A Lightweight Identification System with Keystroke Dynamics and Fingerprint Verification

February 10, 2025

Simone Giovagnoni 1932180 Davide Paossi 1950062 Giuseppe Carnicella 1950329 Federico Cattenari 1950982

https://github.com/giuseppecarnicella/Biometric-System-Project

Abstract

This report presents the development and implementation of a lightweight multimodal identification system. This system uses two biometrics: typing keystrokes and fingerprints. The keystroke metric is used to identify the user and the fingerprint is used for verification.

1. Introduction

In most places, it is becoming increasingly crucial to be able to restrict access. But this need for access control must interface with the economic burden required by such systems. In this project, we develop a lightweight model to identify users with negligible computational cost.

2. Related works

In this section are described related works that have been useful for inspiration and the datasets used for the models.

2.1. CMU Keystroke dynamics dataset(Killourhy & Maxion, 2009)

The CMU Keystroke Dynamics Dataset is a publicly available dataset collected by Carnegie Mellon University to study user authentication based on typing patterns. It contains keystroke timing data from multiple users typing a fixed password multiple times. The dataset records key press and release timestamps, enabling the analysis of dwell time (time a key is held down) and flight time (time between consecutive key presses).

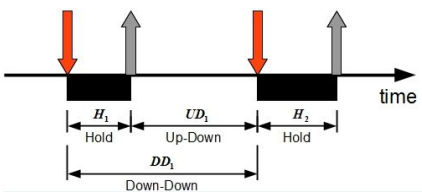

Figure 1. Three time periods used as features in keystroke dataset

2.2. TypeNet(Acien et al., 2022)

TypeNet is a deep learning-based keystroke biometric system designed for large-scale authentication in free-text typing scenarios. It utilizes Long Short-Term Memory (LSTM) networks to analyze users' typing patterns, including key press duration and transition timings. The system has been tested on both physical keyboards and touchscreen devices, achieving state-of-the-art performance.

2.3. Keystroke verification method(rakshithca, 2019)

This code has been our main reference for the development of the keystroke based model. This system uses three different techniques to analyze the keystrokes:

- **Manhattan Distance based**
  Uses the absolute distance between vectors to compare input data with reference data.

- **Euclidean Distance based**
  Uses the traditional Euclidean distance to measure how much the typing pattern differs from the average vector of the recorded subject.

- **KNN based**
  Uses a KNN model to select the most similar sample for the verification task.

2.4. Sokoto Fingerprint dataset(Shehu et al., 2018)

The Sokoto Coventry Fingerprint Dataset (SOCOFing) is a publicly available fingerprint dataset designed for biometric research.
It contains 6,000 fingerprint images from 600 African subjects, with each subject contributing ten fingerprint samples. The data set includes real and synthetically altered fingerprints, featuring modifications such as obliterations, central rotation, and z-cut transformations to simulate spoofing and fingerprint alteration attempts.

Figure 2. Five left-hand fingerprints belonging to the same subject

Figure 3. Images from Figure 2 after being altered into z-cut, obliteration and central rotation

3. Method

In this section, the two methodologies that are part of the final model are to be discussed in depth.

3.1. Keystroke identification

The first stage is an open-set identification. The user types a specific word, and their typing dynamics are analyzed using a K-Nearest Neighbors (KNN) model. The role of this model is to determine whether the user belongs to the set of enrolled subjects or should be classified as an intruder and subsequently rejected.
The identification is performed by comparing the new typing sample with previously enrolled users' samples based on features such as Hold Time (H), Down-Down Time (DD), and Up-Down Time (UD). To enhance performance and reduce the dimensionality of the data, we applied Principal Component Analysis (PCA), retaining the three most significant components.
When a new user types the required word, the KNN model calculates the probability that the typing pattern belongs to a known subject. If this probability exceeds a defined threshold, the user is positively identified as one of the enrolled subjects. Otherwise, they are classified as "Unknown" and rejected.
If the user is positively identified, his assumed identity is passed to the second model.

3.2. Fingerprint verification

In this second step, the identity of the subject proposed by the first model is verified.
The user must provide the system with his or her fingerprint, which will be compared with the fingerprint of the claimed identity. The two fingerprints are compared by matching the SIFT descriptors to determine their equality.
If the two fingerprints are found to be sufficiently similar, the identity is confirmed, otherwise, it is rejected.
The model has been tested on the altered versions of the fingerprints, in order to focus particularly on the robustness of the model. The altered versions are divided into three main categories: hard, medium and easy; each reflecting the alteration level of the images.

4. Results

This section discusses the results in tests of the two models involved by analyzing the metrics of Genuine Acceptance Rate (GAR), False Acceptance Rate (FAR), Genuine Rejection Rate (GAR), False Rejection Rate (FRR) and the ROC curve.

4.1. Keystroke identification results

In the first stage of identification, the model makes a distinction between an enrolled subject and a subject not registered in the system.
So, the identification system checks whether the keystroke typing of the user who wants to log in matches the keystroke typing of an enrolled subject. The model, then, rejects only if the user is recognized as an outsider.

Table 1. Keystroke Test results

Metrics	Results
Genuine Acceptance Rate	78.45%
False Rejection Rate	21.54%
Genuine Rejection Rate	42.00%
False Acceptance Rate	58.00%

4.2. Fingerprint verification results

In the second stage, the verification of the proposed identity from the first model takes place. So, the subject is accepted if the fingerprint matches the one of the claimed identity present in the system.

Table 2. Fingerprint Test results

Metrics	Hard	Medium	Easy
Genuine Acceptance Rate	76.92%	87.78%	94.21%
False Rejection Rate	23.07%	12.21%	5.79%
Genuine Rejection Rate	98.98%	98.91%	97.10%
False Acceptance Rate	1.01%	1.08%	2.89%

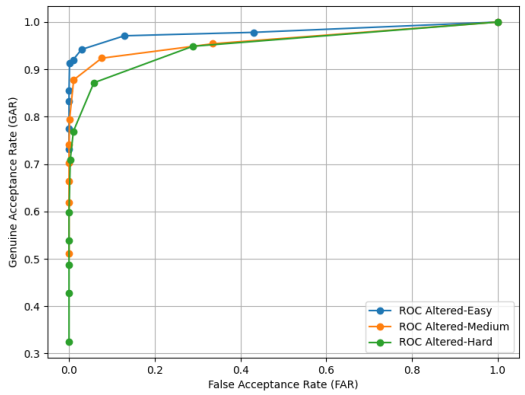

Figure 4. ROC curve of the fingerprint model

4.3. Multimodal system results

In table 3 are presented the results of the whole system.

Table 3. Multimodal system Test results

Metrics	Hard	Medium	Easy
Genuine Acceptance Rate	56.12%	58.03%	68.09%
False Rejection Rate	43.88%	41.97%	32.91%
Genuine Rejection Rate	100%	100%	100%
False Acceptance Rate	0%	0%	0%

5. Discussion and conclusions

The obtained results demonstrate the feasibility of a lightweight yet effective biometric authentication system.
For keystroke identification, the system achieved a Genuine Acceptance Rate (GAR) of 78.45%, which is a reasonable performance for an open-set recognition task. However, the False Acceptance Rate (FAR) of 58.00% indicates that some intruders were incorrectly accepted.
In the fingerprint verification stage, the model performed significantly better, with a GAR ranging from 76.92% (hard alterations) to 94.21% (easy alterations). The False Acceptance Rate remained low across all test conditions, highlighting the robustness of fingerprint matching. The results confirm that fingerprint verification is a reliable second step to validate the user's identity, mitigating errors from the keystroke identification stage.
The final model results have reached the outstanding score of 100% for the Genuine Rejection Rate (GRR), allowing to avoid intruders to access the system keeping the False Acceptance rate at 0%. Despite the excellent results of GRR and FAR, the performances for Genuine Acceptance Rate (GAR) and False Rejection Rate (FRR) are suboptimal, with the range on GAR fluctuating between 56% and 68%, and for FRR between 32% and 44%.
In this type of system, it is crucial to make sure that the number of intruders accessing the system is as small as possible, while as far as genuine subjects are concerned, however tedious, it is possible to retry access a second time.
This study demonstrates that a multimodal biometric system combining keystroke dynamics and fingerprint verification is a viable approach for lightweight user authentication.

6. Future works

Future works should focus on refining the keystroke identification model to reduce the False Acceptance Rate and exploring deep learning techniques for improved biometric recognition. Further evaluation on larger and more diverse datasets would also strengthen the system's generalizability and reliability in real-world scenarios.
Regarding the fingerprint verification phase, it might be interesting not to limit the allowed samples to the right thumb, but to expand the choice of the fingerprint to be registered to all ten fingers, to make fingerprint forgery more complex.
In addition, it would be possible and useful to add another layer of security, represented by the use of a physical token. That would not reduce the lightweight advantages of the developed model, providing high protection from intruders.

References

Acien, A., Morales, A., Monaco, J. V., Vera-Rodriguez, R., and Fierrez, J. Typenet: Deep learning keystroke biometrics. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(1):57–70, 2022. doi: 10.1109/TBIOM.2021.3112540.

Killourhy, K. S. and Maxion, R. A. Comparing anomaly-detection algorithms for keystroke dynamics. In *2009 IEEE/IFIP International Conference on Dependable Systems Networks*, pp. 125–134, 2009. doi: 10.1109/DSN.2009.5270346.

rakshithca. Keystroke-dynamics, 2019. URL https://github.com/rakshithca/KeyStroke-Dynamics.

Shehu, Y. I., Ruiz-Garcia, A., Palade, V., and James, A. Sokoto coventry fingerprint dataset, 2018. URL https://arxiv.org/abs/1807.10609.