Continuous Age Detection using Keystroke Dynamics

Sander Brekke[1] and Patrick Bours[1][0000−0001−5562−6957]

Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Gjøvik, Norway
sanderob@stud.ntnu.no, patrick.bours@ntnu.no

Abstract. When enrolling users into computer systems and applications that are restricted to certain age groups, it is challenging to put trust in the user's provided age. This paper looks into the deployment of continuous analysis of Keystroke Dynamics data captured from the online activity of a user. Using this data the goal is to categorize the user's age into two possible categories: above and below the age of 18, a widespread legal age. We used a dataset captured from 70 adults and 46 children, containing over 780.000 keystrokes. The data is collected when the participants were chatting with a random other participant. Two different statistical methods, using timing features, are presented in both an authentication and an identification scenario. In the authentication scenario we reached an average accuracy of approximately 80% after on average of 180 keystrokes, while in the identification scenario we obtained a 75% True Positive Rate after approximately 20 keystrokes.

Keywords: Keystroke Dynamics · Behavioural Biometrics · Age Determination· Age Detection.

1 Introduction

Certain websites, applications, and computer systems have the need to filter users and their access due to age restrictions, for example because of explicit contents or to stop grooming. While the nature of the computer system or the application does not allow for true identification of age through official identifying documents, an explicit consent from the end user regarding their age may not be sufficient. In these cases, biometric characteristics may be applicable for creating an indication of an end user's age. One such biometric characteristic is keystroke dynamics (KD), where an individual's way of using a computer keyboard is measured [7]. Keystroke dynamics is a *behavioural characteristic* as it measures the way a person performs given actions, in the same way as gait analysis and speech recognition [1, 22].

There are several advantages of using KD to determine age. One of the greatest advantages is the low cost, as it uses the user's keyboard and does not require the introduction of new biometric capture devices. In addition, KD is non-intrusive, as the end user does not need to do anything different from their

normal use of the computer keyboard [13]. Some challenges may include privacy, as the data collected can be used to reconstruct the text that has been written, at the same time as there is no guarantee that the text entered is not sensitive.

As will be discussed in Section 2, there is research strongly suggesting that KD is a biometric characteristic that fulfils the requirements of an identifying characteristic. In addition, there is research proving that KD is a biometric characteristic that can be used for soft biometric evaluation. In other words, KD can be used to predict soft biometric characteristics in an individual. Following this, it is feasible to use KD as a biometric characteristic to predict the soft biometric age.

The remainder of this paper is organized as follows. In Section 2, the state of the art from the field is presented. Section 3 presents the dataset that is used, and Section 4 presents the methodology for the data analysis that is performed. Sections 5 and 6 present the results for the authentication scenario and the identification scenario. Finally, in Section 7 we draw conclusions based on the results in this paper.

2 State of Art

Keystroke dynamics as a behavioural biometric started in the 20th century, where telegraph operators could recognize each other through the way that messages were sent over telegraph lines [2]. This was utilized further during World War II, where military intelligence was able to identify the sender of a morse code and successfully distinguish foe from ally [8]. This was performed by analysing the rhythm, pace, and syncopation of the dots and dashes, a methodology referred to as "The fist of the sender" [16].

There are two main parallel research directions within keystroke dynamics as a biometric characteristic. Firstly, there is the research on KD as an identifying biometric characteristic, often used in the context of authentication [1, 3, 15]. Secondly, there is ongoing research regarding KD as a soft biometric characteristic [14, 19, 20]. Differentiating from the former, employment of the latter does not allow for identification of a person. It can, although, be used to predict properties of individuals and other soft biometric characteristics, like age and gender [6, 21]. Given their natures, soft biometric characteristics can also be used to assist or confirm other biometric characteristics [18].

In the context of age determination using KD, a soft characteristic approach is the most relevant, as age is a soft biometric characteristic. In the scope of this project, the use of continuous detection is chosen instead of static detection. The difference between them lays in the frequency of the analysis and the comparison between the probe and the reference. In continuous detection, the comparison happens at every single keystroke [4, 13], where in static detection, the comparison happens after the user has finished their typing [5, 9]. Continuous detection is chosen due to the nature of the intended application of the results and the identified research gaps on the subject.

As mentioned, keystrokes dynamics is proven to be usable in authenticating individuals. From this, there has also been research attempting to predict and identify age as a soft biometrics of an individual through KD. This research is focused on the use of machine learning (ML). The use of ML is not regarded as feasible in the context of this project, due to the nature of continuous detection, where ML works slower than an individual is typing.

The most relevant research into the use of KD in continuous statistical detection of age is from the master thesis of Tverrå [21], from the Norwegian University of Science and Technology (NTNU), where he looks into the detection of age and gender. The research was performed on a data-set consisting of 56 participants and an average of 1750 keystrokes per participant. In terms of soft biometrics, two age groups (< 30 and > 30) and two gender groups were employed. The participants had a spread of 44 (< 30) versus 12 (> 30) between the age groups, and 31 (male) versus 25 (female) for gender. Tverrå used scaled Manhattan distance (SMD) as the statistical metric, in addition to machine learning. The results yielded were that SMD worked faster than machine learning models. The results yielded by SMD for age prediction was an accuracy of 87% after an average of 1498 keystrokes.

This research aims to become an extension of the current research present by improving the statistical methods employed in existing research, for instance in order to 1) reduce the number of keystrokes needed or 2) increase the accuracy of the methods employed.

3 Dataset

The dataset used in this research is gathered by Aiba, a previous research group at the Norwegian University of Science and Technology[1], and is based on online chat conversations. The chat data is collected using a web-based "Become Acquainted!" application that was built to collect keystroke dynamics data on both laptops and mobile devices. When a user would start using the app, he or she would be connected to a random other person, and they could then chat as long as they wanted on any topic they wanted. When one of them disconnected from the app, both could then again find a new chat partner.

The application was first tested between children from two different schools, where a student from one school was always connected to a student from another school to avoid that they already knew each other. On two different occasions, the app was tested for about 60–90 minutes, to collect data on how children type when they are chatting. Most children used their mobile phone or a tablet for chatting. The second test of the application was between adults. Adults voluntarily participated on a number of fixed timeslots to chat with unknown others. Because this experiment was done in Norway, using also international participants, two different chats were set up. One for chatting in Norwegian and the other for English. Participants could sign up for one or both languages, and

[1] The research group Aiba has later become an independent commercial company (`https://aiba.ai/`, retrieved 24.10.2024)

would for each of the languages only connected to a participant who also opted for that language. The experiment was run 2 days a week for 3 weeks, resulting in a total of 6 days.

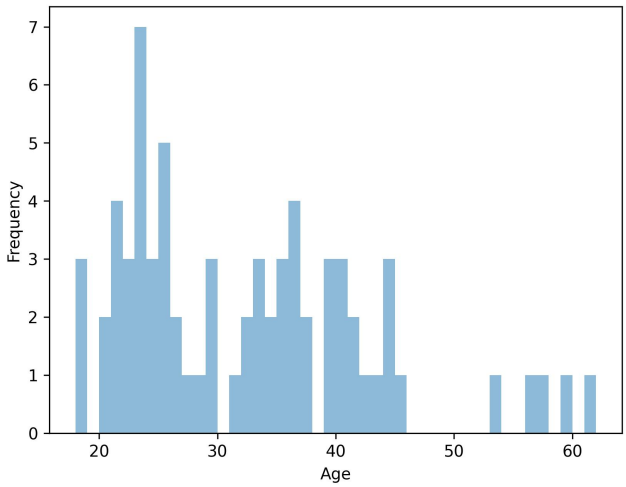


Fig. 1: Age and frequency of participants in adult population

The dataset consists of a total of 116 users, whereof 70 are adults and 46 are children. Of the adults, there are 47 female participants and 23 male participants. Of the children, there are 23 male participants and 23 female participants. In total, there are 70 female and 46 male participants. For the child participants, all were attending 9th grade of Norwegian public school, thus around the age of 14. The age of the adult participants was in the range of [18, 62], and the spread can be seen as a histogram in Figure 1. Each of the users have an average of $\approx$ 6 754.5 keystrokes, from a total of 813 989 keystrokes. These are spread over 45 651 messages within a total of 2 306 conversations. A conversation is a session where either a child or adult participant exchanges messages with a different child or adult participant.

The recorded data set consists of data lines, where each line represents one keystroke. In the data lines, the key value represents the value of the key pressed, the key duration represents the duration from the key is pressed until it is unpressed, and the key latency represents the time from the key is unpressed until the next key is pressed. In cases where a combination of keys is pressed, for example $shift + a$, an upper letter A becomes the respective key value. If the consecutive letter is pressed before the former letter is released, the key latency will become negative. During the creation of the dataset, a keylogger was used. In some cases, some values have not been recorded correctly. These values have

been removed, according to the documentation of the acquired dataset. This resulted in a slightly lower amount of data than the truly typed data.

The dataset is split up into one training-set consisting of children, one training-set consisting of adults, one testing-set containing children and one testing-set containing adults. The split is performed on users, where a user is in either the training-set or the testing-set of a population. This way, a probe will never be compared to a reference trained on that exact user.

3.1 Data Extraction

When working on recorded keystrokes where the latency and duration is described for each key pressed, there are six different features that can be extracted for each digraph. In this case, a digraph is a special case of n-graphs, where n executive key presses are combined into a metric [17]. The six features from the digraphs are: key 1 duration, key 2 duration, total duration, up-down latency, down-down latency, and up-up latency. The calculation for each feature can be seen in Table 1, and each feature is illustrated in Figure 2.

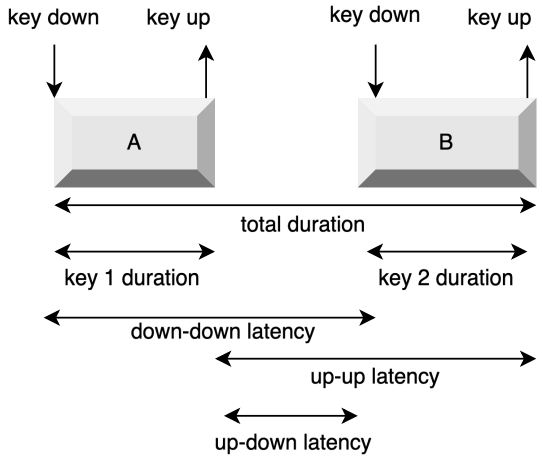


Fig. 2: Illustration of 6 different keystroke features.

When creating digraphs from a data set consisting of data lines of keystrokes, the keystrokes need to be combined in pairs. Although, because the dataset is retrieved from a chatting application, the keystrokes are retrieved from several users writing several messages in several conversations. Thus, the pairs need to be created for each message, for each conversation, and for each user.

When extracting data and creating digraphs from a set of keystrokes, it is possible to utilize different categorizations of the two button values. That is, it is possible to use key value 1 and 2 as the true value of the key pressed,

Tab. 1: Calculations for six different keystroke features.

Feature	**Calculation**
Total duration	Key 1 duration + key 2 duration + key 1 latency
Key 1 duration	Duration for key 1
Key 2 duration	Duration for key 2
Down-down latency	Key 1 duration + key 1 latency
Up-up latency	Key 1 latency + Key 2 duration
Up-down latency	Key 1 latency

giving $42^2 = 1\ 764$ possible digraphs[2]. Other options are to use the side of the keyboard as a value (giving four possible digraphs: *LR*, *LL*, *RL*, *RR*), to use the row of the keyboard as a value (giving $5^2 = 25$ possible digraphs), and to use a combination of side and row (giving $10^2 = 100$ possible digraphs[3]). An argument for choosing any of the three latter options is that it is difficult to find enough data to create templates for 1 764 digraphs. In addition, as will be presented in Section 4.1, efforts have been made to look into the statistical significance of the difference between the two groups' templates. When creating templates based on the keyboard side, keyboard row, or a combination, the statistical analysis of the templates shows close to no difference between the two populations' templates. On the contrary, when creating templates with the key values, the statistical difference between the two populations' templates is greater. Due to this, the use of key values was chosen.

As mentioned, it is known to be hard to find data sets that are large enough to create templates on all 1 764 digraphs when actual key values are employed. However, the digraphs that are rare in the references are also rare in the probes that are compared to the references. This means that, although the table of digraphs present in the reference is sparse, the probes statistically have the same sparseness, making the frequency of an inability to compare due to missing reference data lower than what the reference data suggests. In the comparison of probes against references in this project, the frequency of a probe digraph that did not have a reference digraph was found to be 3.0%.

There are several things that make an individual type differently than they normally do. These things may include: pause for thought, distractions, typing errors, and proofreading. For these cases, the data is captured, but is not wanted in the data set. For this purpose, filtering has been performed. In the case of this project, the filtering has been performed on digraphs, where a digraph is kept if the total duration is less than 2000 ms.

[2] 29 letters in the Norwegian alphabet + 10 numbers + whitespace + comma + period = 42

[3] 5 $rows \cdot 2$ $sides = 10$ $possibilities$

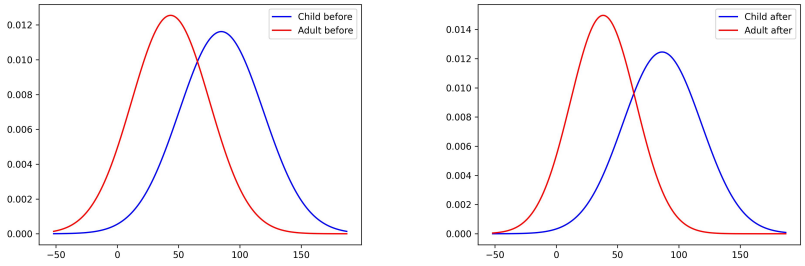
4 Methodology

When creating templates of a dataset for keystroke dynamics, it is possible to choose one of two main pathways. Firstly, it is possible to do so by averaging, where all the template instances for each digraph in the training set is averaged into one instance. Secondly, it is possible to keep all the template instances for each digraph, in order to have a wider set of template instances contained in the references when comparing a probe to the reference. There are no known advantages and disadvantages to either, but in this project, the latter is chosen in order to perform dynamic outlier removal in the template data, as will be further discussed in Section 4.1.

4.1 Outlier Removal

When all templates had been created, a filtering was performed. In order to remove the instances that may not represent the population correctly, all digraph references with less than 10 template instances (i.e. $NumberOfEntries < 10$) were removed.

To statistically distinguish child probes from adult probes, a significant difference between the two populations must be present within their references. When templates for each digraph reference had been created, a two-sided t-test was performed for each digraph. A two-sided t-test is a form of hypothesis testing, where the null hypothesis (h_0) assumes that there is no statistical difference between the populations [11]. The higher the significance of the statistical difference, the lower the p-value. As such, the p-value is the probability of a type one error, where h_0 is wrongfully rejected. In the scope of this project, there is no statistical hypothesis that is being tested, but a t-test and its results can imply whether it is possible to statistically differentiate between the two populations, thus also whether it is possible to statistically determine the age group of a probe.



(a) Before outlier removal.

(b) After outlier removal.

Fig. 3: The digraph with lowest p-value before and after outlier removal.

In Figure 3a, the digraph feature from the training data with the most statistically significant difference between the child and the adult reference for the digraph can be seen. These normal distributions are created from the mean and standard deviations of all template instances within the respective reference feature for both populations. As noticeable, the difference between the samples may not seem high, but the p-value in this case is $3.8156 \cdot 10^{-8}$. This number represents the probability that the two samples are from the same population. In other words, statistically, the difference is considered statistically significant, but the visualization shows that there can be made improvements.

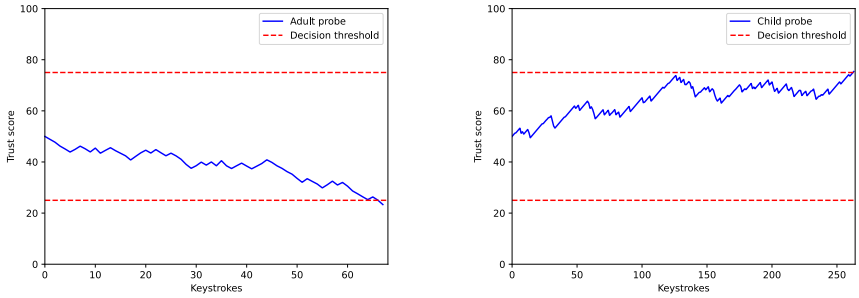
In order to increase the difference between the references created from the adult and the child population templates, a removal of outliers was performed. For each population, templates were created, consisting of all instances of each digraph. From the templates, the standard deviation (σ) and the mean (μ) of each feature is found. Then, iterating over the templates inside the reference of the given digraph for the given population; if any of the values are further than $n \cdot \sigma$ from μ, the template is removed from the reference. Several values of n were tested, and the most optimal is $n = 1.5$. The results of the outlier removal can be seen in Figure 3, where Figure 3a presents the digraph feature with the lowest p-value before the outlier removal, while Figure 3b presents the same digraph feature after the outlier removal. As can be seen, the inter-difference between the populations increases from the outlier removal, and the p-value becomes $1.1204 \cdot 10^{-12}$ after the outlier removal. An improvement factor of 34 059 is achieved[4]. This makes the statistical difference considerably more significant, making it more statistically probable that the references are from two different populations. This makes it easier to statistically determine the correct age group of a probe from comparison with the references.

4.2 Trust Model

In order to continuously determine the population in which a probe belongs, a decision-making model is needed. This is because, different from static keystroke dynamics, it is hard to know when and after how many keystrokes to make a decision, as it is hard to know how many more keystrokes will be entered by the probe user. For this purpose, a trust model developed by Bours was chosen [4, 13]. The basis of the trust model is to continuously calculate a current level of trust, representing the level of trust in a potential decision, based on the deviation and conformity between the probe and the populations' references. The model was developed with continuous authentication in mind, where the initial trust score is 100%, representing the level of trust in which that the current user is the genuine user. A threshold is set, where the trust is too low to continue, and the current user is locked out of the system. In the scope of this project, the starting value will be 50, and there will be set thresholds on where to decide on the population of the probe user. An example of a trust model employment can be seen in Figure 4. Do note the difference in number of keystrokes, where

[4] 34 059 times smaller, as: $\frac{3.8156 \cdot 10^{-8}}{34\ 059} = 1.1204 \cdot 10^{-12}$

the trust model has its strength, allowing for a dynamic number of keystrokes needed for a decision.



(a) Trust model on an adult probe.

(b) Trust model on a child probe.

Fig. 4: Visualization of trust model with unclaimed age detection (identification). Upper threshold equals child and lower threshold equals adult.

As keystrokes are added, the level of trust may increase or decrease based on an algorithm for changing the current trust [4]. Derived from the application of the model by Mondal and Bours [13], the change of trust is calculated with a modified Sigmoid function, where four values are defined: A is the threshold of when a score change is positive or negative, B is the width of the Sigmoid function, C is the upper value for a positive score change, and D is the upper limit for a negative score change. In addition, one threshold value for when to make decisions and one starting value is defined. Different values will be used for different testing methods, distance metrics, and scenarios, as will be discussed closer in Section 5.

5 Authentication

In the domain of biometrics, there are mainly two different use cases; authentication and identification. Authentication involves the verification of something that is claimed, for example an identity, a gender, an age, or similar. Identification, on the other hand, is not verifying a claim, but identifying the biometric property through estimation. This section will look into a proposed method for continuous authentication of a claimed age, while Section 6 will look into a proposed method for continuous identification of the end user's age.

This method consists, in broad strokes, of measuring the distance from the probe digraphs to the reference digraphs of the claimed age group. This method starts by obtaining the six different features from the typed digraphs, as they are being typed. When the features are collected, a point in a six dimensional space is obtained, where each of the features represent one dimension. Likewise,

the features from the digraph is looked upon as a point, and there will be one additional point in the space for each instance of the digraph in the reference. Further, a distance is calculated from the probe point to each of the template points of the given digraph.

For this purpose, several distance metrics were considered. In a comparison review by Killourhy [10], scaled Manhattan distance and nearest neighbour were reviewed to be good distance metrics for keystroke dynamics. From Mondal & Bours [13], a combination of scaled Euclidean distance and correlation distance was utilized. In the master's theses of Tverrå [21] and Moe [12], scaled Manhattan distance, scaled Euclidean distance and nearest neighbour was used. In all mentioned projects, all mentioned distance metrics had yielded a positive result. In this project, the use of scaled Euclidean distance is chosen. The formula for calculating the scaled Euclidean distance between two digraphs with 6 features each can be seen in Equation 1. In this equation, t_n is feature number n of the template digraph, σ_n is the standard deviation of feature n over the reference digraph, and p_n is feature n of the probe digraph. This gives one distance per template in the digraph reference. These are combined into one distance by finding the mean of the distances.

$$SED = \sqrt{\sum_{n=1}^{6} \frac{(t_n - p_n)^2}{\sigma_n}} \tag{1}$$

The results yielded ranged between 63.6% and 93.3%, and can be seen in Table 2. The reference age group is the same as the claimed age group, given the method of testing. This gives an average false match rate (FMR) of 9.1% and a false non match rate (FNMR) of 33.3%.

Tab. 2: Results for claimed age detection, using mean scaled Euclidean distance.

Probe age	Reference age	Probes tested	Correct decisions	Incorrect decisions	Avg. # of keystrokes	Accuracy
child	child	44	28	16	311	63.6%
adult	adult	22	16	6	98	72.7%
child	adult	44	41	3	204	93.2%
adult	child	22	19	3	133	86.4%
TOTAL		**66**	**104**	**28**	**210**	**78.8%**

6 Identification

This approach is based on calculating digraph scores based on probabilities derived from the features of the reference templates. This is done by calculating one probability per feature per reference, giving $2\ references \cdot 6\ features = 12\ probabilitities$. These are subsequently combined into one digraph score, based on the ratio between them.

Each feature score is created from the ratio between the probability that the probe feature belongs to the child population and the probability that the probe feature belongs to the adult population. In order to do this, two normal distributions are created; one from the features of the template instances in the adult reference and one from the features of the template instances in the child reference. The probabilities represent the probability that the probe value of the feature belongs to each of the populations. However, in a normal distribution, the point probability is always zero. This is because the probability is equal to the integral of the interval, and the interval has a width of zero. In order to solve this, one half is added before and after the point, making the width equal to one. Thus, the interval for the probability of value n is $[n - 0.5,\ n + 0.5]$. The probability is then found through calculating two *Z-scores*, as seen in Equation 2 and referencing the probability for each in a *Z-score table*. The exact probability for the interval is found by subtracting the larger probability with the smaller, as seen in Equation 3. The *Z-score table* is equal for all normal distributions.

$$SC_{Z1}(n) = \frac{n + 0.5 - \mu_{feature}}{\sigma_{feature}},\ SC_{Z2}(n) = \frac{n - 0.5 - \mu_{feature}}{\sigma_{feature}} \tag{2}$$

$$P([n - 0.5, n + 0.5]) = MAX(P(SC_{Z1}) - P(SC_{Z2}),\ P(SC_{Z2}) - P(SC_{Z1})) \tag{3}$$

From the two probabilities per feature, one feature score is calculated by dividing the larger with the smaller, as seen in Equation 4. When this is done, the feature scores represent the factor between the probabilities; if $P_{adult,\ feature1}$ is four times greater than $P_{child,\ feature1}$, the initial feature score is 4. The score changes between positive and negative for each feature score depending on which probability that is greater, in order to make the score go towards either decision threshold (i.e. up or down). For instance, when the starting trust score is 50, the adult decision threshold is 25, and the child decision threshold is 75, the digraph score (i.e. the trust score delta) will be negative if the adult probability is larger and positive if the child probability is larger. In the example where the feature score is 4, it would change to -4, as the adult probability is larger.

$$FS = MAX(\frac{P(child|feature_{probe})}{P(adult|feature_{probe})},\ \frac{P(adult|feature_{probe})}{P(child|feature_{probe})}) \tag{4}$$

When one score is created per feature of the probe, a digraph score is derived from the mean of the feature scores. This method does not utilize the Sigmoid function, but do employ the trust model from Section 4.2. Thus, $\Delta\ trust\ score = digraph\ score$.

The application of this method yielded a 77.3% true positive identification rate (TPIR) for child probes and a 72.7% TPIR for adult probes, as seen in Table 3.

Tab. 3: Results for unclaimed age detection using probability.

Probe age group	Probes tested	Correct biometric identification decisions	Incorrect biometric identification decisions	True positive identification rate	Avg. # of keystrokes
child	44	34	10	77.3%	21
adult	22	16	6	72.7%	17

7 Conclusion

When developing, implementing, and testing different methods for proposal, there are several things that may affect the results and the reproducibility. For instance, the contents of the training set have an impact; in this case, the children in the data set all were around the same age, missing the representation of children of other ages. The adult participants, however, were mostly spread between the ages of 20 to 45. The split between the training set and test set may also affect the results. During the data-set collection, the children used their mobile phones and tablets, while the adults mostly used physical keyboards. This makes it hard to know whether the difference in typing between the population was due to difference in age or difference in devices used.

When performing claimed age prediction, the results varied based on the probe age group and the reference age group. This makes the method perform differently based on the application of the method. For instance, if the application of the method is based on keeping adults away from sites designed for children, the false match rate (FMR) is 13.6% and the false non-match rate is (FNMR) 36.3%. If the application is based on keeping children away from adult sites, however, the FMR is 6.7% and the FNMR is 27.3%.

When considering other studies on continuous age detection, few comparisons exist. One such study achieved an 87% accuracy [21]. However, it's worth noting that this research focused solely on identification, not authorization. In contrast, our research achieved comparable results using only 2.6% of the keystrokes, with 21 compared to Tverrå's 825.

This report presents two methods for continuously detecting the age group of an end-user through keystroke dynamics; one method for claimed contexts and one method for unclaimed contexts. The methods yield results that prove the statistical significance of the ability to continuously predict the soft biometrics of age in end-users through keystroke dynamics. Thus, this report contributes to the current state of the art by providing results that proves the feasibility of the approach. However, given its preliminary nature, further research is needed in order to: i) pinpoint significant and influential features, ii) refine the methods to mitigate both false non-match rate and false match rate, and iii) explore strategies to enhance the statistical significance of inter-differences between the populations' references.

References

1. Alzubaidi, A., Kalita, J.: Authentication of Smartphone Users Using Behavioral Biometrics. IEEE Communications Surveys & Tutorials **18**(3), 1998–2026 (2016). https://doi.org/10.1109/COMST.2016.2537748
2. Barghouthi, H.: Keystroke Dynamics: How typing characteristics differ from one application to another. Master's thesis, Gjøvik University College, Norway (2009), `http://hdl.handle.net/11250/143781`
3. Bhatt, S., Santhanam, T.: Keystroke dynamics for biometric authentication — A survey. In: 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering. pp. 17–23. IEEE (Feb 2013). https://doi.org/10.1109/ICPRIME.2013.6496441
4. Bours, P.: Continuous keystroke dynamics: A different perspective towards biometric evaluation. Information Security Technical Report **17**(1-2), 36–43 (Feb 2012). https://doi.org/10.1016/j.istr.2012.02.001
5. Buriro, A., Akhtar, Z., Crispo, B., Del Frari, F.: Age, Gender and Operating-Hand Estimation on Smart Mobile Devices. In: 2016 International Conference of the Biometrics Special Interest Group (BIOSIG). pp. 1–5. IEEE (Sep 2016). https://doi.org/10.1109/BIOSIG.2016.7736910
6. Dantcheva, A., Velardo, C., D'Angelo, A., Dugelay, J.L.: Bag of soft biometrics for person identification: New trends and challenges. Multimedia Tools and Applications **51**(2), 739–777 (Jan 2011). https://doi.org/10.1007/s11042-010-0635-7
7. Gaines, R.S., Lisowski, W., Press, S.J., Shapiro, N.: Authentication by Keystroke Timing: Some Preliminary Results. Tech. rep., RAND institute (1980)
8. Haring, K.: Ham radio's technical culture. Inside technology, MIT press, Cambridge (Mass.) (2007)
9. Hoeijmakers, A., Licitra, G., Meijer, K., Lam, K.H., Molenaar, P., Strijbis, E., Killestein, J.: Disease severity classification using passively collected smartphone-based keystroke dynamics within multiple sclerosis. Scientific Reports **13**(1), 1–12 (Feb 2023). https://doi.org/10.1038/s41598-023-28990-6
10. Killourhy, K.S., Maxion, R.A.: Comparing anomaly-detection algorithms for keystroke dynamics. In: 2009 IEEE/IFIP International Conference on Dependable Systems & Networks. pp. 125–134. IEEE (Jun 2009). https://doi.org/10.1109/DSN.2009.5270346
11. Mishra, P., Singh, U., Pandey, C., Mishra, P., Pandey, G.: Application of student's t-test, analysis of variance, and covariance. Annals of Cardiac Anaesthesia **22**(4), 407–411 (2019). https://doi.org/10.4103/aca.aca_94_19
12. Moe, T.: I still know who you are! Soft Biometric Keystroke Dynamics performance with distorted timing data. Master's thesis, Norwegian University of Science and Technology, Gjøvik, Norway (2021), `https://hdl.handle.net/11250/2781218`
13. Mondal, S., Bours, P.: Continuous Authentication in a real world settings. In: 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR). pp. 1–6. IEEE (Jan 2015). https://doi.org/10.1109/ICAPR.2015.7050673
14. Olasupo, O., Adesina, A.O.: Predicting Age Group and Gender of Smartphone Users Using Keystroke Biometrics. Malaysian Journal of Science and Advanced Technology pp. 124–128 (Oct 2021). https://doi.org/10.56532/mjsat.v1i4.24
15. Pahuja, G., Nagabhushan, T.N.: Biometric authentication & identification through behavioral biometrics: A survey. In: 2015 International Conference on Cognitive Computing and Information Processing(CCIP). pp. 1–7. IEEE (Mar 2015). https://doi.org/10.1109/CCIP.2015.7100681

16. Saini, B.S., Kaur, N., Bhatia, K.S.: Authenticating mobile phone user using keystroke dynamics. Int. J. Comput. Sci. Eng **6**(12), 372–377 (2018)
17. Sim, T., Janakiraman, R.: Are digraphs good for free-text keystroke dynamics? In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE (Jun 2007). https://doi.org/10.1109/cvpr.2007.383393
18. Syed Idrus, S.Z., Cherrier, E., Rosenberger, C., Mondal, S., Bours, P.: Keystroke dynamics performance enhancement with soft biometrics. In: IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015). pp. 1–7 (2015). https://doi.org/10.1109/ISBA.2015.7126345
19. Tsimperidis, I., Rostami, S., Wilson, K., Katos, V.: User Attribution Through Keystroke Dynamics-Based Author Age Estimation. In: Selected Papers from the 12th International Networking Conference, vol. 180, pp. 47–61. Springer International Publishing (2021). https://doi.org/10.1007/978-3-030-64758-2_4
20. Tsimperidis, I., Yoo, P.D., Taha, K., Mylonas, A., Katos, V.: R ^{2} BN: An Adaptive Model for Keystroke-Dynamics-Based Educational Level Classification. IEEE Transactions on Cybernetics **50**(2), 525–535 (Feb 2020). https://doi.org/10.1109/TCYB.2018.2869658
21. Tverrå, O.D.: Continuous Determination of Age and Gender. Master's thesis, Norwegian University of Science and Technology, Gjøvik, Norway (2023), `https://hdl.handle.net/11250/3092833`
22. Zhu, J., Hu, H., Hu, S., Wu, P., Zhang, J.Y.: Mobile Behaviometrics: Models and applications. In: 2013 IEEE/CIC International Conference on Communications in China (ICCC). pp. 117–123. IEEE (Aug 2013). https://doi.org/10.1109/ICCChina.2013.6671100