometric authentication. The results indicate that behavioral eye movements can be effectively employed for biometric verification, particularly in a one-to-one authentication context, achieving success rates between 96% and 100%. In contrast, one-to-many identification yielded an accuracy of only 21.83%. While this demonstrates some potential, the limited sample size restricts the generalizability of the findings and highlights the need for further investigation with larger populations.

With the emergence of Virtual Reality (VR) displays equipped with embedded eye trackers, eye-tracking data is increasingly accessible and can be leveraged for applications such as gaze monitoring, privacy protection, and user authentication. User identification is particularly critical in VR due to security and privacy concerns. Previous work has primarily relied on machine learning applied to motion-based data

(e.g., body, head, eye, and hand tracking), typically requiring explicit tasks and extensive feature sets. In contrast, Asish SM et al. [84] proposed a system for user identification based on minimal eye-gaze features, without the need for identification-specific tasks. To evaluate user identification, gaze data from an educational VR application were processed with machine learning models (RF, kNN) and deep learning models (CNN, LSTM). Results demonstrate that all models achieved identification accuracies exceeding 98% using only six simple gaze features. The findings underscore the potential of eye-gaze-based identification in VR, while emphasizing its security and privacy implications and acknowledging the limitations of the current study.

### 4.3 Methods based on eye/gaze velocity

Bednarik et al. [62] investigated the potential of eye velocity as a distinguishing feature for biometric purposes. They considered left-eye, right-eye, and combined velocity signals. However, the results indicated very limited discriminative power, with identification rates ranging only between 6% and 25%.

Kinnunen et al. [63] proposed an approach based on the distribution of gaze angles over short temporal windows. The local velocity direction of the eyes was estimated using trigonometric transformations and represented in normalized histograms. Seventeen participants watched a 25-minute video while gaze data were collected at 120 Hz. Each participant was modeled using Gaussian Mixture Models, following a methodology reminiscent of speaker recognition. Depending on the amount of training and testing data available, error rates varied between 29.4% and 47.1%, reflecting moderate performance.

Rigas et al. [64] examined velocity and acceleration as biometric indicators by applying the Wald–Wolfowitz statistical test to compare distributions of dynamic eye features. Their experimental design used a moving dot as stimulus, which jumped every 550 ms within grids of either  $3 \times 3$  (Dataset A) or  $2 \times 2$  (Dataset B) positions. Data were collected from 37 and 79 participants respectively, yielding 978 and 4168 signals. Fixations were extracted from these signals, and the classification results were encouraging, with identification rates of 96.6% for Dataset A and 90.4% for Dataset B.

Cuong et al. [65] introduced an alternative representation by encoding gaze features with Mel-Frequency Cepstral Coefficients (MFCCs). Features such as eye position, eye difference, and velocity were transformed into MFCC representations and classified using a single multi-class classifier, where the stimulus was represented by a moving point. This approach emphasized the potential of signal-processing methods borrowed from speech recognition for eye-movement biometrics.

Liang et al. [66] explored acceleration-based features, focusing particularly on diagonal movements and one and two step shifts in both horizontal and vertical directions. Five participants were instructed to repeatedly watch short video clips, and their eye movements were analyzed using both Back-Propagation neural networks and Support Vector Machine classifiers. The highest accuracy achieved in their work reached 82%, suggesting that acceleration properties can provide stable biometric signatures even with relatively small sample sizes.

Darwish and Pasquier [67] concentrated on saccadic dynamics as a biometric source. Their study recruited 22 participants, with eye movements recorded at 120 Hz across four sessions scheduled twice weekly. Stimuli included various patterns, such as a  $4 \times 4$  dot matrix. From the eye-movement data, several features were derived, including angular velocity and acceleration (for both left and right eyes) as well as absolute velocity. Mean, standard deviation, and maximum values of these metrics were then computed and used to train a Random Forest classifier with 10-fold cross-validation. This approach achieved strong results, with a Half Total Error Rate (HTER) of around 5%.

Holland and Komogortsev [68] also incorporated velocity-based features into their framework. Specifically, they extracted horizontal and vertical mean velocity along with peak velocities in both directions. These parameters were used as part of a larger feature set to test classification models across datasets collected with both high-frequency and low-cost trackers.

Yoon et al. [69, 70] proposed a classification framework using velocity categories modeled through Hidden Markov Models (HMMs). In their first study [69], eye movements were captured from 12 participants with

a 60 Hz eye tracker while they observed 250 still images composed of structured dot patterns (proximity grids, aligned grids, glass patterns, large shapes, and small shapes). Velocity data were grouped into five states: extremely low velocity, fixation, smooth pursuit, saccade, and extremely high velocity. Using HMMs with five hidden states, classification accuracies ranged from 53% to 76%, with higher performance observed for participants with lower average gaze velocity. The second study [70] extended the analysis to 16 participants tested over morning and afternoon sessions. Here, HMMs were individualized by selecting the best likelihood across 2–5 hidden states per user. While classification accuracies decreased to between 17% and 41%, the results confirmed that lower gaze velocity correlated with higher accuracy. They also noted better results for morning sessions, suggesting that fatigue might impair biometric consistency.

Juhola et al. [71] and Zhang et al. [72] focused on saccadic features for verification tasks. Juhola et al. compared data acquired from two eye trackers with very different sampling rates (400 Hz vs. 30 Hz) while participants followed a dot moving along a black bar. They extracted amplitude, accuracy, latency, and maximum velocity of saccades and tested Linear Discriminant Analysis, Quadratic Discriminant Analysis, and Naive Bayes classifiers. High-frequency data led to recognition rates of around 90%, whereas low-frequency recordings yielded 70–90%. Zhang et al. extended this line of work by analyzing longer-term stability, with test intervals ranging from half a day to 16 months, using a 30 Hz tracker. They added maximum acceleration and deceleration as features and employed classifiers such as Learning Vector Quantization, Discriminant Analysis, Naive Bayes, and Support Vector Machines. Verification experiments achieved up to 86% accuracy for genuine users and 88% for impostors with SVMs.

Srivastava et al. [73] combined velocity features with position and difference based features in an SVM multi-class classification framework. Using a moving point stimulus, their model consistently achieved high accuracy, exceeding 91% across different test scenarios.

Eberz et al. [74] also incorporated temporal descriptors of saccadic dynamics into their biometric framework, focusing particularly on speed and acceleration. These features were analyzed in conjunction with pupil size, thereby integrating both movement-based and physiological signals. By combining dynamic saccade characteristics with pupillary responses, their study aimed to capture a richer representation of user-specific visual behavior. This multimodal approach demonstrated that temporal aspects of saccades, when fused with pupil-based metrics, can enhance the distinctiveness of biometric profiles and contribute to more reliable user identification.

Ali, Hoque, and Deravi [45] proposed a presentation attack detection method based on directed gaze trajectories, where randomized visual stimuli are used to trigger natural pupillary motion, making spoofing attempts harder to succeed. Their system was tested with photo, 2D mask, and 3D mask artifacts on data collected from 80 volunteers. The results demonstrated that gaze trajectory features are effective in distinguishing genuine users from presentation attacks, confirming the potential of this approach for enhancing biometric security in practical applications.

Spoofing is a major challenge for biometric systems, and a gaze-based method using randomized visual stimuli has been proposed to detect such attacks. Ali et al. [46] propose a method based on the idea that natural head–eye coordination during gaze shifts creates unique accuracy patterns that are difficult for impostors to mimic. Data collected from genuine users and from attackers employing photographs, 2D masks, or replayed videos were analyzed using features derived from collinear and colocated gaze points. Experimental results, including both isolated feature evaluations and score-fusion combinations, demonstrated the effectiveness of the method in distinguishing between authentic and spoofed attempts.

Harezlak K et al. [49] evaluated biometric identification using eye movement signals by testing new features, including frequency-domain representations and the largest Lyapunov exponent, alongside velocities and accelerations. Eye movement data were recorded in 100-ms segments from 24 participants across two sessions, as they observed a point appearing at 29 screen locations. Feature vectors were created for single points and sequences of three points. Two training/testing scenarios were tested: one with 75% of randomly selected vectors and equal participant representation, and another using separate sessions for training and testing. Decision tree and random forest classifiers performed best, with random forest achieving up to 100% accuracy in the first scenario, while performance dropped significantly in the second.



## 4.4 Methods based on pupil size

Despite receiving relatively limited attention as a biometric trait, pupil size has emerged as a promising indicator for both identification and verification, as evidenced by several noteworthy investigations.

Bednarik et al. [62], previously discussed in relation to other ocular features, also examined the role of pupil size in biometric recognition. Their experimental setup included 12 participants exposed to four types of visual stimuli: text, a static cross, a moving cross, and an image. The features extracted comprised pupil size (left, right, and combined), delta pupil size (for left, right, and combined), eye velocity (left, right, and combined), and interocular distance. In addition, aggregate mean values of these attributes were examined, and the analysis employed Fast Fourier Transform (FFT), Principal Component Analysis (PCA), as well as their combined application. Performance was assessed using both individual features and weighted feature fusion. The best single-feature result was achieved with eye distance, yielding a 90% success rate. However, as this measure relates more closely to physical rather than behavioral characteristics, its biometric validity is limited. Among behavior-related features, delta pupil size achieved the strongest outcome (60%). When feature fusion was applied, combining eye distance with other features consistently resulted in accuracies exceeding 90%.

Nugrahaningsih and Porta [75] extended this line of inquiry by evaluating pupil size as a soft biometric in a longitudinal setting. Their study involved 25 participants tested across two sessions separated by one year. The experimental task was minimal, requiring subjects only to fixate on a plus sign displayed at the center of a white background. Key attributes included left and right pupil sizes, along with their ratio and difference. Each attribute was described by 15 statistical measures (e.g., minimum, maximum, mean, variance, median, geometric mean, harmonic mean, interquartile range, kurtosis, skewness, and range). Additional features included the sum of squared differences and the correlation between left and right pupil sizes. Classification was performed using four algorithms: Naïve Bayes, Neural Networks, Support Vector Machines (SVM), and Random Forests. Results indicated higher identification performance when selected subsets of features were employed (0.6194–0.7187), while verification performance remained consistently strong across both full and reduced feature sets ( $\approx 0.97$ ).

Eberz et al. [76] further explored the discriminative power of pupil size by incorporating it into a larger framework of 21 features. Pupil-based measures included maximum, minimum, mean, standard deviation, and range of pupil diameter. Participants were instructed to click as quickly as possible on red dots with white centers that appeared sequentially in random screen locations. Three experimental conditions were used: no prior instruction, textual guidance, and a visual preview of the dot locations. In addition to pupil features, temporal and spatial eye-movement features were also considered. Classification was performed using k-Nearest Neighbors (kNN) and Support Vector Machines. Across the three experimental settings, equal error rates (EER) ranged from 3.98% to 9.27%, underscoring the value of combining pupil size with dynamic ocular features for reliable biometric identification.

The following work, proposed by Casanova et al. [51], develops a web user identification model by analyzing behavioral biometrics from touch dynamics and periocular characteristics such as pupil movements, blinks, and fixations. A feature-level fusion approach is applied, combining different distance measures (Euclidean, Bray-Curtis, Manhattan, Canberra, Chebyshev, Cosine) to determine whether a test sample matches the target subject. Multi-data processing methods such as Canonical Correlation Analysis (CCA) and Principal Component Analysis (PCA) are used to further enhance performance. Results show that fusing these biometric traits significantly improves accuracy, achieving over 92%, surpassing individual biometrics.

## 4.5 Methods based on oculomotor features

Kasproski and Ober [77] investigated the use of Cepstral analysis, the inverse Fourier transform of the logarithm of a signal’s power spectrum, as a biometric feature. Eye movement data were collected with a 250 Hz tracker from nine participants instructed to follow a 3×3 jumping point stimulus, with each participant enrolled on 30 occasions. The dataset consisted of “probes” representing eye movement recordings during eight-second stimulations. From these data, 15 cepstral coefficients were extracted across four waveforms and subsequently used as input to four classifiers: K-Nearest Neighbors (KNN),

Naïve Bayes, Decision Tree, and Support Vector Machine (SVM). Classification performance was assessed using 10-fold cross-validation, yielding an average False Accept Rate (FAR) of 1.48% and a False Reject Rate (FRR) ranging between 6.7% and 43.3%, with an overall average of 22.59%.

Building on different physiological properties, Komogortsev et al. [78] examined oculomotor plant characteristics (OPC) as a verification biometric. Their method extracted saccadic parameters from eye-tracking data and transformed them into nine OPC components using a linear homeomorphic 2D oculomotor plant model. These components included length tension, series elasticity, passive viscosity, agonist and antagonist force-velocity relationships, tension intercept, agonist and antagonist tension slopes, and eye globe inertia. Two statistical techniques—Student’s t-test with Voting and Hotelling’s  $T^2$  test—were employed to measure similarity between OPC vectors, with an additional AND/OR fusion strategy to combine classifier outcomes. Using a 1000 Hz eye tracker, data were collected from 59 participants across two sessions separated by a 20-minute break, during which participants followed a jumping dot stimulus. The best result achieved was a minimum Half Total Error Rate (HTER) of 19 %, using both Hotelling’s  $T^2$  test and the OR Student’s t-test with Voting for horizontal saccadic data.

In a subsequent contribution, Komogortsev et al. [79] explored a multimodal ocular biometric framework by integrating three distinct traits: OPC, Complex Eye Movement (CEM) patterns, and the physical structure of the iris. While iris recognition is not the primary focus of this survey, it was included in their study alongside OPC and CEM features, both derived from eye movement dynamics captured via image sequences. A Sony PlayStation Eye camera was employed to record both iris and eye movement data, using Rorschach inkblot images as stimuli. Data from 87 participants were analyzed, and results demonstrated that combining these ocular traits improved overall accuracy and enhanced the resistance of the system against spoofing attempts.

Liang et al. [80] complemented their use of acceleration-based features by integrating muscle properties into the framework, with particular attention to seven “optimum muscle parameters.” These parameters were derived to capture the biomechanical characteristics of the oculomotor system, reflecting how individual differences in muscular control influence eye movement dynamics. By integrating these physiological descriptors into their biometric model, the authors demonstrated that muscle-based features can provide complementary information to traditional kinematic measures, thereby enhancing the distinctiveness and reliability of gaze-driven biometric identification.

Dillon Lohr et al. [42] investigated the fusion of two eye-centric authentication modalities—eye movements and periocular images—within a calibration-free framework. While both modalities have demonstrated promise individually, their integration in a unified pipeline had not been studied extensively at scale. Using a large in-house dataset of 9,202 subjects with eye-tracking quality comparable to consumer-grade VR devices, the authors showed that their multimodal approach consistently outperformed unimodal systems and surpassed the FIDO benchmark. The use of advanced machine learning models enabled the system to capture complementary discriminative features from both modalities, significantly enhancing authentication performance.

In 2023 Alessandro D’Amelio et al. [43] proposed a principled framework for modeling eye movements as a form of behavioral biometrics, drawing on foraging theory as a basis for capturing the uniqueness of individual gaze behavior. A composite Ornstein–Uhlenbeck process is employed to characterize the exploration–exploitation dynamics of eye movements, with parameters inferred from eye-tracking data through Bayesian analysis. These parameters serve as discriminative features for biometric identification, which is performed using standard classification techniques. A proof of concept on a publicly available dataset demonstrates the effectiveness of the approach, highlighting its potential both for biometric applications and for advancing the analysis of eye-tracking data.

The increasing need for secure digital access, coupled with technological advances and the lack of standardized solutions, continues to drive research into eye movement-based biometrics. This work, suggested by Harezlak and Plucienni[81], presents two neural network-based approaches to user identification from eye movement dynamics. In the first method, a 100-element time series was constructed to represent eye movement features, including velocity, acceleration, jerk, point-to-point percentage changes, and frequency-domain characteristics. These were modeled using a Long Short-Term Memory (LSTM) network. The second method employed the same dynamic features but derived statistical

descriptors from the time series, which were then classified using a dense neural network. Experiments were conducted with the publicly available GazeBase dataset, focusing on recordings obtained under the “jumping point” stimulus condition. Results demonstrate strong discriminative potential: the LSTM model with time series feature vectors achieved 96% accuracy, while the statistical feature-based method achieved 76%. These outcomes were observed across a three-year span of recordings, with analyses also considering different temporal segments and stimulus configurations.

The work suggested by Abbas Seha et al. [85] examines the biometric potential of eye movement patterns from low frame rate eye-tracking devices. It also explores recognition improvements through additional static and dynamic ocular features, such as eye blinking patterns and periocular shape characteristics. These modalities are relevant for applications like continuous driver authentication in law enforcement. Two datasets were collected using different low frame rate eye trackers, recording eye movements from 55 participants during real driving sessions. For eye gaze, fixation and saccade features—including duration, amplitude, and statistical descriptors—were extracted. Blinking features included patterns, speed, acceleration, and power-per-unit-mass profiles, while periocular features comprised eye-opening height, width, and axial ratio. Each modality was evaluated individually, then combined in a multi-modal fusion. While single modalities showed moderate performance, fusion of static and dynamic features significantly improved results, achieving up to a 98.5% recognition rate and 0% error rate. This demonstrates that, although single-modality approaches may lack robustness, multi-modal fusion offers high reliability for continuous driver authentication.

## 4.6 Methods based on head orientation

Cazzato et al. [32] introduced an alternative to traditional eye trackers by employing the depth and RGB streams of the Microsoft Kinect sensor. Their approach did not aim to precisely determine the exact gaze point. Instead, it relied on cases where head orientation could be treated as a reasonable approximation of gaze direction, followed by a rough estimation of pupil position. Because gaze could not be assessed in every video frame due to computational delays and occasional failures in detecting the face or pupil, the missing values were estimated using Kalman Filter predictions. The gaze detection process consisted of two stages. In the first stage, head pose was estimated using the depth and RGB data. In the second stage, pupil position was determined from the RGB stream. These two components were then combined to calculate the final X and Y gaze coordinates. Feature extraction was carried out using Principal Component Analysis (PCA), and the selected features were classified through a Minimax classifier. Experiments were conducted with 12 participants who watched three short video clips. The results showed that precision ranged from 81% to 83%, while recall ranged from 80% to 86%.

M. Wierzbowski et al. [50] conducted an investigation involving 43 participants who viewed multiple 360-degree VR videos while their eye-tracking data and headset orientation were recorded. Using both deep learning and classical statistical methods, the researchers tested whether a viewer of a previously unseen VR video could be identified based on prior viewing data. The experiments achieved a maximum average identification accuracy of 75%, demonstrating the potential for identifying VR users under unforced behavioral conditions.

This research, conducted by Rubo et al. [52], investigates the feasibility of implicit user identification in Social Virtual Reality (VR) based solely on eye gaze patterns during dyadic interactions. Employing Siamese Convolutional Neural Networks on data from 128 participants, the study achieved an identification accuracy of 95.50% by combining models for gaze during speaking and listening. Findings indicate that gaze patterns constitute distinctive, idiosyncratic behavioral traits, largely independent of self-reported personal characteristics. These results underscore significant privacy and security implications for Social VR environments, where gaze data may be collected passively under conditions of limited movement and variable data quality.

Identifying users of a Virtual Reality (VR) headset enables adaptive interface design, personalization of settings, and adjustment of difficulty levels in both gaming and training applications. While most existing identification methods require explicit input, implicit user identification offers a less intrusive alternative that preserves immersion. Liebers et al. [82] proposed a biometric identification system leveraging gaze

patterns and head orientation during stimulus tracking for user recognition. The method was validated through a user study, with a hybrid post-hoc analysis yielding identification accuracies of up to 75% using an interpretable machine learning model and up to 100% with a deep learning approach. Finally, potential application scenarios are discussed, demonstrating how this approach can enable seamless and implicit user identification in VR environments.

Recent research has explored identifying users in virtual reality (VR) based on behavioral data. These approaches often rely on elaborate authentication tasks resistant to “shoulder surfing” but overlook whether identification is possible under natural behavior. In the study by Wierzbowski et al. [86], involving 43 participants, the authors investigated whether a viewer of an unseen 360° VR video could be identified using eye-tracking and headset orientation data recorded while viewing other 360° VR videos. Using deep learning and classical statistical methods, we achieved up to 75% average identification accuracy, demonstrating the potential for VR user identification under unforced conditions.

## 5 Experiment on Gaze-based Biometric with dynamic text

Considering the advantages offered by behavioral biometric approaches, this research aims to investigate their effectiveness in the context of a novel application. In particular the thesis will focus on a specific field of research of this discipline called Gaze-based biometric. Gaze biometric is a field which investigate the behaviour from a user that reacts to a visual stimulus. In this study, we propose to record subjects gaze patterns using an eye-tracking device, capturing a range of descriptive features. The analysis will focus on identifying the specific points of visual attention and extracting meaningful information from these gaze behaviors. This approach not only allows for a detailed characterization of individual visual patterns but also facilitates the development of models to evaluate their performance in both identification and verification tasks. There are two types of stimuli that can be used for Gaze-based biometric: Static stimulus or dynamic stimulus. A static stimulus refers to an object displayed on the screen that remains fixed, such as a still image, whereas a dynamic stimulus involves an animation that changes its properties or behavior over time. In this work, we employed a dynamic stimulus consisting of moving text that changes its position over time. The stimulus was designed with two speed conditions (slow and fast) and two font sizes (small and large), resulting in a total of eight distinct animations. We selected dynamic text as our stimulus because this area remains largely unexplored in the context of gaze-based biometrics. In contrast, previous studies proposed by Holland et al. [54] and Bednarik et al. [62], have primarily focused on static text stimuli for biometric applications. The objective of this research is to apply machine learning techniques to evaluate the performance, for identification and verification, of these kinds of stimuli and determine the most effective animation. Although prior studies have examined static text, we also conducted a preliminary experiment with static text to enable a direct comparison, since the previously reported tests were carried out under different conditions than ours.

### 5.1 Set up

At the beginning, before starting the analysis, a phase used to collect the data and create the dataset was performed. The data collection involved 35 participants, including 22 males and 13 females, the majority of whom were university students. Of the participants, 17 were aged 25 or younger, 13 were between 26 and 40 years old, and 4 were over 40. Additionally, 18 wore eyeglasses, while only one wore contact lenses.

Age	20-30	31-40	41-50	51-60	>60
Number of Testers	26	4	1	3	1

Male	Female
22	13

Eye glasses	Contact lenses
18	1

Figure 5: Testers' data

A detailed overview is provided in Figure 5.

All recordings were carried out under identical artificial lighting conditions, with the same light source activated for each session. The experiments took place in a quiet environment, free from noise and distractions, using an eye tracker mounted on a monitor to record participants' gaze. No additional information regarding the task was provided to the participants aside from the instruction to read or follow the text, even when it appeared at higher speeds. Furthermore, no participant was exposed to any of the texts or animations prior to their first trial, ensuring unbiased and spontaneous responses

throughout the experiment.

The participant was seated on a height-adjustable chair and asked to observe the stimuli presented on the monitor. The monitor, a 23-inch display with a resolution of  $1920 \times 1080$  pixels (full-HD), was positioned in front of a white wall and equipped with an eye tracker. Research carried out in controlled laboratory settings with high-frequency eye trackers, often requiring the use of a chin rest, has demonstrated that very high recognition rates can be achieved [89]. However, such conditions are not realistic for practical applications, as high-frequency devices are costly and the use of a chin rest is inconvenient for everyday scenarios. In real-world contexts, lower-frequency eye trackers should be employed without physically constraining the user’s head. For this reason the eye tracker used was a GazePoint GP3, configured with a data sampling rate of 150 Hz, and employed in our experiments without head restraints. We also chose to conduct the recordings without calibration. While calibration is typically a short procedure, lasting between 15 and 30 seconds, it inevitably lengthens the overall authentication process and can become inconvenient or even frustrating for users. More importantly, although calibration is generally straightforward in controlled laboratory conditions, it is far less reliable in real-world contexts. Users may fail to remain perfectly still, or external factors such as background noise and the presence of other people can easily disrupt the process, rendering the calibration ineffective [35]. This decision, however, restricted the use of gaze coordinates as features in the analysis, due to the lower accuracy compared to a calibrated setup. Nevertheless, we consider this approach to be more practical and representative of real-world conditions, where performing calibration is often impractical and unsuitable for continuous or everyday authentication scenarios.

With regard to the experimental procedure, each participant completed nine trials, organized into three separate sessions (S1, S2, and S3). Each session was conducted on a different day, with the second and third sessions occurring no more than a month after the preceding ones. Consequently, the total testing period for each participant ranged from a minimum of 3 days to a maximum of 30 days, with an average duration of 10 days. During each session, participants completed the test three times, referred to as “trials” (T1, T2, and T3), with a minute breaks between trials. Before each session, participants were simply instructed to observe the content displayed on the screen without receiving any explanation of the experiment’s purpose. The user is positioned approximately 50–60 cm from the monitor, with a possible additional adjustment of up to 15 cm to ensure a comfortable viewing angle for the eye tracker.

## 5.2 Stimuli

Since the field of gaze-based biometrics using dynamic text is largely unexplored, it was challenging to determine which types of stimuli would be most suitable to select. The initial idea was to use a white rectangular block containing black text, displayed against a black background, which moved by bouncing off the screen borders until the animation time expired. At that stage, only four animations were implemented, each differing in the initial position of the block and the direction of its movement. In particular we had:

- **Horizontal moving – left start:** The block of text appeared at the middle-left of the screen ( $x = 0$ ,  $y = \frac{screen\_height}{2} - \frac{block\_height}{2}$ ) and initially moved to the right until it reached the border. At that point, it bounced back toward its initial position. This loop repeated until the animation’s time expired. It is important to note that the block moved exclusively along the x-axis, while the y-axis remained fixed.
- **Horizontal moving – right start:** This animation followed the same behavior as the *Horizontal moving – left start*, with the only difference being the starting position. In this case, the block of text appeared at the middle-right of the screen ( $x = screen\_width$ ,  $y = \frac{screen\_height}{2} - \frac{block\_height}{2}$ ) and initially moved to the left. The motion was restricted to the x-axis, while the y-axis and the bouncing strategy remained unchanged.
- **Diagonal moving – top-left start:** This animation used the same block of text as the previous cases. The block started in the top-left corner of the screen and moved diagonally toward the bottom-right corner. Upon reaching the boundary, it bounced back and returned to its initial position. This cycle was repeated continuously until the allotted time expired.

- **Diagonal moving – bottom-right start:** This animation followed the same pattern as *Diagonal moving – top-left start*, with the only difference being the starting position, which was the bottom-right corner of the screen.

However, after discussion, we were not fully convinced about these animations. While we did not expect their performance to be poor, we agreed that such animations represented an unusual form of dynamic text.

For this initial investigation in the field, we aimed to adhere to a classic format and employ a type of text that is most frequently encountered in everyday contexts. Consequently, we selected two types of animations for the study: Horizontal Scroll and Vertical Block. The Horizontal Scroll animation consists of a single line of text presented continuously. The text initially appears at the middle right of the screen and moves steadily toward the left. As the animation progresses, additional characters are gradually revealed, producing a continuous scrolling effect. This presentation is analogous to the running text commonly utilized in television news broadcasts, thereby providing a familiar stimulus for gaze-based biometric analysis. The Vertical Block animation consists of a block of text occupying approximately two-thirds of the screen width. The block moves steadily from the bottom toward the top of the screen in a continuous manner, similar to the credit roll seen at the conclusion of films. This animation offers a different motion pattern for gaze behavior analysis and serves as a complementary stimulus to the Horizontal Scroll animation, enabling a comparative evaluation of gaze dynamics in distinct movement contexts.

The previously mentioned animation types were further diversified to generate a total of eight distinct stimuli for the experiment. Specifically, each animation type was implemented with two different font sizes (large and small) and two movement speeds (slow and fast), resulting in eight unique animation conditions to be tested.

### 5.2.1 Preliminary experiments

This work also includes a preliminary experiment aimed at selecting appropriate text speed and size parameters. Given the defined animations, it was necessary to determine optimal values for these two variables. To address this question, we conducted a brief exploratory study in the laboratory involving seven participants.

Age	20-30	31-40	41-50	51-60	>60
Number of Testers	3	2	0	2	0

Male	Female
6	1

Eye glasses	Contact lenses
4	0

Figure 6: Testers' data

The demographic data of these participants are presented in Figure 6. The group consisted of six males and one female and four participants wore eyeglasses. Regarding age distribution, three were between 21 and 30 years old, two between 31 and 40 years old, and two between 51 and 60 years old.

The first study aimed to determine the minimum font size at which text remains comfortably readable on the screen during the animation. Participants were asked to select the smallest font size at which they felt confident in reading the text. In addition to the animated stimuli, a static text was presented

to participants to allow them to indicate their preferred font size. Rather than relying on specialized software, we manually selected a range of font sizes and displayed the animations to illustrate how each size would appear. We then solicited opinions from laboratory members to determine the minimum legible size, and based on this collective input, agreed upon the most suitable values. From these suggestions, we identified the largest value among the minimum sizes chosen, which was adopted as the baseline for the experiment. It should be noted that the selected font size may have been influenced by the age of certain participants, as variations in visual acuity and reading comfort can affect preferences. Although this size may appear relatively large, it was considered appropriate for the animation format, ensuring clarity and ease of reading under dynamic viewing conditions. This decision prioritizes both readability and consistency across participants, which is essential for the reliability of gaze-based biometric analysis. Consequently, the font size for the “small text” condition was set to 30, while the “large text” condition was defined as twice this size, at 60.

The objective of the second study was to determine the maximum speed at which a participant can still read text within our dynamic animations. For the purposes of this research, it was not essential for participants to read the entire text. Rather, it was sufficient that they could perceive and process at least a portion of it. This approach reflects the nature of gaze-based biometric authentication, where the goal is to verify an individual’s identity based on eye movement patterns rather than full comprehension of the displayed content. In practical applications, authentication can be successfully achieved without requiring the user to fully understand or read every element of the stimulus, as the biometric signature lies in the gaze behavior itself rather than in the complete reading of the text. Participants were instructed to select the highest speed at which they were still able to follow the text and comprehend it. Specifically, for the horizontal scroll animation, they were asked to read as much as possible, while for the vertical block animation, they were required to read at least the first four lines of text. For this purpose we developed a custom software application in Python to present both types of animations, namely horizontal scroll and vertical block, using the previously selected font sizes. This resulted in a total of four animation conditions: horizontal scroll with small font (HS-small), horizontal scroll with large font (HS-large), vertical block with small font (VB-small), and vertical block with large font (VB-large). During the experiment, participants were instructed to observe each animation and progressively increase its speed by pressing a designated key. They continued this process until the animation became too fast to follow. In this experiment, we deliberately set aside the requirement for participants to read the text comfortably, as our primary interest was not in reading comprehension but in determining the maximum speed at which the text could still be followed. At that point, participants had the option to revert to the previous speed setting and select it as their preferred maximum readable speed. This procedure allowed us to determine a personalized speed threshold for each animation condition, reflecting the highest speed at which participants could still read and track the text.

Rather than assigning a single uniform speed to all animations, we determined a distinct speed for each animation. This choice was motivated by the fact that the perceptual effort required to follow a moving text depends not only on its velocity but also on its format (horizontal scroll vs. vertical block) and font size. For example, a speed that is comfortable for a large-font vertical block may become unreadable when applied to a small-font horizontal scroll.

In our implementation, the scrolling speed of the text stimulus is defined as the number of pixels by which its position is updated at each frame of the animation. This means that speed is expressed in terms of pixels per frame inside our program. At first sight, the use of non-integer values for speed may seem counterintuitive, since pixels are discrete entities on the display and cannot be physically subdivided. The adoption of floating-point values, however, is a standard practice in computer graphics. While the rendering process ultimately positions the stimulus at integer pixel coordinates, the internal position of the object is stored as a floating-point variable. This allows the accumulation of fractional displacements across successive frames, such that sub-pixel movements are realized over time. Without this approach, the animation would appear jerky or imprecise at low speeds, as small displacements would be lost due to rounding. In this way, the use of decimal values for speed ensures the perceptual continuity of motion, despite the discreteness of the display grid.

It is important to note that the measure of speed in pixels per frame is directly dependent on the refresh rate of the display. In the present setup, the frame rate is explicitly fixed at 150 frames per second. This choice is not arbitrary, but rather motivated by the need to synchronize the visual stimulus with the



Tester Number	HS-LITTLE	HS-BIG	VB-LITTLE	VB-BIG
1	3.9	4.8	0.7	0.7
2	1.1	1.9	0.8	0.8
3	2.7	4.0	0.8	1.4
4	2.6	4.0	0.6	0.6
5	1.8	2.6	0.4	0.8
6	1.7	1.7	0.3	0.6
7	2.9	4.7	0.8	1.3

AVG (px/frame)	$\simeq 2.38$	$\simeq 3.4$	$\simeq 0.58$	$\simeq 0.94$
----------------	---------------	--------------	---------------	---------------

AVG (Px/s)	357.0	510.0	87.0	141.0
------------	-------	-------	------	-------

Figure 7: Speeds data per animation

eye-tracking device, which samples gaze data at the same frequency of 150 Hz. Such temporal alignment ensures that each frame of the animation corresponds precisely to one sample of eye-tracking data, thereby facilitating a consistent and interpretable mapping between stimulus motion and recorded gaze behavior.

To convert the per-frame displacement into a more conventional measure of velocity, namely pixels per second, it is sufficient to multiply the displacement per frame by the number of frames per second. Following this procedure, a speed of 2.38 pixels per frame corresponds to 357 pixels per second, while values of 3.4, 0.58, and 0.94 pixels per frame translate to 510, 87, and 141 pixels per second, respectively. These values provide an interpretable indication of the actual velocity of the stimulus as perceived on the screen. All these values are summarized in Figure 7, which first presents the speeds chosen by each participant for each animation, and then reports the average of these speeds expressed both in pixels per frame and in pixels per second.

From the perspective of the participant, these velocities determine the apparent rate at which the text traverses the display from right to left. Higher values, such as 357 or 510 pixels per second, result in a rapid motion where the text crosses the screen in a relatively short amount of time, demanding faster eye movements to track it. In contrast, lower values such as 87 or 141 pixels per second correspond to a slower drift of the text, allowing for more prolonged fixations and smoother pursuit. Thus, expressing the stimulus velocity in terms of pixels per second not only provides a standardized physical measure of speed but also offers a direct connection to the perceptual experience of the task and to the temporal dynamics of oculomotor responses.

Although in the present study the frame rate was fixed at 150 Hz to match the sampling rate of the eye tracker, the same principle applies to different refresh rates. If the animation were instead run at 60 Hz or 120 Hz, the conversion from pixels per frame to pixels per second would need to be recalculated accordingly, leading to proportionally different velocities for the same per-frame displacement. This illustrates that the absolute value of stimulus speed cannot be interpreted independently of the frame rate, and highlights the importance of reporting both parameters in experimental contexts.

In conclusion we decided to adopt the mean speeds computed for each type of animation, averaged across the data collected from the seven participants, as the reference values for the “fast” condition of the experiment. This approach allowed us to derive a standardized measure of fast speed that reflects the natural preferences of the participants while ensuring consistency across stimulus types. For the “slow” condition, we chose to set the speed at exactly half of the corresponding fast speed for each animation. This decision was motivated by the desire to create a clear and perceptually meaningful distinction between the two conditions, while maintaining proportional scaling across animations. Subsequently, we implemented these values in the experimental program and performed a qualitative verification of the resulting motion. The observed scrolling speeds were found to be perceptually acceptable and consistent

with the intended experimental design, providing confidence that the chosen parameters would yield reliable and interpretable results in subsequent testing.

### 5.2.2 Animations

In the previous chapter, we conducted a preliminary study to define the parameters of the experimental setup—specifically, the text speed and character size—ensuring that these values were selected based on empirical evidence rather than chosen arbitrarily. This preparatory step was essential to establish a solid methodological foundation for the subsequent experiments. Having established the parameters, we can now present and describe in detail the eight animations used as stimuli in our gaze-based biometric research. These were systematically constructed by combining the two animation types (Horizontal Scroll and Vertical Block) with two font sizes (small and big) and two movement speeds (slow and fast). This design yielded eight distinct conditions, providing a comprehensive basis for evaluating the effectiveness of dynamic text as a biometric modality.

All the animations are detailed below:

- **VB\_SL\_LIT**: Vertical Block animation with *slow speed* and *small text size*. In this condition, the block of text scrolls upward from the bottom to the top of the screen at a reduced speed, with characters displayed at the minimum legible size identified during the preliminary study. Here, the parameters correspond to a font size of 30 and a scrolling speed of 0.29 px/frame.
- **VB\_SL\_BIG**: Vertical Block animation with *slow speed* and *large text size*. Similar to VB\_SL\_LIT, but the increased character size facilitates readability, while maintaining the slower scrolling speed. In this case, the configuration employs a font size of 60 with a speed of 0.47 px/frame.
- **VB\_FA\_LIT**: Vertical Block animation with *fast speed* and *small text size*. In this case, the upward scrolling occurs at the maximum speed tolerable for participants, with the minimum character size, representing a challenging condition for gaze tracking. The setting combines a font size of 30 with a speed of 0.58 px/frame.
- **VB\_FA\_BIG**: Vertical Block animation with *fast speed* and *large text size*. The rapid upward motion is combined with larger text, allowing participants to more easily perceive and follow the characters despite the increased pace. This version uses a font size of 60 together with a scrolling rate of 0.94 px/frame.
- **HS\_SL\_LIT**: Horizontal Scroll animation with *slow speed* and *small text size*. The text enters from the right edge of the screen and scrolls steadily toward the left at a slower rate, with characters displayed at the smaller size. The parameters here are a font size of 30 and a speed of 1.19 px/frame.
- **HS\_SL\_BIG**: Horizontal Scroll animation with *slow speed* and *large text size*. This condition presents the same horizontal scrolling movement, but with larger text, facilitating readability at the slower pace. The chosen configuration corresponds to a font size of 60 with a scrolling speed of 1.7 px/frame.
- **HS\_FA\_LIT**: Horizontal Scroll animation with *fast speed* and *small text size*. The text moves quickly from right to left, with minimal character dimensions, providing the most visually demanding condition in terms of both speed and size. In this setup, the font size is 30 and the speed reaches 2.38 px/frame.
- **HS\_FA\_BIG**: Horizontal Scroll animation with *fast speed* and *large text size*. In this case, the high-speed scrolling is mitigated by the increased character dimensions, offering a balance between perceptual clarity and dynamic movement. The applied parameters are a font size of 60 combined with a scrolling speed of 3.4 px/frame.

An important aspect to consider in the design of the experiment concerns the textual material used in the animations. Since the primary objective of this research is to investigate gaze-based biometrics, it was

crucial to ensure that the chosen text did not introduce biases related to content interest or familiarity. If certain passages were perceived as more engaging, informative, or familiar than others, this could have inadvertently affected participants' reading behavior and consequently distorted the analysis. To minimize this risk, we opted for neutral and homogeneous material by selecting descriptive texts taken from Wikipedia entries of municipalities in Lombardy, Italy.

A total of 72 unique texts were employed, corresponding to the experimental design ( $8 \text{ animations} \times 3 \text{ trials} \times 3 \text{ sessions}$ ). This ensured that each stimulus presentation featured a distinct text, thereby minimizing the risk of repetition effects or memorization influencing the results. To further enhance methodological rigor, texts were randomly selected for each trial, ensuring that each participant was exposed to a unique sequence of stimuli. This approach prevented systematic biases and increased the generalizability of the findings by diversifying the conditions across participants. Similarly, the order of animation presentation was randomized for every trial, mitigating potential sequence effects such as learning, fatigue, or adaptation. To manage these randomizations and guarantee appropriate stimulus distribution, the experimental software integrated an SQL database to track previously displayed texts. This system automatically assigned unused texts for each trial, ensuring that all participants experienced varied and unbiased stimuli. By combining randomized text selection with randomized animation order, the design maintained both experimental consistency and stimulus neutrality across sessions and participants.

Prior to the presentation of each animation within a trial, a fixation stimulus consisting of a white cross displayed on a black background was shown. This cross was positioned at the center of the screen to act as a visual anchor, directing participants' gaze to a consistent starting point before the animation began. The inclusion of this fixation stimulus serves two important purposes. First, it standardizes the initial gaze position across trials and participants, which reduces variability that could affect gaze tracking accuracy. Second, it ensures that participants are focused and visually prepared for the upcoming stimulus. This procedure enhances the reliability of the data by controlling for differences in initial eye position, which is especially critical in gaze-based biometric research where precise tracking of eye movement patterns is essential.

In conclusion each trial comprised eight animations, each with a duration of 15 seconds. Before each animation, a fixation cross was presented for 2 seconds at the center of the screen to guide the participant's gaze and establish a consistent starting point. The total duration for a single trial was therefore  $(8 \times 15) + (8 \times 2) = 136$  seconds per trial. During the same day, each tester completed three trials, each lasting 136 seconds, for a total of  $136 \times 3 = 408$  seconds, or approximately 6 minutes and 48 seconds. To complete the three sessions, the tester engaged in  $136 \times 3 \times 3 = 1,224$  seconds, which corresponds to a total of 20 minutes and 24 seconds of experimental exposure.

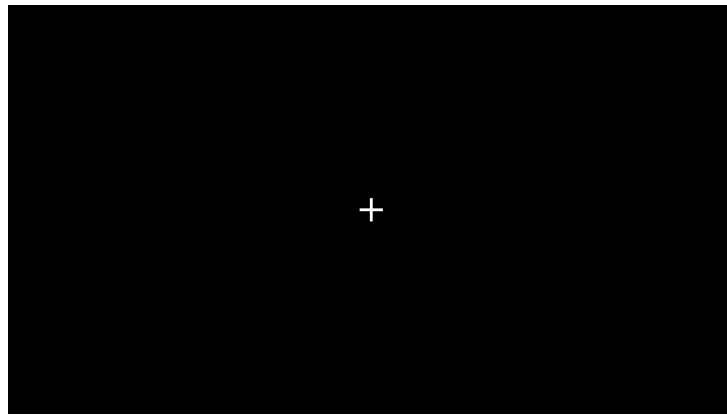


Figure 8: White cross animation

Figure 8 illustrates the white fixation cross that was presented prior to the onset of each animation.

Figure 9 provides a representation of the two types of animations employed in the study. Each animation type was implemented at two distinct speeds and with two font sizes, resulting in a total of eight unique animation conditions.

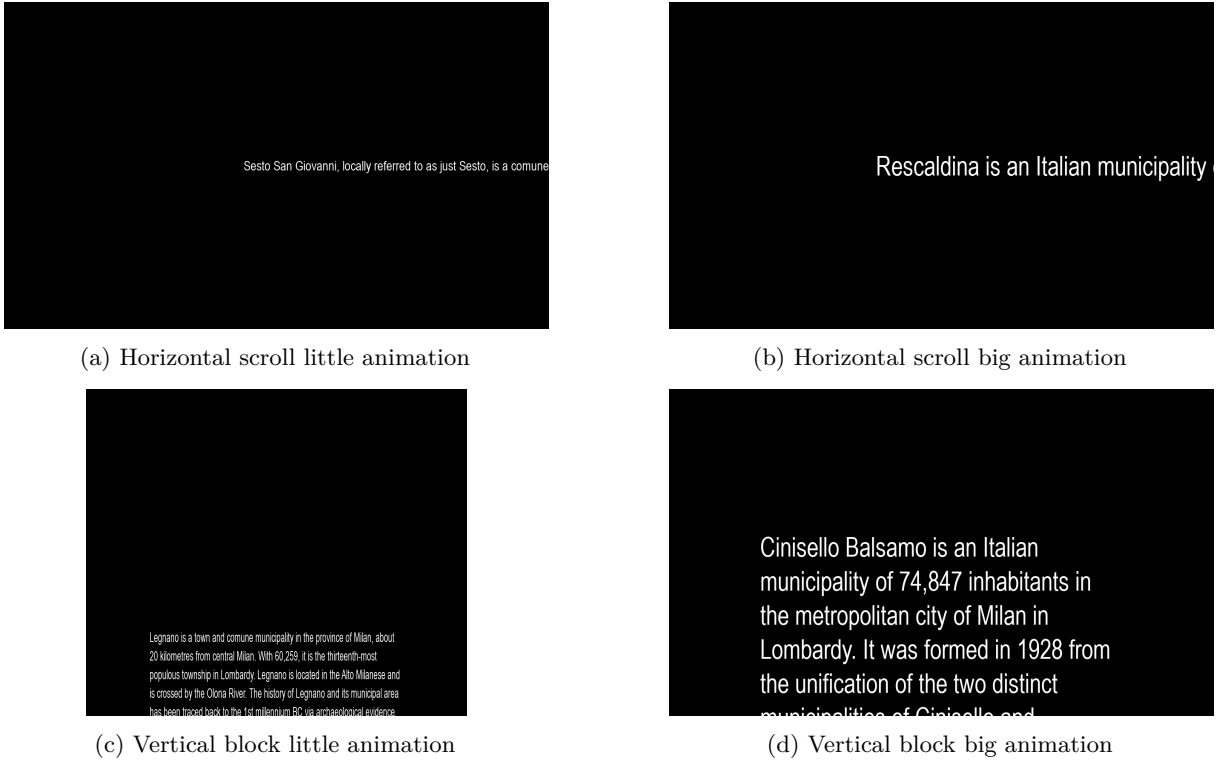


Figure 9: Different types of animations

### 5.3 Captured Eye Tracking Features

In the present study, the set of ocular features was recorded using the GazePoint GP3 eye tracker, which outputs a wide range of parameters designed to characterize visual behavior with both spatial and physiological precision. These features, according to [90], can be organized into several functional categories, namely fixation parameters, raw gaze coordinates, corneal reflection and pupil metrics, three-dimensional eye position, pupil dynamics, and blink-related measures. Each category not only describes a specific aspect of eye activity but also provides potential discriminative cues for biometric modeling.

The first group of features concerns fixations, which are intervals in which the eyes remain relatively stable on a spatial location, thereby enabling information intake and cognitive processing. The GP3 provides normalized screen coordinates of each fixation (FPOGX, FPOGY), which indicate the point of regard on the display. Alongside this spatial information, the system records the onset time of the fixation (FPOGS), its duration (FPOGD), and a unique identifier (FPOGID) for reference across the data stream. An additional validity flag (FPOGV) specifies the confidence of the measurement. Fixation duration is particularly important, as longer fixations are often associated with increased cognitive load or the processing of more complex visual material, while shorter ones may indicate scanning behavior. The spatial distribution of fixations across the screen, in turn, reflects attentional allocation and search strategies. Together, these measures provide a temporal and spatial fingerprint of how an individual interacts with visual stimuli.

Complementing the fixation measures are raw gaze coordinates, which provide the instantaneous location of the eyes on the screen without aggregation into fixation events. These are reported separately for the left (LPOGX, LPOGY, LPOGV) and right eyes (RPOGX, RPOGY, RPOGV), as well as for a binocular average (BPOGX, BPOGY, BPOGV). By distinguishing between monocular and binocular tracking, the system allows for evaluation of convergence behavior (i.e., how well the two eyes align on a target) as well as interocular differences, which may emerge from physiological asymmetries or small errors in ocular coordination. The binocular values represent a fused estimate of the point of regard, often more stable than monocular readings, but the monocular data themselves may carry subtle inter-eye differences that

are individual-specific and thus of biometric relevance.

The third group of features relates to corneal reflection and pupil properties, which are detected by the infrared illumination system of the tracker. For each eye, the GP3 reports the estimated coordinates of the pupil center on the image plane (LPCX, LPCY for the left eye; RPCX, RPCY for the right eye), along with values describing pupil diameter and pupil size (LPD, LPS for the left eye; RPD, RPS for the right eye). Each of these measurements is associated with a validity flag (LPV, RPV), indicating whether the tracking conditions were sufficient for reliable recording. Pupil size is not static: it responds dynamically to changes in illumination, but is also modulated by cognitive and affective processes such as attention, emotional arousal, and mental workload. These fluctuations are partially involuntary, making them difficult to consciously manipulate, and thus potentially stable individual signatures. In addition, the spatial estimation of the pupil center contributes to precise gaze localization, while differences in pupil metrics across individuals can form part of a biometric profile.

Beyond two-dimensional gaze data, the GP3 also captures three-dimensional eye position features, reporting the x-, y-, and z-coordinates of each eye relative to the eye tracker (LEYEX, LEYCY, LEYEZ for the left eye; REYEX, REYCY, REYEZ for the right eye). The x and y values provide lateral and vertical offsets, while the z coordinate specifies the distance between the eye and the tracker, effectively measuring head-to-screen distance. These measures are particularly valuable because they reveal natural head movements, posture adjustments, or minor asymmetries in eye alignment. While such variability might traditionally be considered noise, in the context of biometrics, these natural tendencies may carry person-specific information. For example, an individual may habitually lean closer to or further from the screen, or present consistent horizontal offsets due to ocular dominance.

Closely related to pupil size are the pupil diameter dynamics, reported as LPUPILD and RPUPILD for the left and right eyes, each paired with validity indicators (LPUPILV, RPUPILV). These values offer fine-grained recordings of moment-to-moment changes in pupil diameter, complementing the more general pupil size parameters described earlier. Unlike fixation duration or gaze coordinates, which directly reflect attentional allocation, pupil dynamics provide a physiological signal that is largely involuntary and influenced by a combination of environmental, cognitive, and emotional factors. The stability of these responses across repeated sessions, combined with their resistance to voluntary control, makes them a promising source of biometric information.

Finally, the dataset includes blink-related features, which capture involuntary eyelid closures that briefly interrupt gaze tracking. For each blink, the GP3 assigns a unique identifier (BKID), records its duration (BKDUR), and stores the minimum pupil size observed during the event (BKPMIN), which reflects the degree of occlusion. Blink dynamics are an integral part of ocular behavior, influenced by ocular physiology, fatigue levels, environmental factors, and cognitive state. However, inter-individual variability in blink frequency, duration, and regularity has been documented, suggesting that blinking patterns may themselves provide discriminative information. Importantly, because blinks are largely automatic and difficult to consciously regulate in a consistent manner, they can serve as a robust biometric trait resistant to intentional manipulation.

Taken together, the features recorded by the GP3 provide a comprehensive view of ocular activity. These measurements span spatial aspects of gaze location, temporal aspects of fixations and blinks, as well as physiological characteristics related to pupil dynamics and three-dimensional eye positioning. While some features directly represent visual exploration of stimuli (e.g., gaze coordinates, fixation positions), others capture involuntary physiological responses (e.g., pupil size fluctuations, blink dynamics) or natural head/eye geometry (3D coordinates). The integration of these diverse signals is particularly valuable in a biometric context, as it enables the system to combine behavioral and physiological aspects of eye activity, thereby enhancing robustness, discriminative power, and resistance to spoofing.

## 5.4 Dataset creation

In the current section, we describe the methodology employed to process the gaze-related features introduced in the previous chapter and transform them into the analytical features of the dataset. It is important to note from the outset that the raw gaze coordinates along the x and y axes were not included as direct features. This decision stems from the absence of a calibration procedure, which is intentional. The aim is to design a system suitable for practical, everyday applications where calibration may not be feasible. While this approach enhances usability and applicability in real-world scenarios, it comes at the cost of a slight reduction in the precision of gaze coordinate measurements. In this study, a total of 83 features were derived from the raw gaze and pupil data recorded by the GazePoint GP3 eye tracker.

### 5.4.1 Fixation-Based Features (f0–f10)

Fixations are periods during which the gaze remains relatively stable, enabling detailed visual processing of stimuli. This group of features captures both the frequency and statistical properties of fixation durations.

Feature  $f_0$  represents the total number of fixations recorded during a trial for each animation, reflecting the frequency with which gaze stabilization occurs.

All remaining fixation-based features ( $f_1$ – $f_{10}$ ) are derived from fixation durations, which are calculated from the relative gaze feature FPOGD for each fixation's unique identifier (FPOGID). Statistical descriptors of these durations provide insights into attentional control and scanning behavior:

- $f_1$  — **Minimum fixation duration:** The shortest fixation duration in the trial, indicating rapid scanning behavior.
- $f_2$  — **Maximum fixation duration:** The longest fixation duration, showing sustained attention to a specific location.
- $f_3$  — **Arithmetic mean fixation duration:** Average fixation time, representing typical gaze stability.
- $f_4$  — **Geometric mean fixation duration:** A measure of central tendency less sensitive to extreme values, useful for skewed distributions.
- $f_5$  — **Median fixation duration:** Middle value of fixation durations, robust against outliers.
- $f_6$  — **Standard deviation fixation duration:** Variability in fixation times across the trial.
- $f_7$  — **Median absolute deviation fixation duration:** Robust measure of variability, less influenced by extreme durations.
- $f_8$  — **Skewness fixation duration:** Asymmetry in the distribution of fixation durations.
- $f_9$  — **Interquartile range fixation duration:** Dispersion between the 25th and 75th percentile, excluding extremes.
- $f_{10}$  — **Kurtosis fixation duration:** Describes the "peakedness" of the fixation duration distribution.

### 5.4.2 Fixation Distance Features (f11–f20)

These features quantify spatial distances between consecutive fixations, characterizing gaze exploration and scanning behavior.

Distances are computed from the normalized gaze coordinates, FPOGX and FPOGY, for each fixation. For each pair of consecutive fixations  $(x_i, y_i)$  and  $(x_{i+1}, y_{i+1})$ , the Euclidean distance  $d_i$  is calculated as:

$$d_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}.$$

These distances form a distribution for each trial, from which statistical descriptors are extracted to capture the spatial patterns of gaze movement:

- $f_{11}$  — **Minimum distance**: Smallest shift between consecutive fixations, indicating very local gaze changes.
- $f_{12}$  — **Maximum distance**: Largest shift, reflecting broad scanning movements.
- $f_{13}$  — **Arithmetic mean distance**: Average spatial shift, representing typical scanning magnitude.
- $f_{14}$  — **Geometric mean distance**: Central tendency less influenced by extreme distances.
- $f_{15}$  — **Median distance**: Middle value of fixation-to-fixation distances.
- $f_{16}$  — **Standard deviation distance**: Variability of spatial shifts.
- $f_{17}$  — **Median absolute deviation distance**: Robust measure of spatial variability.
- $f_{18}$  — **Skewness distance**: Asymmetry of the distance distribution.
- $f_{19}$  — **Interquartile range distance**: Dispersion measure excluding extreme values.
- $f_{20}$  — **Kurtosis distance**: Shape of the spatial distance distribution.

#### 5.4.3 Saccade Speed Features (f21–f30)

Saccades are rapid eye movements occurring between fixations, responsible for shifting the line of sight toward new visual targets. To quantify their dynamics, saccade speeds are computed using the normalized gaze coordinates FPOGX and FPOGY, combined with temporal markers from the gaze data. For each pair of successive fixation points, the Euclidean distance between consecutive gaze positions is first determined. The corresponding time interval is calculated as the difference in starting times of consecutive fixations,  $FPOGS_{i+1} - FPOGS_i$ , adjusted by subtracting the duration of the first fixation  $FPOGD_i$ . This adjustment ensures that the interval reflects only the transition time between fixations rather than the fixation itself. Saccade speed for the transition is then defined as the ratio between traveled distance and elapsed time:

$$v_i = \frac{\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}}{t_{i+1} - t_i - d_i},$$

where  $(x_i, y_i)$  and  $(x_{i+1}, y_{i+1})$  denote successive fixation coordinates,  $t_i$  the start time of fixation  $i$ , and  $d_i$  its duration. This procedure yields a distribution of saccade speeds for each trial, which is then summarized through a set of descriptive statistics (f21–f30).

The extracted features are:

- **f21 — Minimum saccade speed**: The lowest velocity observed among saccades, indicating the slowest gaze transitions.
- **f22 — Maximum saccade speed**: The highest velocity recorded, reflecting the most abrupt shifts in gaze.
- **f23 — Arithmetic mean saccade speed**: The average speed across all saccades within the trial.
- **f24 — Geometric mean saccade speed**: Central tendency of saccade speeds, less influenced by extreme values.
- **f25 — Median saccade speed**: The middle speed value, robust against outlier influence.
- **f26 — Standard deviation saccade speed**: The variability of saccadic velocities across the trial.

- **f27 — Median absolute deviation saccade speed:** A robust measure of variation in saccade dynamics.
- **f28 — Skewness saccade speed:** The asymmetry of the saccade speed distribution, indicating bias toward slower or faster movements.
- **f29 — Interquartile range saccade speed:** The spread of saccade velocities within the 25th–75th percentile range.
- **f30 — Kurtosis saccade speed:** The peakedness or tail heaviness of the speed distribution.

#### 5.4.4 Scanpath Feature (f31)

The scanpath length provides a global measure of the total extent of visual exploration during a trial. It is computed as the cumulative sum of Euclidean distances between consecutive fixation points, based on their normalized screen coordinates FPOGX and FPOGY. Formally, given a sequence of fixation points  $(x_i, y_i)$ , the scanpath length  $L$  is defined as:

$$L = \sum_{i=1}^{N-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2},$$

where  $N$  is the total number of fixation points in the trial. This measure integrates all individual fixation-to-fixation transitions into a single trajectory length, capturing the overall amount of gaze displacement. A longer scanpath indicates more extensive or dispersed exploration of the visual stimulus, while a shorter scanpath reflects more localized or focused attention.

The feature extracted is:

- **f31 — Scanpath length:** Total cumulative distance of gaze movements during a trial, representing the overall extent of visual exploration.

#### 5.4.5 Pupil Diameter Features (f32–f51)

Pupil diameter is a physiological indicator influenced not only by luminance but also by cognitive and emotional states. In the present work, pupil diameter values were extracted separately for the left and right eyes, using the gaze features LPD and RPD. To ensure reliability of the measurements, only samples marked as valid by the corresponding validity flags ( $LPV = 1$  or  $RPV = 1$ ) were included in the analysis. For each eye, the collected pupil diameters across the trial were summarized using a set of descriptive statistics, resulting in two parallel groups of features:  $f32$ – $f41$  for the left eye and  $f42$ – $f51$  for the right eye. This approach allows for both independent characterization of each eye and subsequent comparison across them.

The features are defined as follows:

- **f32, f42 — Minimum pupil diameter:** Smallest observed pupil size, reflecting momentary constriction.
- **f33, f43 — Maximum pupil diameter:** Largest observed pupil size, reflecting dilation.
- **f34, f44 — Arithmetic mean pupil diameter:** Average pupil size, representing typical diameter across the trial.
- **f35, f45 — Geometric mean pupil diameter:** Alternative measure of central tendency, less sensitive to outliers.



- **f36, f46 — Median pupil diameter:** Middle value of pupil size distribution, robust to extreme values.
- **f37, f47 — Standard deviation pupil diameter:** Variability in pupil size across the trial.
- **f38, f48 — Median absolute deviation pupil diameter:** Robust measure of dispersion, less influenced by extreme dilation or constriction.
- **f39, f49 — Skewness pupil diameter:** Asymmetry of the pupil size distribution, indicating bias toward constriction or dilation.
- **f40, f50 — Interquartile range pupil diameter:** Spread of the middle 50% of pupil sizes, unaffected by extreme values.
- **f41, f51 — Kurtosis pupil diameter:** Shape descriptor reflecting peakedness or tail heaviness in the pupil size distribution.

#### 5.4.6 Pupil Diameter Ratio Features (f52–f61)

In addition to analyzing the absolute pupil diameters of the left and right eyes, the present study also considered their ratio as a derived measure. The pupil diameter ratio was computed as the quotient between the left and right pupil diameters (LPD/RPD), including only samples where both measurements were valid (LPV = 1 and RPV = 1). This ratio captures relative differences between the two eyes, which may reveal asymmetries in ocular physiology or subtle individual traits. The resulting distribution of ratio values across each trial was summarized through descriptive statistics, generating features *f52–f61*.

The individual features are defined as follows:

- **f52 — Minimum ratio:** The smallest observed value of the LPD/RPD ratio, indicating moments when the left pupil was relatively more constricted.
- **f53 — Maximum ratio:** The largest observed ratio, reflecting moments when the left pupil was relatively more dilated.
- **f54 — Arithmetic mean ratio:** The average ratio, providing a global estimate of relative pupil size balance.
- **f55 — Geometric mean ratio:** Central tendency of the ratio distribution, less influenced by extreme deviations.
- **f56 — Median ratio:** The middle ratio value, robust to outliers.
- **f57 — Standard deviation ratio:** Variability of relative pupil sizes across the trial.
- **f58 — Median absolute deviation ratio:** Robust measure of dispersion in the ratio values, resistant to outliers.
- **f59 — Skewness ratio:** Degree of asymmetry in the ratio distribution, indicating whether the left pupil tends to be systematically larger or smaller than the right.
- **f60 — Interquartile range ratio:** Range of the middle 50% of ratio values, excluding extremes.
- **f61 — Kurtosis ratio:** Descriptor of distribution shape, highlighting whether the ratios are tightly clustered or prone to extreme values.

By focusing on ratios rather than absolute diameters, these features emphasize interocular balance and asymmetry, providing an additional dimension of information that complements absolute pupil measurements.

#### 5.4.7 Pupil Diameter Difference Features (f62–f70)

Complementing the ratio-based analysis, another set of features was derived from the direct difference between left and right pupil diameters, computed as

$$D = \text{LPD} - \text{RPD},$$

where only samples with valid measurements for both eyes ( $\text{LPV} = 1$ ,  $\text{RPV} = 1$ ) were considered. This difference provides a straightforward measure of interocular asymmetry, capturing the extent to which one pupil is larger than the other at a given moment. Positive values indicate instances where the left pupil was larger, whereas negative values correspond to cases where the right pupil was larger. From the distribution of these differences, statistical descriptors were extracted, yielding features  $f62$ – $f70$ . Unlike other feature groups, the geometric mean was not included here, since the distribution of differences may contain negative values, for which the geometric mean is mathematically undefined.

The used features are:

- **f62 — Minimum difference:** The most negative value of  $\text{LPD} - \text{RPD}$ , corresponding to moments when the right pupil was much larger.
- **f63 — Maximum difference:** The largest positive value, when the left pupil exceeded the right in diameter.
- **f64 — Arithmetic mean difference:** The average signed difference, indicating overall bias toward one eye.
- **f65 — Median difference:** The central value of the distribution, robust to outliers.
- **f66 — Standard deviation difference:** Variability of interocular pupil differences.
- **f67 — Median absolute deviation difference:** Robust measure of dispersion for pupil size differences.
- **f68 — Skewness difference:** Asymmetry of the distribution, showing systematic tendencies toward one eye being larger.
- **f69 — Interquartile range difference:** Spread of the middle 50% of differences, insensitive to extremes.
- **f70 — Kurtosis difference:** Shape of the distribution, indicating whether values cluster tightly or include frequent large deviations.

#### 5.4.8 Blink Features (f71–f74)

Blinking constitutes an involuntary yet highly individual ocular behavior, influenced by physiological and cognitive factors such as fatigue, attentional demand, and emotional state. In the present analysis, blink-related features were derived from the identifiers and durations recorded by the GP3 eye tracker.

The total number of blinks in a trial ( $f71$ ) was determined by counting the distinct blink identifiers (BKID), restricted to values greater than zero to exclude missing or invalid entries. This feature provides a measure of blink frequency, which can vary substantially across individuals and contexts.

In addition, the temporal properties of blinks were captured using the duration values (BKDUR), measured in milliseconds. Only valid, nonzero durations were considered. From this distribution, three descriptors were extracted: the mean blink duration ( $f72$ ), the minimum blink duration ( $f73$ ), and the maximum blink duration ( $f74$ ). Together, these features describe not only the overall temporal profile of blinks but also their variability within a trial, reflecting both individual oculomotor tendencies and task-related influences.

Formally, the features are defined as follows:

- **f71 — Number of blinks:** Total count of distinct blink events during the trial.
- **f72 — Blink mean duration:** Average duration of recorded blinks.
- **f73 — Blink minimum duration:** Shortest recorded blink, representing rapid eyelid closures.
- **f74 — Blink maximum duration:** Longest recorded blink, corresponding to prolonged eyelid closures.

#### 5.4.9 Directional Movement Features (f75–f82)

In addition to fixations, saccades, and blinks, gaze dynamics can also be characterized by the predominant directions of movement across the visual field. These directional features capture the extent and frequency of horizontal and vertical displacements, providing insight into scanning strategies, asymmetries, and biases in exploration patterns.

For each pair of consecutive gaze points within a trial, the horizontal ( $\Delta x$ ) and vertical ( $\Delta y$ ) displacements were computed. The direction of movement was determined by the sign of the displacement: a negative  $\Delta y$  corresponds to upward movement, a positive  $\Delta y$  to downward movement, a negative  $\Delta x$  to leftward movement, and a positive  $\Delta x$  to rightward movement. For each direction, two types of measures were accumulated: (i) the total distance traveled in that direction, obtained as the sum of the absolute displacements, and (ii) the total number of directional transitions.

Formally, for a sequence of gaze points  $(x_i, y_i)$  with  $i = 1, \dots, N$ , the displacement vectors are given by:

$$\Delta x_i = x_{i+1} - x_i, \quad \Delta y_i = y_{i+1} - y_i.$$

Directional totals are then computed by summing the absolute values of displacements in each respective direction, while counts are obtained by tallying the number of occurrences where the displacement sign meets the condition for that direction.

The features are defined as follows:

- **f75 — Total upward distance:** Sum of vertical displacements in the upward direction ( $\Delta y < 0$ ).
- **f76 — Total downward distance:** Sum of vertical displacements in the downward direction ( $\Delta y > 0$ ).
- **f77 — Total leftward distance:** Sum of horizontal displacements in the leftward direction ( $\Delta x < 0$ ).
- **f78 — Total rightward distance:** Sum of horizontal displacements in the rightward direction ( $\Delta x > 0$ ).
- **f79 — Number of upward transitions:** Count of gaze shifts with  $\Delta y < 0$ .
- **f80 — Number of downward transitions:** Count of gaze shifts with  $\Delta y > 0$ .
- **f81 — Number of leftward transitions:** Count of gaze shifts with  $\Delta x < 0$ .
- **f82 — Number of rightward transitions:** Count of gaze shifts with  $\Delta x > 0$ .

## 5.5 Machine Learning Analysis

This section presents the machine learning analysis conducted on the collected data, which has been represented through the diverse set of features described in the previous section. Given that the domain of behavioral biometrics based on dynamic text remains largely unexplored, our study also investigates the feasibility of performing user identification.

It is well established that biometric identification based on behavioral signals often achieves limited performance, making it unsuitable for reliable use in soft biometric applications. Nevertheless, prior research has shown that certain modalities, such as keystroke dynamics [9], represent notable exceptions. Accordingly, our first objective is to assess whether gaze- and pupil-based features in our setting also exhibit such exceptionality in the identification task.

In addition to this evaluation, the identification analysis is leveraged as a methodological step to optimize hyperparameters and to determine the most informative subsets of features via the SelectKBest method [91]. The insights obtained from this stage are subsequently applied to the verification analysis, which constitutes the core of our biometric evaluation. Specifically, by transferring the optimized hyperparameters and reduced feature sets from the identification task, we significantly lower the computational cost of verification. This step is crucial, as verification requires training a separate model for each individual in the dataset, and performing hyperparameter optimization for every subject would otherwise lead to prohibitive computational demands.

In our analysis, we employ seven different machine learning classifiers to construct and evaluate predictive models. As an initial step, we used the Naïve Bayes classifier to obtain a preliminary understanding of the potential performance that could be achieved. However due to its strong assumption of feature independence, which is rarely satisfied in practice, this model is utilized solely for exploratory purposes. Consequently, Naïve Bayes results are not included in the main performance tables, as the classifier is primarily used for a rapid baseline estimation, and its performance is typically inferior to more advanced methods.

Among the classifiers employed in this study, the k-Nearest Neighbors (kNN) algorithm [94] represents a well-established and widely used method across various application domains. One of the principal strengths of k-NN lies in its simplicity and interpretability, making it an effective baseline model before the adoption of more complex techniques. It often provides satisfactory performance without extensive parameter tuning, which makes it particularly suitable for exploratory analysis. While the construction of a k-NN model is computationally efficient, prediction time can become relatively slow when dealing with large datasets, either in terms of the number of features or the number of samples [93].

The next classifier employed is Logistic Regression [95], which is widely used due to its simplicity, interpretability, and efficiency. One of its main advantages is that it is easy to implement and does not require strong assumptions about the distribution of the classes in the feature space. Moreover, Logistic Regression can be naturally extended to handle multiple classes through multinomial regression, providing a probabilistic interpretation of class predictions that is both intuitive and informative. However, the primary limitation of Logistic Regression lies in its assumption of linearity between the independent variables and the log-odds of the dependent variable. As a result, it can only model linear decision boundaries and therefore struggles to capture more complex, nonlinear relationships that may exist within the data [96].

To proceed we used SVC [97] and one of its variant Nusvc [98]. According to [93] Kernelized Support Vector Machines (SVMs) are highly effective models capable of learning complex decision boundaries, even with few features. They perform well on both low- and high-dimensional data but scale poorly with the number of samples. While datasets up to around 10,000 samples are manageable, larger ones can cause significant runtime and memory issues. Moreover, SVMs require careful data preprocessing and precise parameter tuning to achieve optimal performance. Unlike SVC, which relies on the parameter  $C$  to balance the trade-off between maximizing the margin and minimizing classification errors, NuSVC introduces the parameter  $\nu$  (nu). This parameter defines an upper bound on the proportion of training errors and a lower bound on the proportion of support vectors. As a result, NuSVC offers a more interpretable and direct means of controlling model complexity and the number of support vectors involved in the decision

function [99].

The next classifier is one of the most well-known and widely used: Random Forest [100]. Random forests are powerful machine learning methods for both regression and classification, requiring minimal parameter tuning and no data scaling. They retain the strengths of decision trees while overcoming many of their limitations. However, their complexity—arising from combining numerous deep trees—reduces interpretability, making single decision trees preferable when a simple, visual explanation of the decision process is needed [93].

Finally, the last classifier used is the multilayer perceptron (MLP) [101]. Neural networks have regained prominence in machine learning due to their ability to capture complex patterns from large datasets. With sufficient data, computation, and parameter tuning, they often outperform other algorithms in both classification and regression tasks. However, they require extensive training time, careful data preprocessing, and are best suited for homogeneous data [93].

The classifiers presented above will be employed throughout the analysis. A single classifier is not used because, in machine learning, there is no universally best algorithm. In fact performance depends on the characteristics of the data and the specific problem being addressed.

As previously stated, the first part of the analysis focuses on the identification problem. This phase employs two distinct data-splitting strategies to create training and testing sets. The first is the classic random split using 80% of the data for the train set and 20% for the test set. The second is a session-based split, where data from the first two sessions (S1 and S2) form the training set, and data from the third session (S3) is used as the test set. Initially, a general identification will be performed using the entire dataset, both with and without feature selection, to evaluate any potential improvements in performance. An analysis involving every single animation taken singularly is then performed to assess which is the best performing animations and see if a single animation can be used for identification purpose. This analysis is divided into two types, based on the data used in the training set. The first method is called single train (ST), where the model is trained and tested on data from a single animation. While the second approach is called full train (FT), where the model is trained on data from all animations combined and tested on a single animation. The latter method has been used in previous studies and, in some cases, achieved better results due to the larger amount of training data. However, in other cases, performance worsened because the inclusion of all animations introduced confusion between different data distributions.

Afterward, we proceed to the verification phase, which represents the main objective of this work. In this stage, we employ only the session-based split, applied in the same manner as in the identification analysis. This choice reflects a more realistic, real-world scenario, where user verification occurs across different sessions. Similar to the identification phase, we first conduct a global analysis using the entire dataset across all animations, both with and without feature selection, to obtain a general overview of the achievable performance. Subsequently, we focus on single-animation analyses, maintaining the two training strategies previously adopted for identification. However, this part introduces an additional division. Specifically, we evaluate two modes, *st* and *ft*, to compare a standard verification approach with one that uses only the best-performing model for testing. These modes will be discussed in detail in the dedicated chapter. For all these configurations, we also performed an open-set analysis to evaluate how the system responds when an unregistered and previously unseen user attempts authentication. Finally, to complete the verification analysis, we conducted a temporal scaling experiment, reducing the duration of all animations. The objective was to assess how the best-performing animation behaves when using only the first 10 or 5 seconds instead of the full 15-second duration.

Finally, we perform a comparative analysis between identification and verification using the full dataset, rather than single animations, in order to assess the performance differences between static and dynamic text under similar experimental conditions.

### 5.5.1 Identification without feature selection

We begin the identification analysis using the entire dataset, without distinguishing between different animations. In this initial phase, feature selection through SelectKBest [91] is intentionally omitted, as

the objective is to establish a baseline reference. These results will serve as a comparison point to evaluate potential improvements obtained when applying feature selection in subsequent analyses. The program goal is to use two different types of split: random split and session split.

The random split is performed using the `train_test_split` function [102], where 80% of the dataset is randomly selected to form the training set, and the remaining 20% is used as the test set. To ensure balanced representation across individuals, stratification based on the tester identifier is applied. This guarantees that approximately 80% of each participant’s data is included in the training set, while the remaining 20% is allocated to the test set, maintaining consistent class proportions across both subsets. To minimize the randomness inherent in this approach, the procedure was repeated ten times, and the results were averaged across all runs. This repetition reduces the impact of particularly favorable or unfavorable data splits that could otherwise bias the performance evaluation.

The session-based split corresponds to the three recording sessions conducted on different days. This design was chosen primarily to minimize the influence of temporary conditions such as fatigue or momentary distraction, ensuring that the data reflect a more stable and representative ocular behavior. Moreover, performing the recordings on separate days allows us to average out day-to-day variability. The second motivation for this setup is methodological: the data from the first two sessions (S1 and S2) are used to train the model, while the third session (S3) serves as the test set, enabling an evaluation of the model’s generalization capability across sessions.

A hyperparameter tuning procedure was implemented using a pipeline [103] composed of three main stages. The first stage employs the `SimpleImputer` [105] to handle missing values by replacing them with the mean of the corresponding feature column. The second stage performs feature scaling, tested with three different normalization methods: `MinMaxScaler` [106], `StandardScaler` [107], and `RobustScaler` [108]. Finally, the third stage involves hyperparameter optimization, using the parameter grids reported in figure 13 and `GridSearchCV` [104]. It is important to note that each time a `GridSearchCV` strategy was applied to determine the optimal hyperparameters, the complete pipeline configuration was consistently employed. The hyperparameter optimization was conducted using a 5-fold cross-validation procedure ( $cv = 5$ ), in which the training data were divided into five equally sized folds. In each iteration, four folds were used for model training and the remaining one for validation. This process was repeated five times so that every fold served once as the validation set, and the resulting performance metrics were averaged to obtain a more reliable and unbiased estimate of the model’s performance.

We’re now ready to present the results obtained during this analysis.

Identification without feature selection		
Classifier	Random split	Session split
KNN	0.6974	0.6
Logistic Regression	0.7036	0.6524
Nusvc	0.7552	0.6405
Random Forest	0.7762	0.6167
SVC	0.7668	0.6369
MLP	0.7645	0.6655

Figure 10: Identification results without feature selection

Figure 10 illustrates that, for the random split type, the best performance is achieved by the Random Forest classifier, yielding an accuracy of 77.62%. In contrast, for the session split type, the highest accuracy is obtained by the Multilayer Perceptron (MLP) model, with a value of 66.55%.

A notable performance drop is observed when transitioning from the random split to the session split. This decline is expected and can be attributed to the reduced amount of training data available in the session split scenario. Specifically, the random split allocates 80% of the data to the training set, whereas the session split allocates only 66.6%. Moreover, the random split introduces a potential data leakage issue, as samples from the same session may appear in both the training and testing sets. This overlap provides the model with indirect information about the test data, thereby inflating performance compared

to the session split, where such leakage is avoided.

### 5.5.2 Identification with feature selection

We now proceed to perform the identification analysis with an additional step: feature selection. The overall procedure and implementation remain identical to those described in Section 5.5.1. In particular two data split strategies were used, random and session, together with GridSearchCV for hyperparameter optimization. The main difference lies in the inclusion of the feature selection stage, illustrated in Figure 14. Specifically, we employed the SelectKBest method[91] to identify the subset of features (k) that yielded the highest mean cross-validation accuracy during the grid search. This approach not only enhances model interpretability but also reduces the dimensionality of the feature space, thereby potentially improving generalization performance.

Feature selection with identification		
Classifier	Random split	Session split
KNN	0.7044	0.6071
Logistic Regression	0.7026	0.65
Nusvc	0.7619	0.6464
Random Forest	0.7696	0.6
SVC	0.7696	0.6548
MLP	0.7587	0.6762

Figure 11: Identification results with feature selection

Figure 11 presents the results obtained from this analysis. For the random split type, the Random Forest classifier remains the best performer, achieving an accuracy of 76.69%. This represents a slight decrease compared to the accuracy of 77.62% obtained without the feature selection step, corresponding to a reduction of 13 features (from 83 to 70). Interestingly, the Support Vector Classifier (SVC) achieved the same accuracy of 76.69% using the reduced feature set of 70 features, matching the performance of the Random Forest under these conditions.

For the session split, a modest improvement was observed with the Multilayer Perceptron (MLP) classifier, whose accuracy increased from 66.55% to 67.62% when using only 70 features.

The number of selected features, along with their corresponding identities, are presented in Appendix 6.2.

The table 1 displays the top 30 features selected across the entire analysis. These features were ranked according to their cumulative F-score, aggregated across all classifiers and session splits. Each time a feature appeared during the selection process, its F-score was added to an accumulator. If the same feature appeared again in subsequent analyses, its score was added cumulatively to the previous total. As shown in the table, the top 10 positions are predominantly occupied by pupil-related features, either referring to a single eye or derived from ratios or differences between the two eyes.

In contrast, Table 2 displays the set of features that were never selected in any model or split type. We can observe that many of these are saccade speed features, along with two distance-between-consecutive-fixations features. Interestingly, the kurtosis ratio between left and right pupil diameter is also never selected, despite pupil diameter being among the best features. This suggests that kurtosis may not be a representative metric in this context.

Table 1: Top 30 features for Identification with feature selection ranked by cumulative F-score

Feature ID	Feature name	Cumulative F-score
f44	Arithmetic mean right pupil diameter	2777.22
f45	Geometric mean right pupil diameter	2724.38
f46	Median right pupil diameter	2658.24
f34	Arithmetic mean left pupil diameter	2482.82
f35	Geometric mean left pupil diameter	2443.10
f36	Median left pupil diameter	2384.14
f64	Arithmetic mean difference between left and right pupil diameter	1734.38
f65	Median difference between left and right pupil diameter	1697.57
f33	Maximum left pupil diameter	1644.71
f43	Maximum right pupil diameter	1561.86
f55	Geometric mean ratio between left and right pupil diameter	1139.40
f56	Median ratio between left and right pupil diameter	1125.79
f54	Arithmetic mean ratio between left and right pupil diameter	1065.89
f71	Number of blinks	760.39
f52	Minimum ratio between left and right pupil diameter	599.55
f62	Minimum difference between left and right pupil diameter	529.04
f58	Median absolute deviation ratio between left and right pupil diameter	437.26
f60	Interquartile range ratio between left and right pupil diameter	410.47
f67	Median absolute deviation difference between left and right pupil diameter	385.03
f69	Interquartile range difference between left and right pupil diameter	374.81
f66	Standard deviation difference between left and right pupil diameter	357.18
f63	Maximum difference between left and right pupil diameter	340.73
f39	Skewness left pupil diameter	250.94
f53	Maximum ratio between left and right pupil diameter	234.36
f37	Standard deviation left pupil diameter	228.57
f57	Standard deviation ratio between left and right pupil diameter	227.71
f49	Skewness right pupil diameter	223.79
f80	Number of movements in descending direction	210.63
f42	Minimum right pupil diameter	196.55
f32	Minimum left pupil diameter	196.55

Table 2: List of features never selected in Identification with feature selection

Feature ID	Feature name	Description	
f11	Minimum distance between consecutive fixations	Saccade-related metric	
f16	Standard deviation distance between consecutive fixations	Saccade-related metric	
f21	Minimum saccade speed	Saccade speed metric	
f23	Arithmetic mean saccade speed	Saccade speed metric	st
f24	Geometric mean saccade speed	Saccade speed metric	
f25	Median saccade speed	Saccade speed metric	
f28	Skewness saccade speed	Saccade speed metric	
f30	Kurtosis saccade speed	Saccade speed metric	
f61	Kurtosis ratio between left and right pupil diameter	Pupil diameter ratio metric	

### 5.5.3 Identification with single animation - single train(ST) type

The main goal of this stage was to identify which animation performs best among the eight selected stimuli. We began with a full identification analysis to establish a general baseline of the performance achievable when using all animations together. Building on this, we refined the analysis to examine the performance of each animation individually. This refined procedure, referred to as the single train (ST) approach, trains the model using data from one animation only and then tests it on data from the same animation.

Two data splitting strategies were used to ensure robust evaluation. The first, a random split, allocates 80% of the data for a given animation to training and the remaining 20% to testing. This process was stratified by tester ID to ensure a fairer distribution and was repeated ten times to reduce variability caused by random sampling.

The second, a session-based split, uses data from the first two sessions of an animation for training and the third session for testing. This method was executed once to simulate a real-world scenario of session progression.



The analysis was conducted through a structured pipeline composed of four sequential steps. First, missing values were addressed using an imputation procedure that replaced them with the mean value of each respective feature, ensuring completeness of the dataset. Second, a scaling step was applied to standardize the feature values, bringing them onto a comparable range and preventing scale-related bias in the model. Third, a feature selection process was performed, following the same methodology as in the previous analysis, to identify and retain the most informative features for the model. Finally, hyperparameter tuning was carried out using GridSearchCV with the parameter grid presented in Figure 14, optimizing the model's performance.

This analysis serves a dual purpose. Primarily, it evaluates the performance of each animation within the context of an identification task, providing insight into their relative effectiveness. Additionally, it establishes the optimal parameters for the subsequent verification analysis, which is considerably more computationally intensive as it requires constructing a separate model for each individual. By performing parameter tuning within the single-animation identification analysis, this process becomes significantly more efficient. The parameters identified through this analysis, together with the top k features selected, constitute the foundation for the verification analysis, in which feature selection determines the final set of features used for model evaluation.

Identification single animation (single animation as train set) and random split								
Classifier	HS_FA_BIG	HS_FA_LIT	HS_SL_BIG	HS_SL_LIT	VB_FA_BIG	VB_FA_LIT	VB_SL_BIG	VB_SL_LIT
KNN	0,6143	0,5794	0,5349	0,5333	0,7698	0,719	0,7524	0,7
Logistic Regression	0,6429	0,6206	0,5778	0,6	0,8	0,7746	0,746	0,7254
MLP	0,6683	0,6143	0,6048	0,6175	0,7968	0,7698	0,7683	0,7349
NuSVC	0,6635	0,6175	0,5508	0,5984	0,7937	0,7825	0,746	0,7127
Random Forest	0,7397	0,6302	0,6635	0,6317	0,8175	0,781	0,819	0,7698
SVC	0,6794	0,646	0,5683	0,6032	0,7968	0,7714	0,7683	0,7286
Identification single animation (single animation as train set) and session split								
Classifier	HS_FA_BIG	HS_FA_LIT	HS_SL_BIG	HS_SL_LIT	VB_FA_BIG	VB_FA_LIT	VB_SL_BIG	VB_SL_LIT
KNN	0,4095	0,4952	0,4286	0,3905	0,7905	0,6667	0,7238	0,619
Logistic Regression	0,6	0,5524	0,5143	0,4762	0,8	0,8	0,7429	0,6952
MLP	0,5333	0,5714	0,4286	0,4952	0,7333	0,7238	0,6952	0,6476
NuSVC	0,5333	0,6	0,4857	0,4095	0,7524	0,7238	0,7143	0,6667
Random Forest	0,5905	0,6095	0,5048	0,5333	0,781	0,781	0,7238	0,7524
SVC	0,4857	0,5143	0,5048	0,4762	0,8286	0,7238	0,7333	0,6476

Figure 12: Identification single ST

Figure 12 presents a comprehensive collection of the results obtained during the analysis. For each type of data split, the figure includes a table showing the performance of each animation using the selected classifier.

Considering the random split, the best-performing animations are VB\_SL\_BIG with an accuracy of 81.9% and VB\_FA\_BIG with an accuracy of 81.75%. In contrast, when evaluating the session split, VB\_FA\_BIG emerges as the best performer, with accuracy increasing from 81.75% to 82.86%. This is followed by VB\_FA\_LIT, whose accuracy rises from 78.25% to 80%, while VB\_SL\_BIG shows a notable decrease from 81.9% to 74.29%.

At first glance, this behaviour may seem counterintuitive, as one would generally expect performance to decrease under a session split compared to a random split. This expectation is based on the fact that the training set for the random split contains 80% of the data, whereas the training set for the session split contains only 66.6%, and in the random split samples from the same session can appear in both the training and test sets. However, the results reveal a more nuanced pattern: certain animations experience a drop in accuracy, particularly horizontal scroll types, while others show slight improvements despite having a smaller training set.

This improvement may be attributed to the reduced risk of overfitting when training on more distinct session data. By excluding samples from the same session in training and testing, the model may generalise better to unseen data, particularly in animations with more consistent and robust patterns. This behaviour appears to be a characteristic of dynamic text stimuli, and further evidence supporting this observation

will be discussed in Section 5.6.

## 5.6 Static vs Dinamic text

# 6 Appendix

This section includes supplementary materials that could not be accommodated within the main chapters due to their length and level of detail. Specifically, it provides a subsection presenting the complete parameter grid used during hyperparameter tuning, listing all parameters explored for each classifier. In addition, it contains extended result tables that complement the main findings discussed in the core analysis, offering a broader and more detailed view of the experimental outcomes.

## 6.1 Parameter grid

For every `GridSearchCV` execution, the cross-validation parameter was set to `cv = 5`, applying a 5-fold cross-validation procedure. The evaluation metric was defined as `scoring = 'accuracy'` to optimize the models based on classification accuracy. Figure 13 display the full parameter grid used everytime an hyperparameter tuning is performed, without using feature selection step.

Classifier	Parameters
Gaussian Naive Bayes	<input type="checkbox"/> <b>Scaler</b> : MinMaxScaler, StandardScaler, RobustScaler <input type="checkbox"/> No additional hyperparameters tuned
K-Nearest Neighbors (k-NN)	<input type="checkbox"/> <b>Scaler</b> : MinMaxScaler, StandardScaler, RobustScaler <input type="checkbox"/> <b>n_neighbors</b> : 3, 5, 7, 9, 11 <input type="checkbox"/> <b>weights</b> : uniform, distance <input type="checkbox"/> <b>metric</b> : minkowski, euclidean, manhattan
Logistic Regression	<input type="checkbox"/> <b>Scaler</b> : MinMaxScaler, StandardScaler, RobustScaler <input type="checkbox"/> <b>C</b> : 0.001, 0.01, 0.1, 1, 10, 100
Nu-Support Vector Machine (NuSVC)	<input type="checkbox"/> <b>Scaler</b> : MinMaxScaler, StandardScaler, RobustScaler <input type="checkbox"/> <b>nu</b> : 0.25, 0.5, 0.75 <input type="checkbox"/> <b>kernel</b> : rbf, poly, sigmoid <input type="checkbox"/> <b>gamma</b> : scale, auto
Random Forest	<input type="checkbox"/> <b>Scaler</b> : MinMaxScaler, StandardScaler, RobustScaler <input type="checkbox"/> <b>n_estimators</b> : 20, 30, 50, 100, 200 <input type="checkbox"/> <b>max_features</b> : sqrt <input type="checkbox"/> <b>max_depth</b> : 5, 10, 20, 30
Support Vector Machine (SVC)	<input type="checkbox"/> <b>Scaler</b> : MinMaxScaler, StandardScaler, RobustScaler <input type="checkbox"/> <b>C</b> : 0.001, 0.01, 0.1, 1, 10, 100 <input type="checkbox"/> <b>gamma</b> : 0.001, 0.01, 0.1, 1, 10, 100 <input type="checkbox"/> <b>kernel</b> : rbf, poly
Multilayer Perceptron (MLP)	<input type="checkbox"/> <b>Scaler</b> : MinMaxScaler, StandardScaler, RobustScaler <input type="checkbox"/> <b>hidden_layer_sizes</b> : (100,), (100, 50), (150, 100, 50) <input type="checkbox"/> <b>activation</b> : tanh, relu <input type="checkbox"/> <b>alpha</b> : 0.0001, 0.001, 0.01 <input type="checkbox"/> <b>learning_rate_init</b> : 0.001, 0.01 <input type="checkbox"/> <b>solver</b> : adam

Figure 13: Parameter grid without feature selection step

Conversely, Figure 13 presents the parameter grid that includes the additional feature selection step.

Classifier	Parameters
Gaussian Naive Bayes	<input type="checkbox"/> <b>Scaler:</b> MinMaxScaler, StandardScaler, RobustScaler <input type="checkbox"/> <b>Feature_selection_k:</b> [30, 40, 50, 60, 70] <input type="checkbox"/> No additional hyperparameters tuned
K-Nearest Neighbors (k-NN)	<input type="checkbox"/> <b>Scaler:</b> MinMaxScaler, StandardScaler, RobustScaler <input type="checkbox"/> <b>Feature_selection_k:</b> [30, 40, 50, 60, 70] <input type="checkbox"/> <b>n_neighbors:</b> 3, 5, 7, 9, 11 <input type="checkbox"/> <b>weights:</b> uniform, distance <input type="checkbox"/> <b>metric:</b> minkowski, euclidean, manhattan
Logistic Regression	<input type="checkbox"/> <b>Scaler:</b> MinMaxScaler, StandardScaler, RobustScaler <input type="checkbox"/> <b>Feature_selection_k:</b> [30, 40, 50, 60, 70] <input type="checkbox"/> <b>C:</b> 0.001, 0.01, 0.1, 1, 10, 100
Nu-Support Vector Machine (NuSVC)	<input type="checkbox"/> <b>Scaler:</b> MinMaxScaler, StandardScaler, RobustScaler <input type="checkbox"/> <b>Feature_selection_k:</b> [30, 40, 50, 60, 70] <input type="checkbox"/> <b>nu:</b> 0.25, 0.5, 0.75 <input type="checkbox"/> <b>kernel:</b> rbf, poly, sigmoid <input type="checkbox"/> <b>gamma:</b> scale, auto
Random Forest	<input type="checkbox"/> <b>Scaler:</b> MinMaxScaler, StandardScaler, RobustScaler <input type="checkbox"/> <b>Feature_selection_k:</b> [30, 40, 50, 60, 70] <input type="checkbox"/> <b>n_estimators:</b> 20, 30, 50, 100, 200 <input type="checkbox"/> <b>max_features:</b> sqrt <input type="checkbox"/> <b>max_depth:</b> 5, 10, 20, 30
Support Vector Machine (SVC)	<input type="checkbox"/> <b>Scaler:</b> MinMaxScaler, StandardScaler, RobustScaler <input type="checkbox"/> <b>Feature_selection_k:</b> [30, 40, 50, 60, 70] <input type="checkbox"/> <b>C:</b> 0.001, 0.01, 0.1, 1, 10, 100 <input type="checkbox"/> <b>gamma:</b> 0.001, 0.01, 0.1, 1, 10, 100 <input type="checkbox"/> <b>kernel:</b> rbf, poly
Multilayer Perceptron (MLP)	<input type="checkbox"/> <b>Scaler:</b> MinMaxScaler, StandardScaler, RobustScaler <input type="checkbox"/> <b>Feature_selection_k:</b> [30, 40, 50, 60, 70] <input type="checkbox"/> <b>hidden_layer_sizes:</b> (100,), (100, 50), (150, 100, 50) <input type="checkbox"/> <b>activation:</b> tanh, relu <input type="checkbox"/> <b>alpha:</b> 0.0001, 0.001, 0.01 <input type="checkbox"/> <b>learning_rate_init:</b> 0.001, 0.01 <input type="checkbox"/> <b>solver:</b> adam

Figure 14: Parameter grid with feature selection step

## 6.2 Identification with feature selection - Best k features

The following section presents the optimal number of features (k) selected for each classifier during the identification phase with feature selection. It also provides a complete list of these features along with their corresponding f-scores, which quantify the relative importance of each feature throughout the evaluation process.

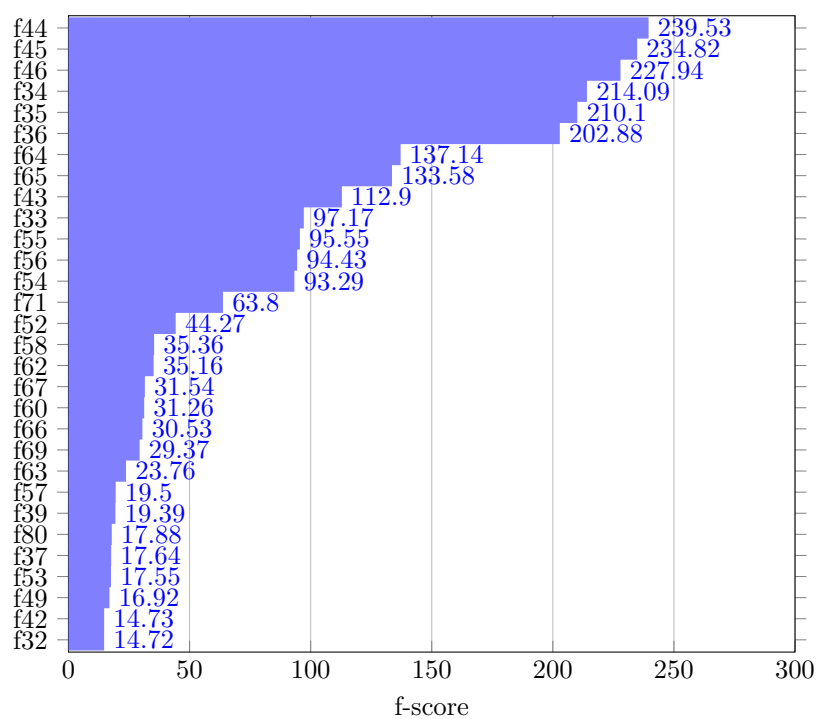


Figure 15: Knn classifier with random split and  $k = 60$

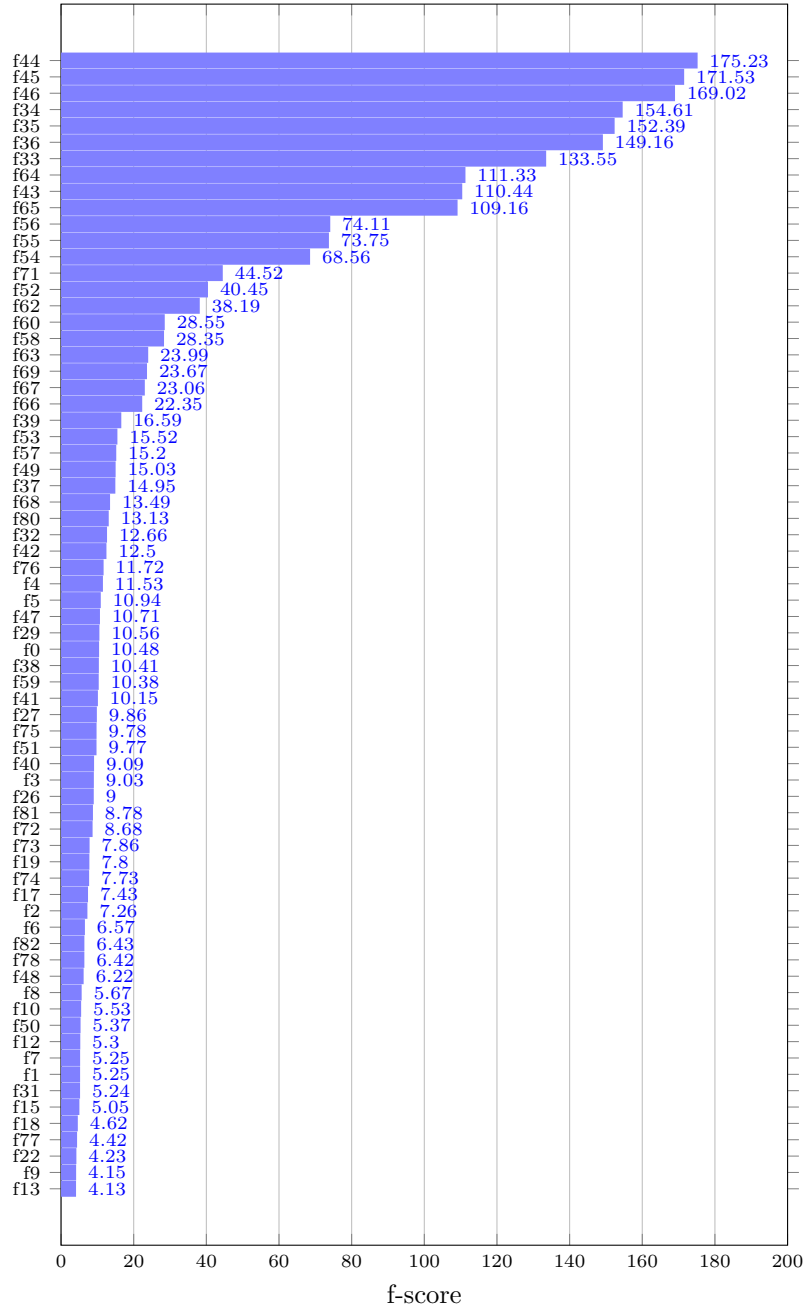


Figure 16: Knn classifier with session split and  $k = 70$

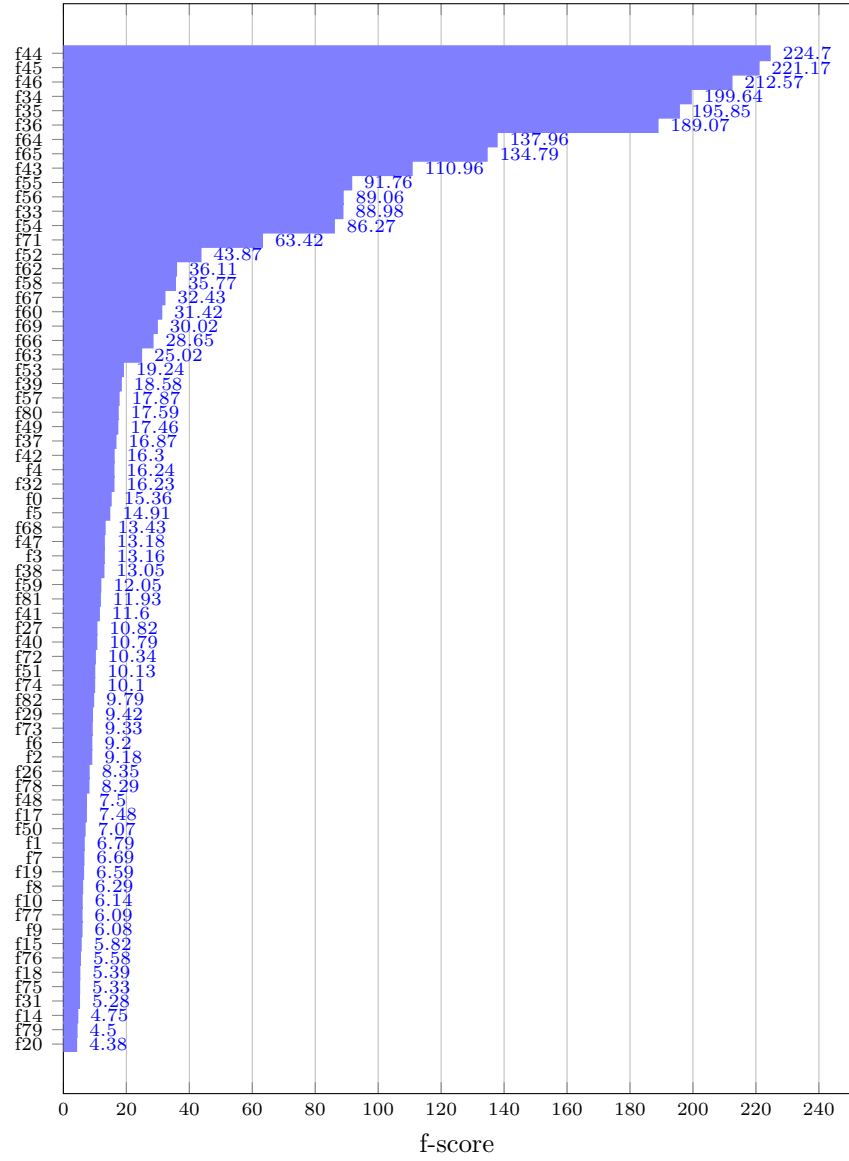


Figure 17: Logistic Regression classifier with random split and  $k = 70$ .

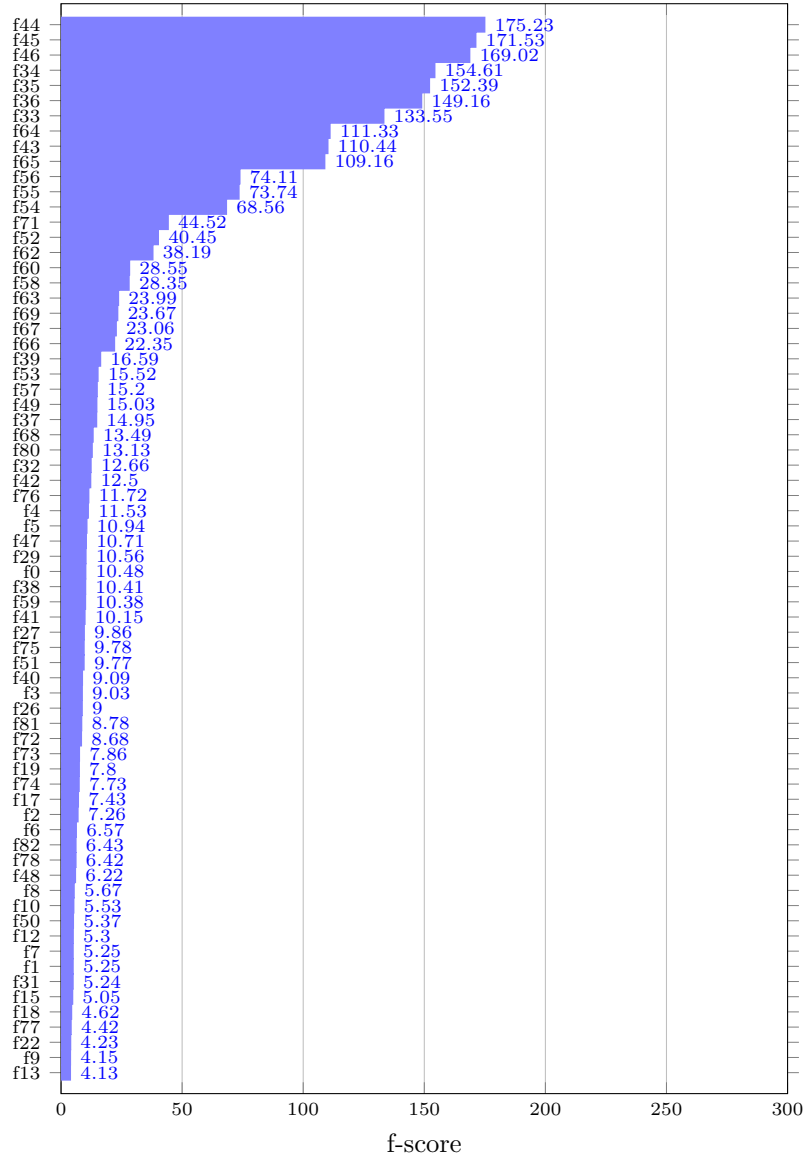


Figure 18: Logistic Regression classifier with session split and  $k = 70$

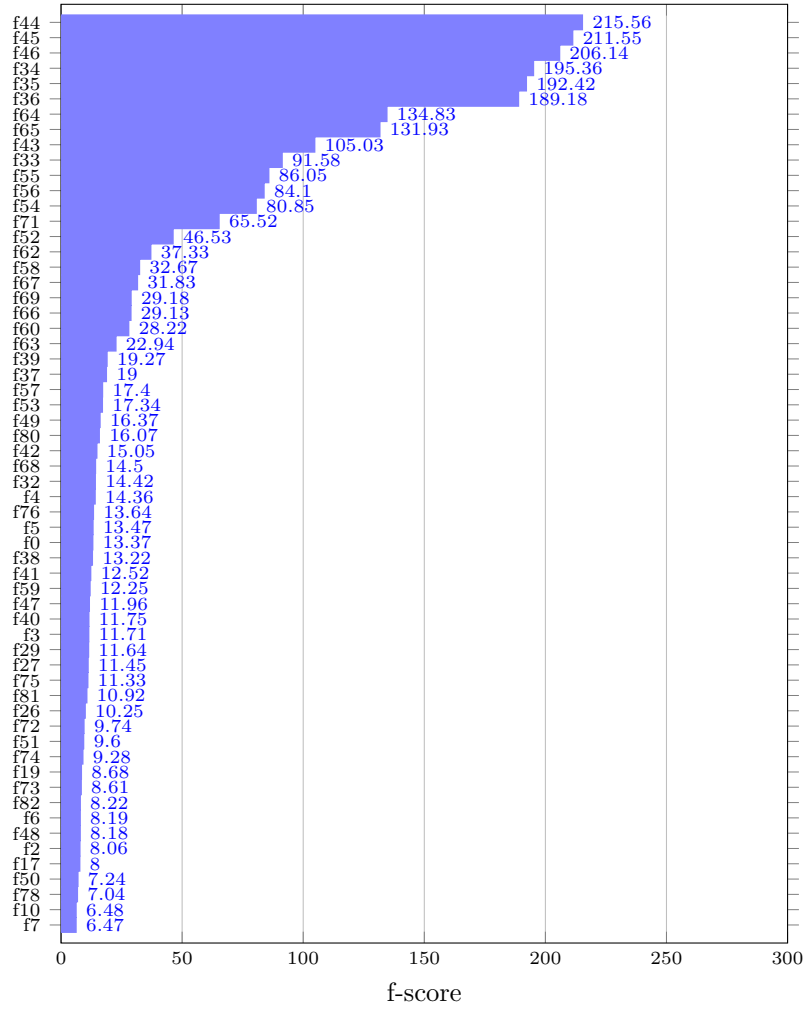


Figure 19: NuSVC with random split and  $k = 60$



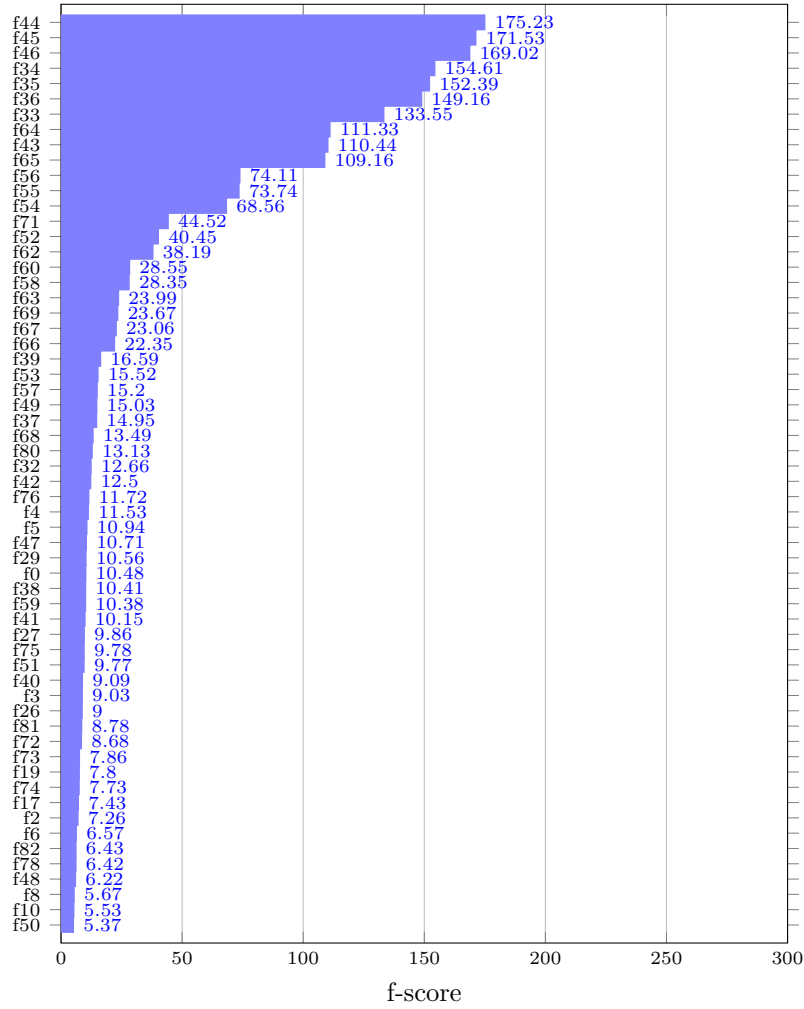


Figure 20: NuSVC with session split and  $k = 60$

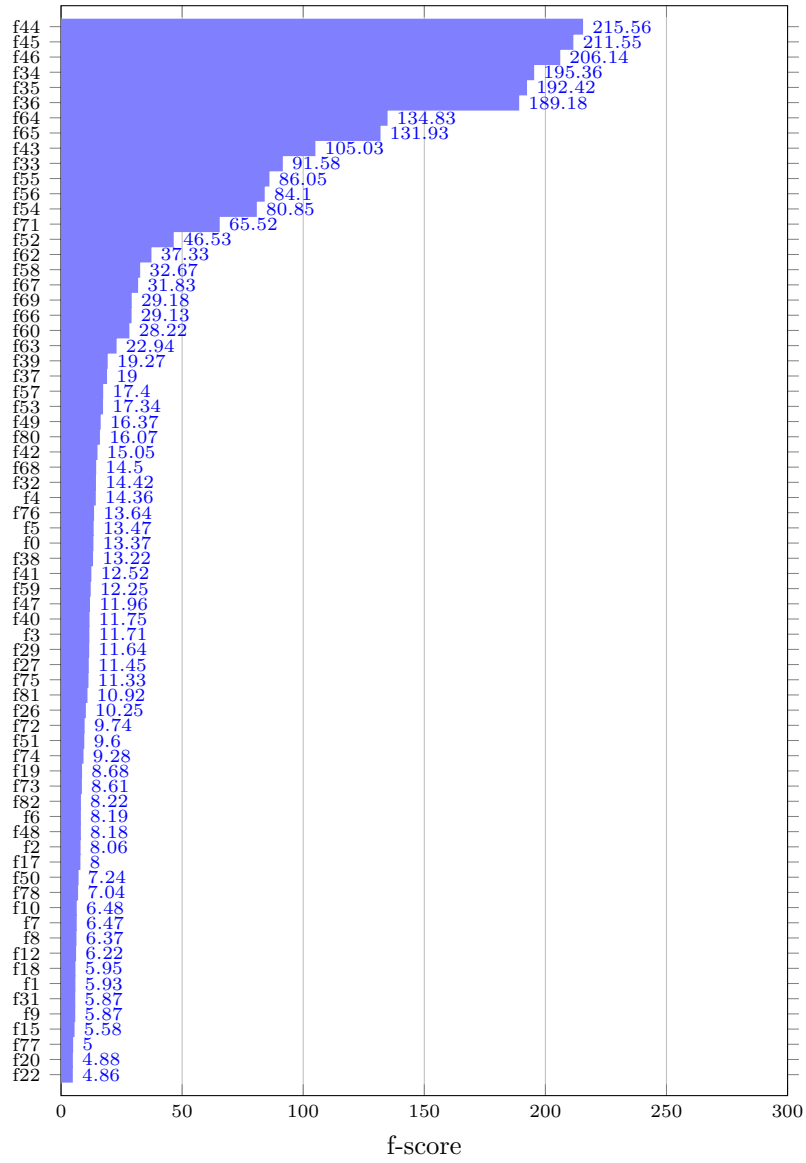


Figure 21: SVC with random split and  $k = 70$

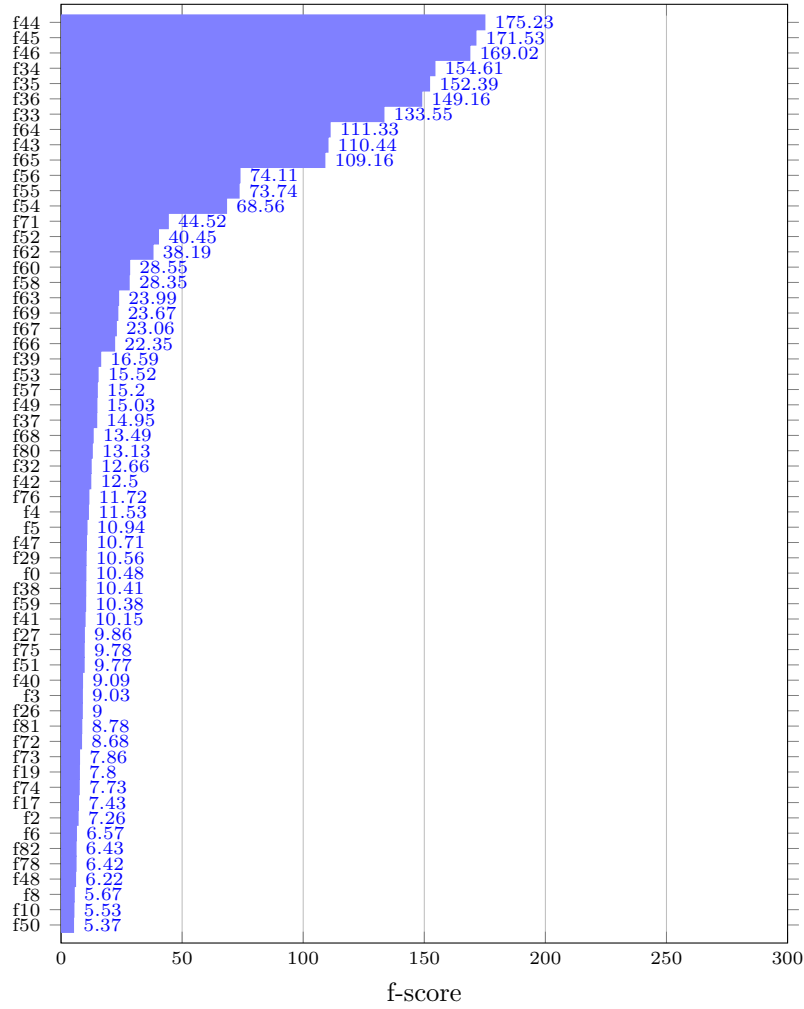


Figure 22: SVC classifier with session split and  $k = 60$

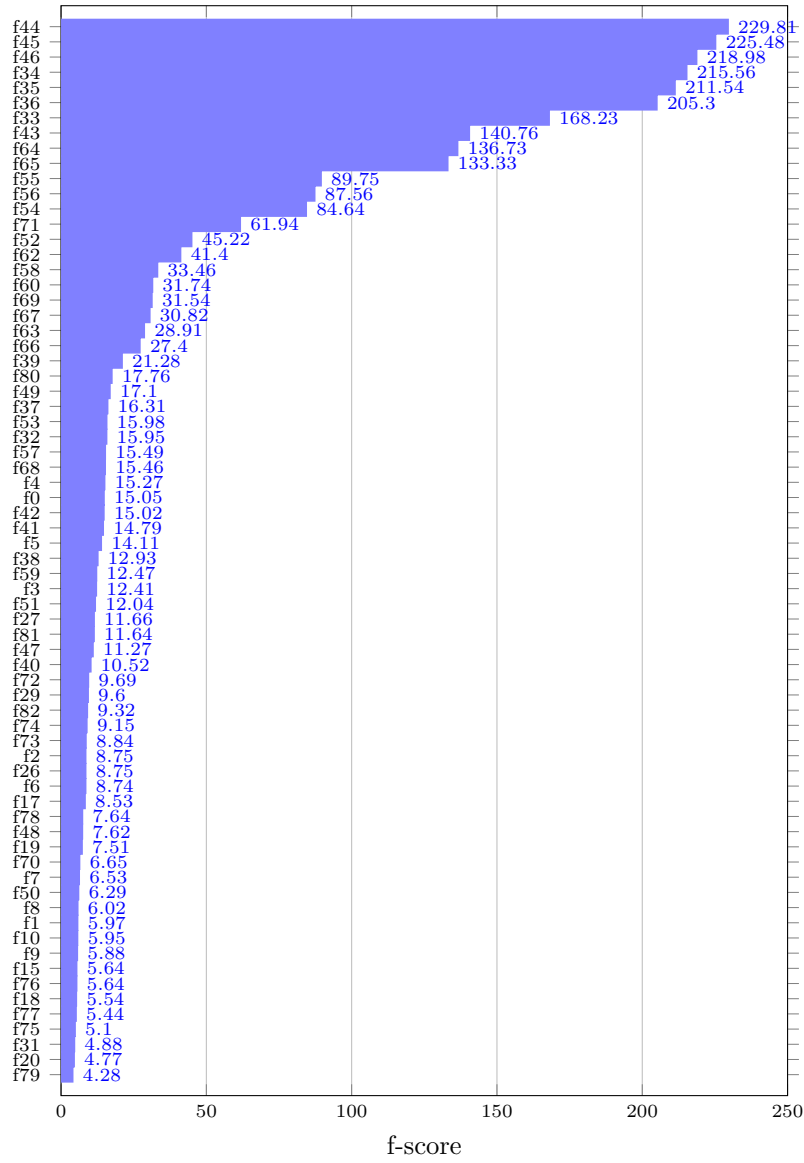


Figure 23: Random Forest with random split and  $k = 70$

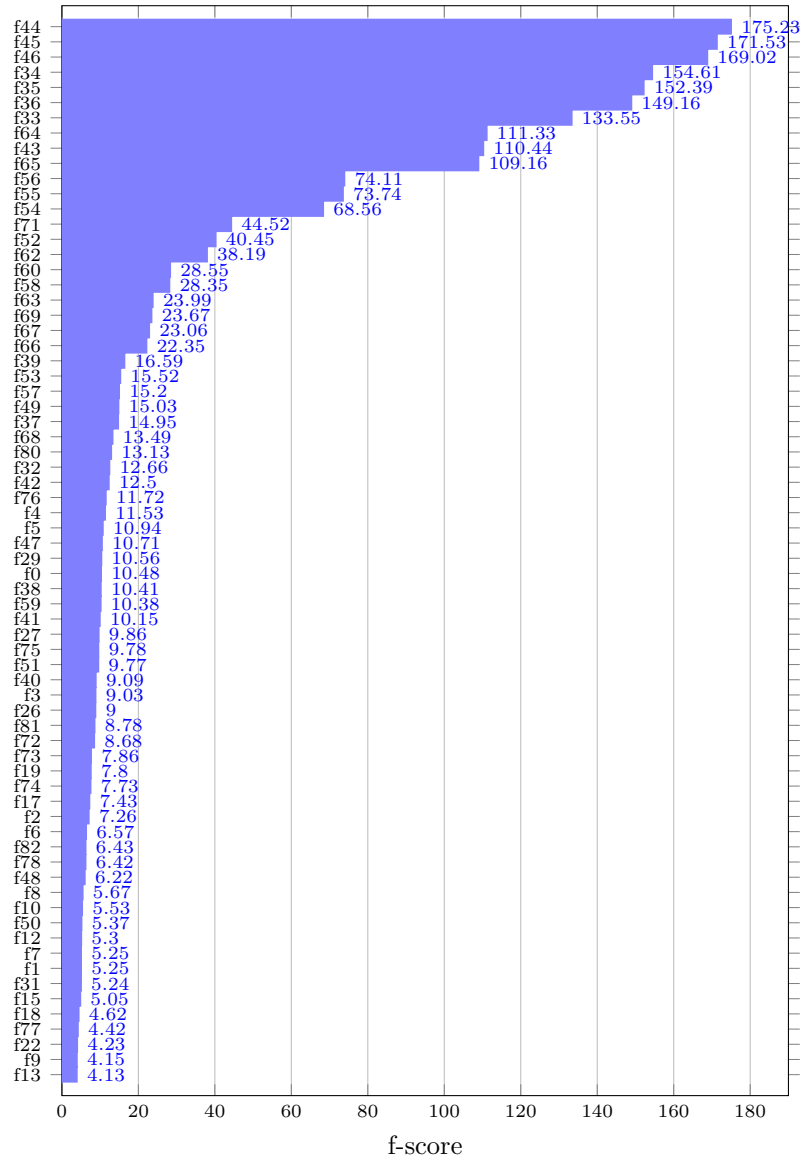


Figure 24: Random Forest with session split and  $k = 70$

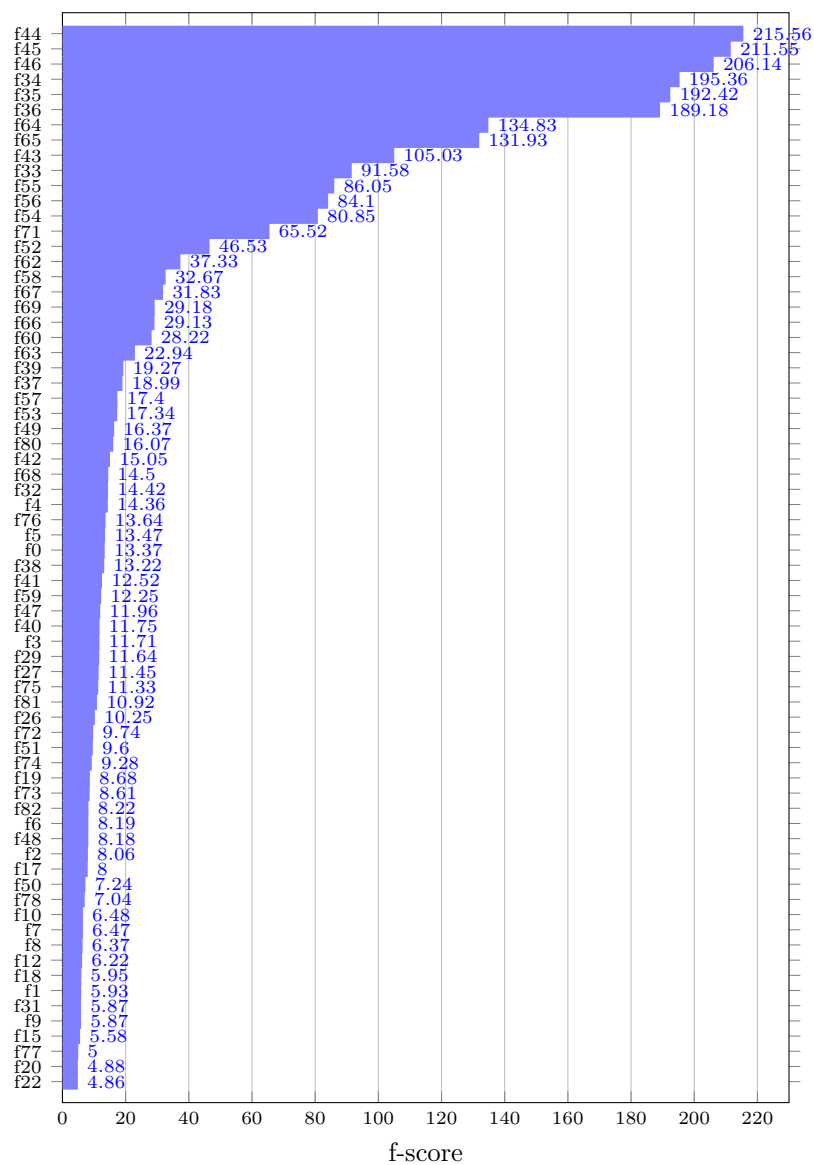


Figure 25: MLP with random split and  $k = 70$

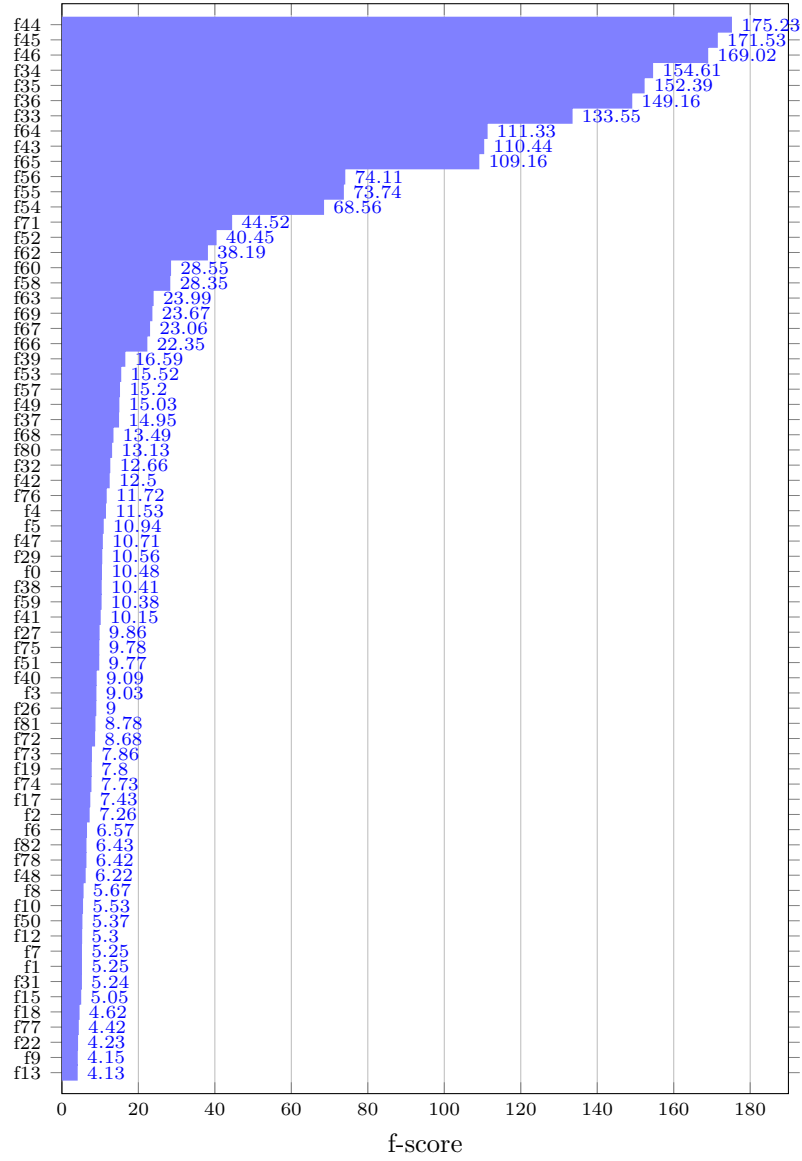


Figure 26: MLP classifier session split and  $k = 70$

## 7 Conclusion

## 8 Acknowledgment



## 9 References

- [1] Number of smartphone mobile network subscriptions worldwide from 2016 to 2023, with forecasts from 2023 to 2028, <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>
- [2] NIST study shows computerized fingerprint matching is highly accurate, <https://www.nist.gov/news-events/news/2004/07/nist-study-shows-computerized-fingerprint-matching-highly-accurate>
- [3] NIST study shows computerized fingerprint matching is highly accurate, [https://www.nec.com/en/press/202109/global\\_20210922\\_02.html#:~:text=ln%20this%20benchmark%20test%2C%20the,a%20million%20people%20\(\\*2\).](https://www.nec.com/en/press/202109/global_20210922_02.html#:~:text=ln%20this%20benchmark%20test%2C%20the,a%20million%20people%20(*2).)
- [4] Behavioral Biometrics for Human Identification: Intelligent Applications, Liang Wang University of Melbourne, Australia Xin Geng Southeast University, China
- [5] Al-Zubi, S., Brömme, A., & Tönnies, K. (2003, September 10-12). Using an active shape structural model for biometric sketch recognition. Paper presented at the DAGM, Magdeburg, Germany (pp. 187-195).
- [6] Stolfo, S. J., Hu, C.-W., Li, W.-J., Hershkop, S., Wang, K., & Nimeskern, O. (2003b). Combining behavior models to secure e-mail systems (No. CU Tech. Rep.).
- [7] Spafford, E. H., & Weeber, S. A. (1992, October). Software forensics: Can we track code to its authors? Paper presented at the 15th National Computer Security Conference (pp. 641-650).
- [8] Yampolskiy, R. V., & Govindaraju, V. (2006b, April 17-22). Use of behavioral biometrics in intrusion detection and online gaming. Paper presented at the Biometric Technology for Human Identification III. SPIE Defense and Security Symposium, Orlando, FL.
- [9] Bergadano, F., Gunetti, D., & Picardi, C. (2002). User authentication through keystroke dynamics. [TISSEC]. ACM Transactions on Information and System Security, 5(4), 367–397. doi:10.1145/581271.581272
- [10] Denning, D. E. (1987). An intrusion-detection model. IEEE Transactions on Software Engineering, 13(2), 222–232. doi:10.1109/TSE.1987.232894
- [11] Garg, A., Rahalkar, R., Upadhyaya, S., & Kwiat, K. (2006, June 21-23). Profiling users in GUI based systems for masquerade detection. Paper presented at The 7th IEEE Information Assurance Workshop (IAWorkshop 2006), West Point, NY.
- [12] Novikov, D. (2005). Neural networks to intrusion detection. Unpublished MS thesis, Rochester Institute of Technology, Rochester, NY.
- [13] Westeyn, T., Pesti, P., Park, K., & Starner, T. (2005, July). Biometric identification using song-based eye blink patterns. Paper presented at the Human Computer Interaction International (HCII), Las Vegas, NV.
- [14] Westeyn, T., & Starner, T. (2004). Recognizing song-based blink patterns: Applications for restricted and universal access. Paper presented at the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (p. 717).
- [15] BenAbdelkader, C., Cutler, R., & Davis, L. (2002). Person identification using automatic height and stride estimation. Paper presented at the IEEE International Conference on Pattern Recognition.
- [16] Kale, A., Sundaresan, A., Rajagopalan, A. N., & Cuntoor, N., RoyChowdhury, A., Kruger, V., et al. (2004). Identification of humans using gait. IEEE Transactions on Image Processing, 13(9). doi:10.1109/TIP.2004.832865
- [17] Erdogan, H., Ercil, A., Ekenel, H., Bilgin, S., Eden, I., & Kirisci, M. (2005a). Multimodal person recognition for vehicular applications. LNCS, 3541, 366–375.

- [18] Erdogan, H., Ozyagci, A. N., Eskil, T., Rodoper, M., Ercil, A., & Abut, H. (2005b, September). Experiments on decision fusion for driver recognition. Paper presented at the Biennial on DSP for in-vehicle and mobile systems, Sesimbra, Portugal.
- [19] Erzin, E., Yemez, Y., Tekalp, A. M., Erçil, A., Erdogan, H., & Abut, H. (2006). Multimodal person recognition for human-vehicle interaction. *IEEE MultiMedia*, 13, 18–31. doi:10.1109/MMUL.2006.37
- [20] Brause, R., Langsdorf, T., & Hepp, M. (1999). Neural data mining for credit card fraud detection. Paper presented at the 11th IEEE International Conference on Tools with Artificial Intelligence (pp. 103–106).
- [21] Andrew T. Duchowski. 2017. *Eye Tracking Methodology: Theory and Practice* (3rd. ed.). Springer Publishing Company, Incorporated.
- [22] Young, L.R., Sheena, D. Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation* 7, 397–429 (1975). <https://doi.org/10.3758/BF03201553>
- [23] Crane, H. D. (1994). The purkinje image eyetracker, image stabilization, and related forms of stimulus manipulation. In D. H.Kelly(Ed.), *Visual science and engineering: Models and applications*(pp. 13–89). New York: Marcel Dekker.
- [24] Crane,H.D.,&Steele,C.M.(1985).Generation-Vdual-purkinje-imageeyetracker.*AppliedOptics*,24, 527–537.
- [25] Chapter 9 - Biometric Authentication to Access Controlled Areas Through Eye Tracking, doi: <https://doi.org/10.1016/B978-0-08-100705-1.00009-9>
- [26] A. Maeder, C. Fookes, S. Sridharan, Gaze based user authentication for personal computer applications, in: *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing*, 2004, pp. 727–730.
- [27] O.V. Komogortsev, A. Karpov, L.R. Price, C. Aragon, Biometric authentication via oculomotor plant characteristics, in: *2012 5th IAPR International Conference on Biometrics (ICB)*, 2012, pp. 413–420
- [28] A. De Luca, R. Weiss, H. Hußmann, X. An, Eyepass– eye-stroke authentication for public terminals, in: *CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08)*, ACM, New York, NY, USA, 2008, pp. 3003–3008.
- [29] P. Dunphy, A. Fitch, P. Olivier, Gaze-contingent passwords at the ATM, in: *4th Conference on Communication by Gaze Interaction (COGAIN)*, Citeseer, 2008, pp. 59–62.
- [30] J. Weaver, K. Mock, B. Hoanca, Gaze-based password authentication through automatic clustering of gaze points, in: *2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2011, pp. 2749–2754.
- [31] D.H. Cymek, A.C. Venjakob, S. Ruff, O.H.-M. Lutz, S. Hofmann, M. Roetting, Entering PIN codes by smooth pursuit eye movements, *J. Eye Mov. Res.* 7 (2014) 1–11.
- [32] D. Cazzato, A. Evangelista, M. Leo, P. Carcagnì, and C. Distantè, “A low-cost and calibration-free gaze estimator for soft biometrics: An explorative study,” *Pattern Recognition Letters*, vol. 82, pp. 196–206, 2016, doi: 10.1016/j.patrec.2015.10.015.
- [33] Cantoni, V., Galdi, C., Nappi, M., Porta, M., & Riccio, D. (2015). GANT: Gaze analysis technique for human identification. *Pattern Recognition*, 48(4), 1027–1038. <https://doi.org/10.1016/j.patcog.2014.02.017>
- [34] J. Yin, J. Sun, J. Li, and K. Liu, “An effective gaze-based authentication method with the spatiotemporal feature of eye movement,” *Sensors*, vol. 22, no. 8, p. 3002, 2022, doi: 10.3390/s22083002.
- [35] M. Porta, P. Dondi, N. Zangrandi and L. Lombardi, "Gaze-Based Biometrics From Free Observation of Moving Elements," in *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 1, pp. 85–96, Jan. 2022, doi: 10.1109/TBIOM.2021.3130798.

- [36] Omair Shahzad Bhatti, Michael Barz, and Daniel Sonntag. 2021. EyeLogin - Calibration-free Authentication Method for Public Displays Using Eye Gaze. In ACM Symposium on Eye Tracking Research and Applications (ETRA '21 Short Papers). Association for Computing Machinery, New York, NY, USA, Article 14, 1–7. <https://doi.org/10.1145/3448018.3458001>
- [37] Mohamed Khamis, Ludwig Trotter, Ville Mäkelä, Emanuel von Zezschwitz, Jens Le, Andreas Bulling, and Florian Alt. 2018. CueAuth: Comparing Touch, Mid-Air Gestures, and Gaze for Cue-based Authentication on Situated Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 174 (Dec. 2018), 21 pages. <https://doi.org/10.1145/3287052>
- [38] ADIL HAMID MALLA, A GAZE-BASED AUTHENTICATION SYSTEM: FROM AUTHENTICATION TO INTRUSION DETECTION. Submitted to the Office of Graduate and Professional Studies of Texas A&M University in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE.
- [39] Ivo Sluganovic, Marc Roeschlin, Kasper B. Rasmussen, and Ivan Martinovic. 2018. Analysis of Reflexive Eye Movements for Fast Replay-Resistant Biometric Authentication. *ACM Trans. Priv. Secur.* 22, 1, Article 4 (February 2019), 30 pages. <https://doi.org/10.1145/3281745>
- [40] Matin Fallahi, Patricia Arias-Cabarcos, Thorsten Strufe. Beyond Gaze Points: Augmenting Eye Movement with Brainwave Data for Multimodal User Authentication in Extended Reality. <https://doi.org/10.48550/arXiv.2404.18694>
- [41] N. Alsufyani, A. Ali, S. Hoque and F. Deravi, "Biometric presentation attack detection using gaze alignment," 2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA), Singapore, 2018, pp. 1-8, doi:10.1109/ISBA.2018.8311472.
- [42] Dillon Lohr, Michael J. Proulx, Mehedi Hasan Raju, Oleg V. Komogortsev. Ocular Authentication: Fusion of Gaze and Periocular Modalities. <https://doi.org/10.48550/arXiv.2505.17343>
- [43] D'Amelio A, Patania S, Bursic S, Cuculo V, Boccignone G. Using Gaze for Behavioural Biometrics. *Sensors*. 2023; 23(3):1262. <https://doi.org/10.3390/s23031262>
- [44] Virginio Cantoni, Tomas Lacovara, Marco Porta, and Haochen Wang. 2018. A Study on Gaze-Controlled PIN Input with Biometric Data Analysis. In *Proceedings of the 19th International Conference on Computer Systems and Technologies (CompSysTech '18)*. Association for Computing Machinery, New York, NY, USA, 99–103. <https://doi.org/10.1145/3274005.3274029>
- [45] Ali A, Hoque S, Deravi F. Directed Gaze Trajectories for Biometric Presentation Attack Detection. *Sensors*. 2021; 21(4):1394. <https://doi.org/10.3390/s21041394>
- [46] Ali, A., Hoque, S. & Deravi, F. Gaze stability for liveness detection. *Pattern Anal Applic* 21, 437–449 (2018). <https://doi.org/10.1007/s10044-016-0587-2>
- [47] Haochen Wang, Prof. Virginio Cantoni. Gaze-Based Biometrics: Some Case Studies.
- [48] Yasmeen Abdrabou, Ahmed Shams, Mohamed Omar Mantawy, Anam Ahmad Khan, Mohamed Khamis, Florian Alt, and Yomna Abdelrahman. 2021. GazeMeter: Exploring the Usage of Gaze Behaviour to Enhance Password Assessments. In *ACM Symposium on Eye Tracking Research and Applications (ETRA '21 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 9, 1–12. <https://doi.org/10.1145/3448017.3457384>
- [49] Harezlak K, Blasiak M, Kasprowski P. Biometric Identification Based on Eye Movement Dynamic Features. *Sensors*. 2021; 21(18):6020. <https://doi.org/10.3390/s21186020>
- [50] M. Wierzbowski, G. Pochwatko, P. Borkiewicz, D. Cnotkowski, M. Pabiś-Orzeszyna and P. Kobylński, "Behavioural Biometrics in Virtual Reality: To What Extent Can We Identify a Person Based Solely on How They Watch 360-Degree Videos?," 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Singapore, Singapore, 2022, pp. 417-422, doi: 10.1109/ISMAR-Adjunct57072.2022.00090.

- [51] Andrea Casanova, Lucia Cascone, Aniello Castiglione, Weizhi Meng, Chiara Pero, User recognition based on periocular biometrics and touch dynamics, *Pattern Recognition Letters*, Volume 148, 2021, Pages 114-120, ISSN 0167-8655. <https://doi.org/10.1016/j.patrec.2021.05.006>.
- [52] Rubo, M., Son, G. Social gaze fingerprints: identifying social virtual reality users by their eye gaze patterns. *Virtual Reality* 29, 144 (2025). <https://doi.org/10.1007/s10055-025-01210-4>
- [53] D.L. Silver, A. Biggs, Keystroke and eye-tracking biometrics for user identification, in: *IC-AI*, 2006, pp. 344–348.
- [54] C. Holland, O.V. Komogortsev, Biometric identification via eye movement scanpaths in reading, in: *2011 International Joint Conference on Biometrics (IJCB)*, 2011, pp. 1–8.
- [55] C.D. Holland, O.V. Komogortsev, Complex eye movement pattern biometrics: analyzing fixations and saccades, in: *2013 International Conference on Biometrics (ICB)*, 2013.
- [56] I. Rigas, G. Economou, S. Fotopoulos, Biometric identification based on the eye movements and graph matching techniques, *Pattern Recognit. Lett.* 33 (2012) 786–792.
- [57] C. Galdi, M. Nappi, D. Riccio, V. Cantoni, M. Porta, A new gaze analysis based soft biometric, in: J.A. Carrasco-Ochoa, J.F. Martínez-Trinidad, J.S. Rodríguez, G.S. di Baja (Eds.), *Pattern Recognition*, in: *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2013, pp. 136–144.
- [58] V. Cantoni, M. Porta, C. Galdi, M. Nappi, H. Wechsler, Gender and age categorization using gaze analysis, in: *2014 10th International Conference on Signal-Image Technology and Internet-Based Systems (SITIS)*, 2014, pp. 574–579.
- [59] C. Galdi, H. Wechsler, V. Cantoni, M. Porta, M. Nappi, Towards demographic categorization using gaze analysis, *Pattern Recognit. Lett.* (2015) (accessed 16 September 2015).
- [60] R. Biedert, M. Frank, I. Martinovic, D. Song, Stimuli for gaze based intrusion detection, in: J.J. Park, H. Jong, V.C.M. Leung, C.-L. Wang, T. Shon (Eds.), *Future Information Technology, Application, and Service*, in: *Lecture Notes in Electrical Engineering*, Springer, Netherlands, 2012, pp. 757–763.
- [61] A. George, A. Routray, A score level fusion method for eye movement biometrics, *Pattern Recognit. Lett.* (2015) (accessed 2 December 2015).
- [62] R. Bednarik, T. Kinnunen, A. Mihaila, P. Fränti, Eye-movements as a biometric, in: H. Kalviainen, J. Parkkinen, A. Kaarna (Eds.), *Image Analysis*, in: *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2005, pp. 780–789.
- [63] T. Kinnunen, F. Sedlak, R. Bednarik, Towards task-independent person authentication using eye movement signals, in: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10)*, 2010, pp. 187–190.
- [64] I. Rigas, G. Economou, S. Fotopoulos, Human eye movements as a trait for biometrical identification, in: *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2012, pp. 217–222.
- [65] N.V. Cuong, V. Dinh, L.S.T. Ho, Mel-frequency cepstral coefficients for eye movement identification, in: *2012 IEEE 24th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2012, pp. 253–260.
- [66] Z. Liang, F. Tan, Z. Chi, Video-based biometric identification using eye tracking technique, in: *2012 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC)*, 2012, pp. 728–733.
- [67] A. Darwish, M. Pasquier, Biometric identification using the dynamic features of the eyes, in: *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013, pp. 1–6.
- [68] C.D. Holland, O.V. Komogortsev, Complex eye movement pattern biometrics: analyzing fixations

and saccades, in: 2013 International Conference on Biometrics (ICB), 2013.

[69] H.J. Yoon, T.R. Carmichael, G. Tourassi, Gaze as a Biometric (Medical Imaging 2014: Image Perception, Observer Performance, and Technology Assessment), in: Proc. SPIE, vol. 9037, International Society for Optics and Photonics, 2014.

[70] H.J. Yoon, T.R. Carmichael, G. Tourassi, Temporal stability of visual search-driven biometrics, in: Medical Imaging 2015: Image Perception, Observer Performance, and Technology Assessment, International Society for Optics and Photonics, in: Proc. SPIE, vol. 9416, 2015.

[71] M. Juhola, Y. Zhang, J. Rasku, Biometric verification of a subject through eye movements, *Comput. Biol. Med.* 43 (2013) 42–50.

[72] Y. Zhang, J. Laurikkala, M. Juhola, Biometric verification of a subject with eye movements, with special reference to temporal variability in saccades between a subject's measurements, *Int. J. Biom.* 6 (2014) 75

[73] N. Srivastava, U. Agrawal, S.K. Roy, U.S. Tiwary, Human identification using linear multiclass SVM and eye movement biometrics, in: 2015 Eighth International Conference on Contemporary Computing (IC3), 2015, pp. 365–369.

[74] S. Eberz, K.B. Rasmussen, V. Lenders, I. Martinovic, Preventing lunchtime attacks: fighting insider threats with eye movement biometrics, in: 2015 Network and Distributed System Security (NDSS) Symposium, 2015.

[75] N. Nugrahaningsih, M. Porta, Pupil size as a biometric trait, in: V. Cantoni, D. Dimov, M. Tistarelli (Eds.), *Biometric Authentication*, in: Lecture Notes in Computer Science, Springer International Publishing, 2014, pp. 222–233

[76] S. Eberz, K.B. Rasmussen, V. Lenders, I. Martinovic, Preventing lunchtime attacks: fighting insider threats with eye movement biometrics, in: 2015 Network and Distributed System Security (NDSS) Symposium, 2015.

[77] P. Kasprowski, J. Ober, Eye movements in biometrics, in: D. Maltoni, A.K. Jain (Eds.), *Biometric Authentication*, in: Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2004, pp. 248–258.

[78] O.V. Komogortsev, A. Karpov, L.R. Price, C. Aragon, Biometric authentication via oculomotor plant characteristics, in: 2012 5th IAPR International Conference on Bio metrics (ICB), 2012, pp. 413–420.

[79] O.V. Komogortsev, A. Karpov, C.D. Holland, H.P. Proenca, Multimodal ocular biometrics approach: a feasibility study, in: 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2012, pp. 209–216.

[80] Z. Liang, F. Tan, Z. Chi, Video-based biometric identification using eye tracking technique, in: 2012 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC), 2012, pp. 728–733.

[81] Harezlak K, Pluciennik E. Towards Improved Eye Movement Biometrics: Investigating New Features with Neural Networks. *Sensors*. 2025; 25(14):4304. <https://doi.org/10.3390/s25144304>

[82] Jonathan Liebers, Patrick Horn, Christian Burschik, Uwe Gruenefeld, and Stefan Schneegass. 2021. Using Gaze Behavior and Head Orientation for Implicit Identification in Virtual Reality. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology (VRST '21)*. Association for Computing Machinery, New York, NY, USA, Article 22, 1–9. <https://doi.org/10.1145/3489849.3489880>

[83] B. L. Tait, "Behavioural Biometrics Authentication Tested Using EyeWriter Technology," 2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3), London, UK, 2019, pp. 1-9, doi: 10.1109/ICGS3.2019.8688257.

[84] Asish SM, Kulshreshth AK, Borst CW. User Identification Utilizing Minimal Eye-Gaze Features in Virtual Reality Applications. *Virtual Worlds*. 2022; 1(1):42-61.

<https://doi.org/10.3390/virtualworlds1010004>

[85] Sherif Nagib Abbas Seha, Dimitrios Hatzinakos, Ali Shahidi Zandi, Felix J.E. Comeau, Improving eye movement biometrics in low frame rate eye-tracking devices using periocular and eye blinking features, Image and Vision Computing, Volume 108, 2021, 104124, ISSN 0262-8856, <https://doi.org/10.1016/j.imavis.2021.104124>.

[86] M. Wierzbowski, G. Pochwatko, P. Borkiewicz, D. Cnotkowski, M. Pabiś-Orzeszyna and P. Kobyliński, "Behavioural Biometrics in Virtual Reality: To What Extent Can We Identify a Person Based Solely on How They Watch 360-Degree Videos?," 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Singapore, Singapore, 2022, pp. 417-422, doi: 10.1109/ISMAR-Adjunct57072.2022.00090.

[87] Cazzato D, Carcagnì P, Cimorelli C, Voos H, Distanto C, Leo M. Ocular Biometrics Recognition by Analyzing Human Exploration during Video Observations. Applied Sciences. 2020; 10(13):4548. <https://doi.org/10.3390/app10134548>

[88] Jim Holdsworth, Matthew Kosinski, What is behavioral biometrics?  
<https://www.ibm.com/think/topics/behavioral-biometrics>

[89] A. George and A. Routray, "A score level fusion method for eye movement biometrics," Pattern Recognit. Lett., vol. 82, pp. 207–215, Oct. 2016.

[90] OPEN GAZE API BY GAZEPOINT,  
<http://andrewd.ces.clemson.edu/courses/cpsc881/manuals/Gazepoint/Gazepoint%20API.pdf>

[91] SelectKBest,  
[https://scikit-learn.org/stable/modules/generated/sklearn.feature\\_selection.SelectKBest.html](https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html)

[92] Naive Bayes, [https://scikit-learn.org/stable/modules/naive\\_bayes.html](https://scikit-learn.org/stable/modules/naive_bayes.html)

[93] Introduction to Machine Learning with Python, Andreas C. Müller and Sarah Guido

[94] KNeighborsClassifier,  
<https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

[95] Logistic Regression,  
[https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)

[96] Advantages and Disadvantages of Logistic Regression,  
<https://www.geeksforgeeks.org/data-science/advantages-and-disadvantages-of-logistic-regression/>

[97] SVC, <https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>

[98] NuSVC, <https://scikit-learn.org/stable/modules/generated/sklearn.svm.NuSVC.html>

[99] Underrated and Implementing Nu-Support Vector Classification (NuSVC), <https://www.geeksforgeeks.org/machine-learning/underrated-and-implementing-nu-support-vector-classification-nusvc/>

[100] Random Forest,  
<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

[101] MLP, [https://scikit-learn.org/stable/modules/generated/sklearn.neural\\_network.MLPClassifier.html](https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html)

[102] train\_test\_split,  
[https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.train\\_test\\_split.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html)

[103] Pipeline, <https://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html>

[104] GridSearchCV,  
[https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html)

[105] SimpleImputer, <https://scikit-learn.org/stable/modules/generated/sklearn.impute.SimpleImputer.html>

- [106] MinMaxScaler,  
<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>
- [107] StandardScaler,  
<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>
- [108] RobustScaler,  
<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html>