Optimal Passwordless Continuous Authentication Measures for Remote Employees

Manya Narkar
University of Twente
Tel.: 57-7-5685303, Fax: 57-7-5685303, Ext. 144
E-mail: m.a.narkar@student.utwente.nl

Abstract: Considering the increased reliance on remote employment systems, the need to prioritise safety and foster a productive environment is imperative. The choice of authentication measure plays a crucial role in maintaining this balance. However, current research in the field lacks an exploration of the trade-offs between intrusive and non-intrusive authentication measures. Additionally, most research at the moment consists of comparing authentication measures within similar categories (such as touchscreen and motion-based authentication with other touchscreen and motion-based authentication), or across dissimilar categories (such as accelerometer-based gait authentication with device profiling authentication). The latter consists of not only varying performance metrics (such as Equal Error Rate (EER) and False Rejection Rate (FRR)), but also varying usability testing sample sizes and integrated development environments, among others. This results in challenges in conducting a fair and unbiased comparison of the systems. Our study addresses these notable gaps by assessing the trade-offs between intrusive and non-intrusive authentication measures within a homogenised environment to produce unbiased results. Through an extensive review of existing literature, we first identify existing authentication measures that are better suited for remote employees. Then, a web-app that implements each of these measures is developed, by minimising discrepancies across each system, so that they can be evaluated through testing. Finally, the results are analysed to conclude which type of system is a better fit for continuous authentication for remote employees.

Keywords: Continuous Authentication, Security, Intrusive authentication, Non-intrusive authentication, Hybrid authentication, Remote Employees.

1. INTRODUCTION AND LITERATURE REVIEW

The existence of COVID-19 has brought about a change in workplace dynamics, with Eurostat [2022][24] reporting a notable increase from 5.5% in 2019, to 13.5% in 2021 among individuals aged 20-64 that began working from home. With this significant increase, the demand for efficacious authentication measures is critical. Password vulnerabilities have been highlighted as one of the most significant risk factors for security [29]. Considering passwords are the root cause of over 80% of breaches, passwordless authentication is encouraged. Continuous authentication (CA) emerges as a robust security measure since unlike traditional measures that exist solely at entry points, CA monitors user/device behaviour *continuously* [7]. It accounts for changing risk factors such as network anomalies, environmental changes and other behavioural data. This authentication, can then be divided into: intrusive and non-intrusive authentication. The former consists of methods that require active participation and interaction from the user (such as answering a security question), and the latter involves minimal disruption (such as extraction of device posture). While numerous proposals for continuous authentication exist [2], a gap persists in comparing intrusive and non-intrusive authentication measures. Furthermore, existing studies exhibit much heterogeneity in methodologies, hindering a fair assessment of the effects of the authentication measures compared. Furthermore, according to a study by Al-Sharafi et al.[2016][3], users may not express significant concerns about security vulnerabilities unless they are seen to have a direct impact on a users' account. This makes user preferences a critical aspect to explore. Motivated by these gaps in current research, our study aims to compare the trade-offs between intrusive and non-intrusive authentication measures whilst considering security as well as user-convenience. We aim to do so by developing a web-app system that implements these authentication measures for testing through usability testing and performance metrics. The system further overcomes challenges posed by methodological heterogeneity by means such as homogenising the Integrated Development Environment (IDE) and language for developing the systems, using similar performance metrics, etc. This research introduces a novel approach in not only analysing the better fit for remote employees between intrusive, non-intrusive and hybrid systems, but by developing a uniform framework that enables fair assessment of the same. The main contributions of this research study are as follows:

- Proposes research questions to address the different aspects this study aims to cover.
- Provides an in-depth analysis to answer the aforementioned research questions.
- Summarises the findings and discusses the most viable authentication measure.

- Examines research limitations and identifies areas for future improvement.

2. RESEARCH QUESTIONS

The main research question for this proposal is as follows: What are the trade-offs between intrusive and nonintrusive authentication measures and how can their security and convenience be assessed in terms of continuous authentication? This main research question can be further divided into sub-research questions. They are as follows:

RQ1: On what basis can representative intrusive and non-intrusive authentication measures be selected?

RQ2: How can a system be developed and evaluated for these authentication measures?

RQ2.1: How will these authentication measures be implemented?

RQ2.2: How can evaluating these measures be homogenised during and after implementation?

RQ3: What measures are considered for evaluating the security and convenience of the aforementioned systems?

3. ADDRESSING RESEARCH QUESTIONS

This section addresses the methodologies used to answer each research question. It further concludes each subsection with results, consequently answering their respective questions.

3.1. On Answering RQ1

To understand the bases on which we select the intrusive and non-intrusive authentication measures, we engaged in a literature review of some common biometric authentication measures that can be collected by users. These include: DNA verification, facial recognition, iris and retina recognition, vein structure and signature [28]. Now, considering the five characteristics of biometrics [17]: universal,distinctive,persistent,collectable and unique, the initial eight biometrics can be narrowed down to three, namely, facial, voice and iris recognition. DNA verification, vein structure and signature can be disregarded since these biometrics are not 'collectable' by an average laptop possessed by a remote employee. A filter which is relevant when considering *remote* work is a *lightweight* security solution [12]. Since common devices possessed by users have limited resources such as computing power, memory, and sensor quality, among others, the solution must be adept at working within these constraints. Iris recognition, while collectable, according to the National CyberSecurity Centre (NCSC)[22], relies on infrared cameras for capturing images with sufficient detail. Although not hard to find, most laptops today do not come equipped with infrared cameras making this a non-practical solution. Feasible behavioral biometrics as suggested by Tripathi [2011][28] also consist of: keystroke analysis - which considering the task, can be considered a non-intrusive authentication measure. Ometov et al. [2018][21], mentions three types of factor groups to connect individuals to credentials: knowledge, ownership and biometric factors. Among these, due to their inherent probabilistic nature, biometric authentication measures are not very straightforward in satisfying the binary decision mechanism. A suggested measure that does satisfy this, resulting in increased security, is the One Time Password (OTP)[21]. Time-based OTP was also one of the highest-ranked two-factor authentication measures in a usability survey Reese et al. [2019][23] - rendering it an appropriate fit for remote authentication.

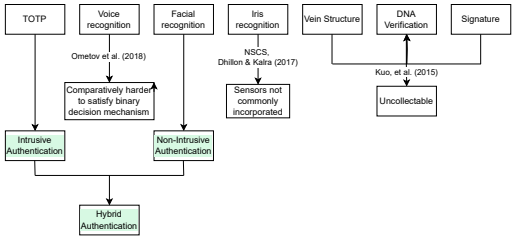

Figure 1: . *Authentication measure selection process.*

3.2. On Answering RQ2.1

After the selection of authentication measures, a system implementing each of these non-intrusive (system A, with facial recognition), intrusive (system B, with OTP) and hybrid (system C, with both facial recognition and OTP) measures, in the same environment was developed [1] [21], using Python, HTML and Javacript. The framework used to implement the systems is Flask. Figure 2 displays the sequence of interactions a user can have with the system. Users started by creating an account with a unique username, password and clicked 4 pictures of themselves via a prompt with instructions. After that, they are introduced to the pages of systems A,B and C, sequentially as they complete tasks. Each page had the same tasks, namely, image description and math problem solving, that users were required to complete during user testing to mimic a productive environment. Each page had its respective authentication measure continuously authenticating the user in the background as they completed the tasks.

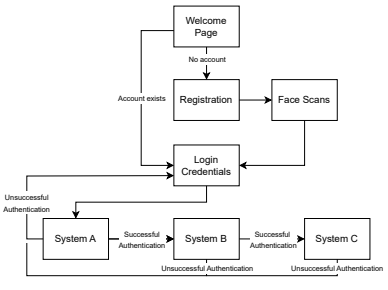

Figure 2: . *User-system interaction sequence .*

[1] https://gitlab.utwente.nl/s2297809/AuthenticationMeasureResearch

3.2.1. Facial Recognition

The first system is that of facial recognition which utilises the Local Binary Patterns Histogram (LBPH) algorithm. It's robustness and accuracy enable it to outperform other Euclidian distance-based algorithms such as Eigenfaces and Fisherfaces [25]. Considering the diversity of remote employees' camera capabilities, LBPH's effectiveness in scenarios with low-resolution cameras - producing 90% accuracy at 35 pixels and 94% at 45 pixels - as demonstrated by Ahmed et al. [2018][1] is crucial. Training the classifier involves using a subset of the Labeled Faces in the Wild Home (LFW)[30] dataset due to its inclusion of multiple images per person, which in turn enhances robustness [27]. Although the complete LFW dataset consists of 13,323 images, the model being retrained after each registration necessitated a manageable subset due to time constraints. Therefore, the dataset used in training consists of 53 individuals, each contributing 3-5 pictures. It is noteworthy, that training on even this limited dataset required approximately 11.191 seconds. In a real-world setting, users would submit appropriate photos for registration ahead of time, allowing for extended processing time without disrupting their workflow.

3.2.2. OTP Authentication

The second system is that of an OTP. The libraries used for this include the random and smtplib modules. In a real-world scenario, users would log in using their individual email accounts. However, during usability testing, to expedite the testing process and minimize the collection of personal information, a dedicated dummy email account is created. Users are given access to this email, where they will receive OTPs via Simple Mail Transfer Protocol, secured by Transport Layer Security which encrypts the same message. They are required to enter these OTPs into a prompt on their main work page with the task.

3.2.3. Hybrid Authentication

The third system combines these authentication measures. The user is authenticated by means of facial recognition at first, and in the case that it fails, they are asked to input the OTP emailed to them. Figure 3 demonstrates each system with its authentication measure.

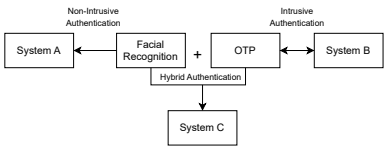

Figure 3: *Composition of each system with respective authentication measures*

[2]https://apscheduler.readthedocs.io/en/3.x/

3.3. On Answering RQ2.2

Although the authentication measures have been selected, uniformity must be realised during and after the development of the web-app they are integrated in. The main challenge currently is that different research proposals have different setups for their researches, so results are obtained under non-uniform conditions. There are a few ways to maintain homogeneity and ensure this did not happen in this research:

1. Programming environment: During the development stage, it is important to deploy the authentication measures in the same environment. This helps overcome biases that can be introduced by differences in complexities due to libraries, syntax, etc. For this reason, the authentication measures are both implemented in Python due to its vast ecosystem of libraries and frameworks, in the VSCode Integrated Development Environment. They are also integrated into the Flask web application framework, establishing uniform communication protocols between the pages and the logic for both authentication measures. The threading logic implemented in both measures is nearly identical and makes use of global locks. Furthermore, the OTP and facial recognition authentication are both scheduled tasks, being invoked at equal intervals by leveraging Pythons APScheduler library [2].

2. Size and composition of users for testing: To promote an unbiased assessment, the user base during the testing phase is kept the same for all systems being compared. This ensures consistency across the number of users as well as their demographics. Alroobaea and Mayhew [2014][4] recommend moderate sample size of 16±4 is sufficient to ensure a comprehensive exploration of the system and is more appropriate for comparative studies. Since our research which is based on comparing various authentication systems aligned with their recommendation, we chose a sample size of 21 which was approximately equal to 16±4. Every user is asked to test all three systems: intrusive, non-intrusive, and hybrid. This ensures the same variation in demographics across all systems.

3. User interaction with the system: It should be ensured that, save for the interaction required for authentication, the users are required to complete the same tasks and interact with the systems in the same way. In the current system, every user is asked to complete two tasks for each type of system: image description, and a few basic math problems. The kind of image and problems vary across each system, but the type of tasks are the same throughout.

4. Performance metrics: It is crucial to employ the same type of evaluation metrics across the systems to assess their performance fairly. For example, the survey by Al-Naji and Zagrouba [2020][2] consists of some proposals using only

accuracy, some using only EER, some only FAR and so on to evaluate their systems. In order to promote comparability, this research has the same metric used for both systems, namely, accuracy. The reason for this selection is discussed in the next section

3.4. On answering RQ3

Now that the system had been developed, it could be evaluated. Evaluation of the previously selected authentication measures boils down to two main factors: security and convenience. This section will explore the methodology and results of both aspects.

3.4.1. Security

For the security evaluation of the system, aligning with the goals of the research, it is important to establish a common ground between the systems for comparison. Although according to V.L.B. De Mel [2023][11], metrics such as False Acceptance Rate (FAR), False Positive Rate (FPR), etc. are common in facial recognition systems, it is difficult to apply these to an OTP based system. Therefore, for the purpose of a fair comparison, the primary security evaluation metric considered in this study is accuracy. Carvalho et al. [2019][8] describe accuracy as how well a system predicts unseen data. The facial recognition-based authentication system, despite parameter fine-tuning, exhibited a modest accuracy of 54%. We experimented with the detectMultiScale() [3][9] methods parameters: scaleFactor and minNeighbours as can be seen in Table 1.

Table 1: This table visualizes parameter fine-tuning for detectMultiScale()

No.	SF	MN	Accuracy	Images detected
1.	1.8	10	75%	4
2.	1.5	7	66%	9
3.	1.3	5	60%	10
4.	1.2	10	54%	19
5.	1.2	5	38%	19
6.	1.1	7	37.5%	19
2.	1.01	10	23%	26

The method is responsible for detecting objects of varying sizes in the input images. The scaleFactor parameter specifies how much the image size is reduced - higher values favour larger faces while lower make the system more sensitive to smaller faces. The minNeighbours parameter determines the number of neighbours a region needs to be considered a face. Lower values result in more detections but with a higher FPR, and higher values detect fewer faces but with increased accuracy. The last two columns in Table 1 reveal a trade-off between accuracy and total faces detected. Higher scaleFactor values increased accuracy but reduced the total number of faces detected (out of 19 test images). This was observed regardless of changes in minNeighbours. To address this trade-off, we selected highest values that detected all faces successfully and introduced confidence intervals. A study by Hadi et al.[2022][13] is relevant in considering the confidence levels due to their usage of n Asus x455LJ laptop with Intel(R) Core(TM) i3-5010U CPU. The confidence level of 62% identified in their study on the lower-range laptop suggests that a similar level may perform well on mid-range laptops commonly used by remote employees. Therefore our test system utilised a scaleFactor of 1.2, minNeighbour value of 10 and confidence level of above 60%.

In contrast to this, the OTP-based authentication system, while lacking a traditional accuracy metric due to the nature of its functioning, still displays higher security than facial recognition in the sense that it has an accuracy of 100% because it will never allow unauthorised access. This assertion, however, is based on the premise that no external account, such as an e-mail ID, is additionally compromised.

3.4.2. Convenience

During usability testing, after the users were done interacting with the three systems, they were asked to complete a questionnaire [4] [20]. The convenience of the system was measured via this survey which included Likert scale questions gauging user perceptions regarding aspects such as productivity, hindrance and ease of use, among others. A set of 6 questions recording the aforementioned factors were repeated for every system. The questions asked can be observed in Appendix A. According to Lazar et al. [2017][18], some of the most common criteria for determining the representativeness of users include: age, gender, education and job among others. Our target group being remote workers was encompassing of users of different age groups as well as occupational and ethnic backgrounds. This enabled the accommodation of a representative sample.

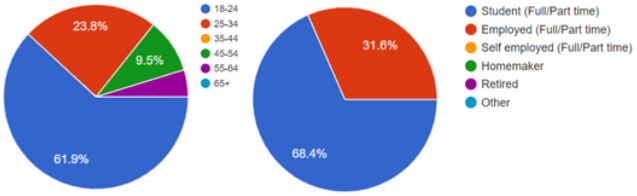

Figure 4: . *Distribution of participant ages (left) and educational/occupational status.*

[3] https://docs.opencv.org/3.4/d1/de5/classcv_11_CascadeClassifier.html

[4] https://docs.google.com/spreadsheets/d/15uZumNpBTCf105ThQmNoVVZLPPqfzMphSeGYbOqFuW4/edit?usp=sharing

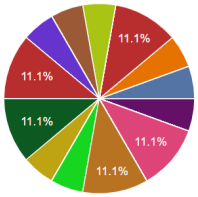

Figure 5: . *Distribution of participant ethnicities including: UK, Germany, India, Jordan, Nigeria, Romania, Russia, Pakistan, Vietnam, Qatar, Cyprus, Indonesia, Japan.*

3.4.3. Approach to Analysing Survey Results

Once the survey results were collected, it was time to analyse them. "Within-subject" design is that which requires each participant of usability testing to be exposed to multiple systems [18]. Since the users were asked to interact with all three systems and then answer the questionnaire, this study used a within-subject approach to the usability testing. The goal of the statistical analysis is to understand whether the users have differing opinions about either of the proposed systems. Further according to Lazar et al. (2017)[18], significance tests to compare the means of multiple groups include: t-tests and Analysis of Variance (ANOVA) tests. Since the t-test is a pairwise test, we utilise the ANOVA test in order to compare the three systems. Since we have only one independent variable (ie., type of system: A,B,C), we are to conduct a one-way ANOVA test.

3.4.4. Statistical Analysis of Survey Results

The one way ANOVA test was done using the Statistical Package for the Social Sciences (SPSS) software [5]. The data which was originally saved in the .xlsx spreadsheet file was then opened with SPSS. In SPSS, the equivalent of a within-group one-way ANOVA test is a 'Repeated Measures ANOVA test'. Executing said test resulted in multiple tables, two of which were of considerable importance. These were Mauchly's Test of Sphericity [6][6] and the Multivariate test [7]. For an alpha (α) - which is the threshold for significance - value of 5% or 0.05, The Mauchly test results (as in Figure 9 in the Appendix) indicated consistent user preferences across three systems. The Multivariate test results (Figure 10. in the appendix) indicated a significant difference between results of each surveyed system [8][14]. To get more information regarding the differences among systems, a class of post-hoc tests known as Multiple Comparison Analysis was used, the most commonly used MCA statistics being: Tukey, Bonferroni and Dunnett [31]. Notable findings for some questions (Figure 11 of the appendix) include:

- Q1: For the first question, a significant difference of p=0.026 [9] was observed between systems B and C. System C further showed a higher mean score than A and B signifying more positive opinions.
- Q2: For the second question, significant differences between systems A and B, as well as B and C were observed. System A was preferred in comparison to system B, as was system C. System A further had the highest mean score among the three.
- Q3: For the third question, a significant difference of p=0.010 was observed between systems B and C. System C was preferred to both A and B, although not significantly to A.

For the rest of the questions, ie., Q4,Q5 and Q6, although not significantly, the mean score of C was always higher than its counterparts. This entails an overall more positive outlook towards system C. It is preferred over the other two in the last three questions as well. Referring to the two close-ended questions in the survey that asked users which system they thought struck a balance between convenience and safety, 80.9% answered with system C, 14% with system A and 14% with system B, since users were allowed to select more than one option. Regarding deployment preference, 38% favoured system A emphasizing ease of use while specifying their opinion would only hold if accuracy was guaranteed. No individuals from the 25-34 group preferred this. System C was favoured by 66.6%, citing safety and convenience in the sense that they would not be too inconveniently affected (logged out) in the case of a faulty face detection, and would have to rely on a second authentication measure which they mentioned did not affect their productivity much. Individuals who chose system B, 9.5%, also agreed by mentioning that they chose the same because it enabled them to stay on the webpage in case of a faulty face detection.

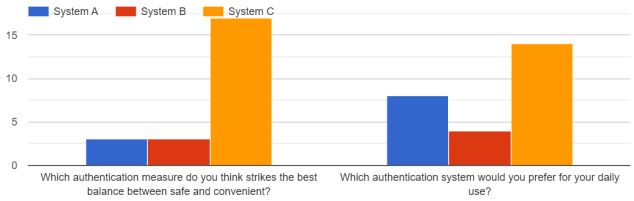

Figure 6: . *This graph displays responses to the system they think best strikes the balance, and which one they would actually deploy.*

[5] https://www.ibm.com/products/spss-statistics
[6] Blanca et al. (2023).
[7] https://www.ibm.com/support/pages/ibm-spss-statistics-28-documentation
[8] https://www.ibm.com/support/pages/ibm-spss-statistics-28-documentation
[9] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4111019/

3.4.5. User Feedback

While the questionnaire results played a significant role in the research study, the open question at the end of the survey, and one-on-one conversations with the participants after their testing also provided helpful insights of the user experience. This highlights user concerns and suggestions.

- Duration of authentication: This feedback was provided not only during the one-on-one conversations with a few users, but also in the open-ended question response in the survey. Some users expressed concerns regarding the frequency of authentication, suggesting it could be distracting during extended work sessions. We clarified that the intervals were shortened for usability testing, but in practical scenarios they would be extended to minimise disruptions in workflow.

- Type of tasks: Around the 10-12 user mark, some users began expressing their distaste for math. Regrettably, changing tasks at that stage would necessitate restarting the usability testing process. Due to time constraints, task modifications at this stage were not possible. Potential solutions are further discussed in the 'discussion' section for future research.

4. CONCLUSION

In conclusion, this research set out to examine the most appropriate type of authentication system between intrusive, non-intrusive and hybrid measures for remote employees by evaluating their security and convenience. While considering the 54% accuracy of system A, it may be essential to note that the training set for this study was no more than 300, with 53 individuals having 3-5 pictures of them each. Another factor worth noting is the quality of the images used in training from the LFW dataset. Most of the pictures were of low resolution, posing as a potential challenge in achieving a higher accuracy. Despite the theoretical accuracy, system A reported a lower fail rate: with only 5/21 users experiencing failed facial recognition during usability testing conducted in the study. Nevertheless, 66% of users vouched for the deployment of hybrid system C in their daily lives. Although users reported to being able to multitask better with system A (Q5), system C further possessed a higher mean score for every other question asked, indicating a more positive user response than the other systems did. For questions 1,2 and 3, significant differences were observed between systems B and C, the latter possessing a higher mean. In terms of security, the facial recognition system, system A, was able to produce a 54% accuracy as previously mentioned. System B and the hybrid system C, produced a 100% accuracy, preventing any unauthorized access. It is noteworthy to consider that most user responses that vouched for the deployment of system A indicated their preference would hold if the system had higher accuracy. Achieving an accuracy of 100% in facial recognition has inherent challenges that underscore its limitations [5, 16]. Furthermore, the intrusive authentication aspect of system C aligns with the binary decision framework [21] which ensures a clear distinction between correct and incorrect inputs, eliminating the possibility of probabilistic outcomes which contribute to ambiguity. Although it compromises on convenience, convenience results from the survey (specifically for Q5) prove users would prefer to compromise on convenience for safety, than the other way round. Hence, we can conclude that hybrid systems of authentication are the most appropriate authentication measure for remote employees in terms of security as well as user preference.

5. DISCUSSION

Although findings provide valuable insights, it is essential to acknowledge the limitations and propose scope for future research and improvement. This section will do the same.

- Sample Size for Usability Testing: Although 16±4 was indeed a recommended size, the current research was time-bound, limiting the extent to which an even distribution across demographics could be ensured. Future research could benefit not only from a larger sample space, but also an evenly distributed demographic that ensures an approximately even number of employed individuals, males, females, and individuals of differing age groups are selected.

- User feedback: As mentioned in the previous feedback section, it was found midway through usability testing that some of the users found some of the tasks to be stressful. In the future, a system could be developed where users are given the option to choose between various tasks, such that they may not be required to complete those that they find difficult or pressurising. For example, tasks could vary between literature, math, sciences, spotting the difference exercises, etc.

- Refining facial recognition: Given the time constraints of the current research, there exists an opportunity for future investigations to dedicate more time to refining the facial recognition system. This could involve experimenting with various classifying algorithms and datasets (the Yale dataset, for example) or simply fine-tuning different parameters for a specific algorithm further. An additional suggestion would be to use various cascades for facial recognition instead of just one. The side profile and smile cascades could be used to further improve the accuracy and account for not only varying orientations of the face, but features as well.

- Security Assumption: This study assumes the accuracy of OTP-based systems provided there is no other account such as an email being compromised. Future studies should take into account the likelihood of these accounts being compromised to further realise and enhance OTP-based authentication.

6. REFERENCES

[1] Ahmed, Aftab, Jiandong Guo, Fayaz Ali, Farha Deeba, and Awais Ahmed. 2018. "LBPH based improved face recognition at low resolution." In *2018 international conference on Artificial Intelligence and big data (ICAIBD)*, 144–147. IEEE.

[2] Al-Naji, Fatimah Hussain, and Rachid Zagrouba. 2020. "A survey on continuous authentication methods in Internet of Things environment." *Computer Communications* 163: 109-133. Elsevier.

[3] Al-Sharafi, Mohammed A, Ruzaini A Arshah, EA Abo-Shanab, and N Elayah. 2016. "The effect of security and privacy perceptions on customers' trust to accept internet banking services: An extension of TAM." *Journal of Engineering and Applied sciences* 11(3): 545–552.

[4] Alroobaea, Roobaea, and Pam J Mayhew. 2014. "How many participants are really enough for usability studies?" In *2014 Science and Information Conference*, 48–56. IEEE.

[5] Anwarul, Shahina, and Susheela Dahiya. 2020. "A comprehensive review on face recognition methods and factors affecting facial recognition accuracy." In *Proceedings of ICRIC 2019: Recent Innovations in Computing*, 495–514. Springer.

[6] Blanca, María J, Jaume Arnau, F Javier García-Castro, Rafael Alarcón, and Roser Bono. 2023. "Repeated measures ANOVA and adjusted F-tests when sphericity is violated: which procedure is best?" *Frontiers in Psychology* 14. Frontiers Media SA.

[7] Brocardo, Marcelo Luiz, Issa Traore, and Isaac Woungang. 2014. "Toward a framework for continuous authentication using stylometry." In *2014 IEEE 28th international conference on advanced information networking and applications*, 106–115. IEEE.

[8] Carvalho, Diogo V, Eduardo M Pereira, and Jaime S Cardoso. 2019. "Machine learning interpretability: A survey on methods and metrics." *Electronics* 8(8): 832. MDPI.

[9] CV::Cascadeclassifier class reference (no date) OpenCV. Available at: `https://docs.opencv.org/3.4/d1/de5/classcv_1_1CascadeClassifier.html` (Accessed: 18 January 2024).

[10] Dahiru, T. (2011) 'P-value, a true test of statistical significance? A cautionary note', Annals of Ibadan Postgraduate Medicine, 6(1). doi:10.4314/aipm.v6i1.64038.

[11] De Mel, VLB. "Survey of Evaluation Metrics in Facial Recognition Systems."

[12] Dhillon, Parwinder Kaur, and Sheetal Kalra. 2017. "A lightweight biometrics based remote user authentication scheme for IoT services." *Journal of Information Security and Applications* 34: 255–270. Elsevier.

[13] Hadi H, Radiles H, Susanti R, Mulyono, M. "Human Face Identification Using Haar Cascade Classifier and LBPH Based on Lighting Intensity." 2022. *ejournal*. Jul. `https://ejournal.uin-suska.ac.id/index.php/IJAIDM/article/view/15245`.

[14] "IBM SPSS Advanced Statistics 28." IBM, Jan 2024. `https://www.ibm.com/support/pages/ibm-spss-statistics-28-documentation`.

[15] IBM SPSS statistics, 17 Jan 2024. Available at: https://www.ibm.com/products/spss-statistics (Accessed: 28 January 2024).

[16] Introna, Lucas D., and Helen Nissenbaum. 2009. "Facial Recognition Technology: A survey of policy and Implementation Issues." *SSRN*. Jul. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1437730`.

[17] Kavianpour, Sanaz, Bharanidharan Shanmugam, Sami Azam, Mazdak Zamani, Ganthan Narayana Samy, Friso De Boer, and others. 2019. "A systematic literature review of authentication in Internet of Things for heterogeneous devices." *Journal of Computer Networks and Communications* 2019. Hindawi.

[18] Lazar, Jonathan, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research methods in human-computer interaction*. Morgan Kaufmann.

[19] Narkar, M. (2024) Authentication Measure Research Gitlab Repository `https://gitlab.utwente.nl/s2297809/AuthenticationMeasureResearch`

[20] Narkar, M. (2024). Survey Results. `https://docs.google.com/spreadsheets/d/15uZumNpBTCf105ThQmNoVVZLPPqfzMphSeGYbOqFuW4/edit?usp=sharin`

[21] Ometov, Aleksandr, Sergey Bezzateev, Niko M"akitalo, Sergey Andreev, Tommi Mikkonen, and Yevgeni Koucheryavy. 2018. "Multi-factor authentication: A survey." *Cryptography* 2(1): 1. MDPI.

[22] "Biometric recognition and authentication systems." Jan 2019. `https://www.ncsc.gov.uk/collection/biometrics/iris`.

[23] Reese, Ken, Trevor Smith, Jonathan Dutson, Jonathan Armknecht, Jacob Cameron, and Kent Seamons. 2019. "A usability study of five two-factor authentication methods." In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, 357–370.

[24] "Rise in EU population working from home - Products Eurostat News - Eurostat." Nov 2022. `https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20221108-1`.

[25] Samet, Refik, and Muhammed Tanriverdi. 2017. "Face recognition-based mobile automatic classroom attendance management system." In *2017 International conference on cyberworlds (CW)*, 253–256. IEEE.

[26] St›hle, L. and Wold, S. (1989) ‘Analysis of variance (ANOVA)’, Chemometrics and Intelligent Laboratory Systems, 6(4), pp. 259–272. doi:10.1016/0169-7439(89)80095-4.

[27] Tan, Xiaoyang, Songcan Chen, Zhi-Hua Zhou, and Fuyan Zhang. 2006. ”Face recognition from a single image per person: A survey.” *Pattern recognition* 39(9): 1725–1745. Elsevier.

[28] Tripathi, KP. 2011. ”A comparative study of biometric technologies with reference to human interface.” *International Journal of Computer Applications* 14(5): 10–15. Citeseer.

[29] Yıldırım, M, and Ian Mackie. 2019. ”Encouraging users to improve password security and memorability.” *International Journal of Information Security* 18: 741–759. Springer.

[30] Ramesh, M. et al. (2007) Labeled faces in the wild home, LFW Face Database : Main. Available at: https://vis-www.cs.umass.edu/lfw/ (Accessed: 20 December 2023).

[31] McHugh, M.L. (2011) ‘Multiple comparison analysis testing in ANOVA’, Biochemia Medica, pp. 203–209. doi:10.11613/bm.2011.029.

7. APPENDIX

A.
Post-Hoc Tests Table

Multiple Comparisons

Dependent Variable		(I) Type of System	(J) Type of System	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Q1_Num	Tukey HSD	A	B	.4762	.33962	.346	-.3400	1.2924
			C	-.4286	.33962	.422	-1.2448	.3876
		B	A	-.4762	.33962	.346	-1.2924	.3400
			C	-.9048*	.33962	.026	-1.7209	-.0886
		C	A	.4286	.33962	.422	-.3876	1.2448
			B	.9048*	.33962	.026	.0886	1.7209
	Bonferroni	A	B	.4762	.33962	.498	-.3603	1.3127
			C	-.4286	.33962	.636	-1.2650	.4079
		B	A	-.4762	.33962	.498	-1.3127	.3603
			C	-.9048*	.33962	.030	-1.7412	-.0683
		C	A	.4286	.33962	.636	-.4079	1.2650
			B	.9048*	.33962	.030	.0683	1.7412
	Sidak	A	B	.4762	.33962	.420	-.3580	1.3104
			C	-.4286	.33962	.510	-1.2628	.4056
		B	A	-.4762	.33962	.420	-1.3104	.3580
			C	-.9048*	.33962	.029	-1.7390	-.0706
		C	A	.4286	.33962	.510	-.4056	1.2628
			B	.9048*	.33962	.029	.0706	1.7390
Q2_Num	Tukey HSD	A	B	.7619*	.23002	.004	.2091	1.3147
			C	.0476	.23002	.977	-.5052	.6004
		B	A	-.7619*	.23002	.004	-1.3147	-.2091
			C	-.7143*	.23002	.008	-1.2671	-.1615
		C	A	-.0476	.23002	.977	-.6004	.5052
			B	.7143*	.23002	.008	.1615	1.2671
	Bonferroni	A	B	.7619*	.23002	.005	.1954	1.3284
			C	.0476	.23002	1.000	-.5189	.6142
		B	A	-.7619*	.23002	.005	-1.3284	-.1954
			C	-.7143*	.23002	.009	-1.2808	-.1478
		C	A	-.0476	.23002	1.000	-.6142	.5189
			B	.7143*	.23002	.009	.1478	1.2808
	Sidak	A	B	.7619*	.23002	.005	.1969	1.3269
			C	.0476	.23002	.996	-.5174	.6126
		B	A	-.7619*	.23002	.005	-1.3269	-.1969
			C	-.7143*	.23002	.009	-1.2793	-.1493
		C	A	-.0476	.23002	.996	-.6126	.5174
			B	.7143*	.23002	.009	.1493	1.2793

Figure 7: . *Post-Hoc questions 1,2. Leftmost column is representative of the question asked. Columns I and J represent the systems being compared in each row. p-values are to be found under the 'Sig' column.*

Q3_Num	Tukey HSD	A	B	.5238	.31419	.226	-.2313	1.2789
			C	-.4286	.31419	.366	-1.1836	.3265
		B	A	-.5238	.31419	.226	-1.2789	.2313
			C	-.9524*	.31419	.010	-1.7074	-.1973
		C	A	.4286	.31419	.366	-.3265	1.1836
			B	.9524*	.31419	.010	.1973	1.7074
	Bonferroni	A	B	.5238	.31419	.302	-.2500	1.2976
			C	-.4286	.31419	.533	-1.2024	.3453
		B	A	-.5238	.31419	.302	-1.2976	.2500
			C	-.9524*	.31419	.011	-1.7262	-.1785
		C	A	.4286	.31419	.533	-.3453	1.2024
			B	.9524*	.31419	.011	.1785	1.7262
	Sidak	A	B	.5238	.31419	.273	-.2479	1.2955
			C	-.4286	.31419	.444	-1.2003	.3431
		B	A	-.5238	.31419	.273	-1.2955	.2479
			C	-.9524*	.31419	.011	-1.7241	-.1807
		C	A	.4286	.31419	.444	-.3431	1.2003
			B	.9524*	.31419	.011	.1807	1.7241

Figure 8: . *Post-Hoc questions 3,4. This figure is a continuation of the previous, hence rows and columns can be interpreted in the same way.*

B.
Mauchly's Test of Sphericity Table

Mauchly's Test of Sphericity[a]

Within Subjects Effect	Measure	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
						Greenhouse-Geisser	Huynh-Feldt	Lower-bound
SystemType	Quest1	.976	.464	2	.793	.976	1.000	.500
	Quest2	.999	.020	2	.990	.999	1.000	.500
	Quest3	.881	2.405	2	.300	.894	.976	.500
	Quest4	.901	1.971	2	.373	.910	.997	.500
	Quest5	.775	4.854	2	.088	.816	.879	.500
	Quest6	.898	2.044	2	.360	.907	.993	.500

Figure 9: . *Mauchly's Test of Sphericity.*

C.
Multivariate Test Table

Multivariate[a,b]

Within Subjects Effect		Value	F	Hypothesis df	Error df	Sig.
SystemType	Pillai's Trace	.594	2.537	12.000	72.000	.008
	Wilks' Lambda	.492	2.484[c]	12.000	70.000	.009
	Hotelling's Trace	.858	2.430	12.000	68.000	.011
	Roy's Largest Root	.522	3.130[d]	6.000	36.000	.014

Figure 10: . *Multivariate Test.*

D.
Usability Testing Survey Questions

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
This system hindered my productivity	○	○	○	○	○
I was comfortable multitasking with the authentication measures while completing the tasks	○	○	○	○	○
This system causes inconvenient pauses in my workflow	○	○	○	○	○
It was easy for me to adapt to the authentication measure in this system	○	○	○	○	○
I am willing to compromise on convenience for increased safety with this system	○	○	○	○	○
I feel safe with this authentication measure	○	○	○	○	○

Figure 11: