Artificial Intelligence Review (2024) 57:272
https://doi.org/10.1007/s10462-024-10893-1

Deep learning-based authentication for insider threat detection in critical infrastructure

Arnoldas Budžys[1] · Olga Kurasova[1] · Viktor Medvedev[1]

Accepted: 25 July 2024
© The Author(s) 2024

Abstract
In today's cyber environment, threats such as data breaches, cyberattacks, and unauthorized access threaten national security, critical infrastructure, and financial stability. This research addresses the challenging task of protecting critical infrastructure from insider threats because of the high level of trust and access these individuals typically receive. Insiders may obtain a system administrator's password through close observation or by deploying software to gather the information. To solve this issue, an innovative artificial intelligence-based methodology is proposed to identify a user by their password's keystroke dynamics. This paper also introduces a new Gabor Filter Matrix Transformation method to transform numerical values into images by revealing the behavioral pattern of password typing. A siamese neural network (SNN) with the branches of convolutional neural networks is utilized for image comparison, aiming to detect unauthorized attempts to access critical infrastructure systems. The network analyzes the unique features of a user's password timestamps transformed into images and compares them with previously submitted user passwords. The obtained results indicate that transforming the numerical values of keystroke dynamics into images and training an SNN leads to a lower equal error rate (EER) and higher user authentication accuracy than those previously reported in other studies. The methodology is validated on publicly available keystroke dynamics collections, the CMU and GREYC-NISLAB datasets, which collectively comprise over 30,000 password samples. It achieves the lowest EER value of 0.04545 compared to state-of-the-art methods for transforming non-image data into images. The paper concludes with a discussion of findings and potential future directions.

Keywords Critical infrastructure · Deep learning · Keystroke dynamics · Cybersecurity · Behavioral biometrics · Siamese neural network

1 Introduction

Today's cyber environment offers cybercriminals and intruders multiple opportunities to attack a country's networks and critical infrastructure, demand money for ransom data, facilitate large-scale fraud schemes, and threaten national security. It is important

Olga Kurasova and Viktor Medvedev have contributed equally to this work.

Extended author information available on the last page of the article

Published online: 29 August 2024

for critical infrastructure and businesses to protect their respective facilities from digital threats. The consequences of these threats can be serious, resulting in significant financial losses, reputational damage, and loss of customer confidence. As cyberattacks and cyberfraud continue to impact our daily lives, the FBI's internet crime complaint center (IC3) plays an essential role in dealing with cyberthreats. The IC3 serves as a public resource for submitting reports of cyberattacks and incidents, allowing them to collect data, identify trends, and address threats at hand. In 2022, IC3 received 800,944 complaints, a 5% decrease from 2021. However, the potential total loss increased from $6.9 billion in 2021 to more than $10.2 billion in 2022 (Federal Bureau of Investigation 2023).

Stolen credentials account for 80% of the financial losses attributed to cybercrime (Verizon 2022). Phishing is a form of cyberattack that uses fraudulent emails, text messages, and phone calls, masquerading as messages from a trusted institution, to steal personal, financial, or credential information from an unwary recipient (Basit et al. 2021; Jain and Gupta 2022). Recently, critical infrastructure such as electricity grids have been facing a particularly serious cybersecurity challenge due to persistent advanced threats using illegally obtained employee passwords (Rajkumar et al. 2023). Protecting critical infrastructure against malicious insider threats is a challenging task, as these individuals are usually given a high level of trust and access (Al-Mhiqani et al. 2022). These attacks are highly sophisticated and insidious forms of cyberattacks carried out by well-funded and resilient threat actors. Insider threat attacks manifest themselves in various ways, including stealing other users' passwords and exposing the system. Critical infrastructure often has technological limitations, such as the inability to install expensive new equipment for advanced security measures such as physiological biometrics. In critical infrastructure, some of these threats can be mitigated by denying the use of mobile phones, tablets, and gadgets or disabling cameras on the equipment. Nevertheless, if a phishing attack succeeds, an insider can use his or her login credentials on various systems. Despite the changing cyber environment, implementing a new authentication system is essential. This system should capture the biometric characteristics of keystrokes when a user enters a password and compare them to a previously entered password. While insiders can potentially obtain a system administrator's password by "looking over their shoulder" (Krombholz et al. 2015; Mattera and Chowdhury 2021) or implement software to gather the password, they cannot replicate the original user's unique typing behavior. This indicates that keystroke biometrics can effectively protect system access. However, in today's environment, physiological biometrics such as fingerprint scanners, voice authentication, and iris recognition are in high demand. These solutions provide a high level of security. However, deploying such solutions often requires the purchase and installation of new, potentially costly hardware. Physiological biometrics has several disadvantages. Facial recognition can be affected by hats, glasses, and changes in hairstyle. Iris recognition can be deceived by photographs. Additionally, fingerprints can be replicated to impersonate another person. (Abdulrahman and Alhayani 2023). In today's wars, phones are accessed using the physiological biometric data of the deceased person to take their money or cause damage to their social networks (Ugwuoke et al. 2021; Gofman and Villa 2023).

New authentication methods, particularly keystroke dynamics, which essentially authenticate users based on their typing behavior, are the subject of ongoing studies and discussions (Roy et al. 2022; Budžys et al. 2023; Medvedev et al. 2023). Keystroke biometrics is a behavioral authentication method used to verify, for authentication purposes, the unique biometric data of a user's behavior related to their typing patterns. The technology originated in the 19th century when telegraph operators were able to identify the sender by their distinctive writing style (Giancardo et al. 2015). Finding better and new ways to authenticate users when using digital resources is an ongoing challenge in a constantly evolving cybersecurity threat landscape. Organizations, critical infrastructure, cloud storage platforms (Manthiramoorthy et al. 2024), and individuals can improve their security posture and protect valuable digital assets by remaining proactive and adopting advanced authentication methods based on keystroke dynamics.

Keyboard behavior biometrics can be divided into two main categories: Static authentication (SA) and continuous authentication (CA), also known as dynamic authentication. Static authentication requires users to enter passwords or passphrases. Continuous authentication, on the other hand, monitors the user's behavior throughout the entire session. In continuous authentication, real-time data are collected, analyzed, and used to create a profile of the user based on the behavioral patterns of the user during the session (Liang et al. 2020). Before gaining access to critical infrastructure systems, users are required to enter a strong password in their habitual typing pattern. This prerequisite leads us to the area of static authentication. Static authentication based on user typing behavior operates like a sentry, requiring individuals to type their password accurately, as they usually do. In this research, static authentication was selected to facilitate a direct comparison of the results of the newly proposed methodology with the results of previous studies. The equal error rate (EER) was chosen for this comparison. It provides a standardized basis to assess and compare the performance of different methods. To narrow the scope of this paper, this study does not explore the many machine learning techniques applied to static authentication to analyze keystroke patterns as numerical inputs as described in previous studies (Killourhy and Maxion 2009; Zhong et al. 2012; Krishnamoorthy et al. 2018; Elliot et al. 2019; Bicakci et al. 2020). This research investigates the transformation of numerical values into images and the use of deep learning neural networks on keystroke dynamics for static authentication. Processing images for convolutional neural networks (CNNs) is considered superior due to the mathematical capabilities inherent in neural networks (Sharma et al. 2019; Zhu et al. 2021).

The aim of this research is to propose a new method to transform non-image or tabular data into images and to develop a methodology for authenticating users of critical infrastructure based on their keystroke dynamics using the siamese neural network (SNN). The SNN architecture, consisting of CNN branches, is employed to achieve this goal. The selection of this neural network architecture is motivated by its proven ability to enhance the accuracy of static user authentication, as evidenced by its high performance in anomaly detection (Zhou et al. 2020) and in comparison of two or more objects (Ondrašovič and Tarábek 2021). Given the unique characteristics of CNNs, this study focuses on user identification using visual images, which are transformed from tabular data representing keyboard input.

Cybersecurity requires advanced authentication systems, especially in the face of increasing digital threats. This study presents a novel gabor filter matrix transformation (GAFMAT) method for transforming keystroke dynamics into images, which significantly enhances the capabilities of SNNs in user authentication. This addresses a gap in behavioral biometrics, providing a basis for protecting critical infrastructure from insider threats. The proposed approach combined with an SNN is expected to improve the accuracy of user authentication using keystroke biometrics. This approach has been confirmed and verified in the research presented in this paper. The versatility of the proposed method is also considered to be applicable to solving other related problems in the field.

The main contributions of this research are as follows:

- A novel method, GAFMAT, is introduced for transforming non-image data into images. This transformation of keystroke dynamics into images reveals essential behavioral features associated with password typing.
- This research proposes an artificial intelligence-based user authentication methodology that integrates the GAFMAT method for keystroke biometrics. This solution utilizes an SNN in combination with CNNs to compare the features of currently and previously entered passwords, thereby effectively identifying insiders in critical infrastructure.
- The proposed methodology demonstrates its effectiveness on publicly available datasets such as CMU and GREYC-NISLAB. It achieves EERs that are competitive with, and often superior to, other state-of-the-art methods.

The paper is structured as follows. Section 2 summarizes related works on fixed-text keystroke dynamics for user authentication. Section 3 proposes a methodology for user authentication. Section 4 describes the existing techniques and presents the newly developed method for transforming text into images. The experimental setup and the obtained results are presented and discussed in Sect. 5. Section 6 provides an overview of the challenges and future research directions. Finally, Sect. 7 concludes the paper, highlighting the main findings.

2 Related works

Recently, there has been an increasing focus on research on artificial intelligence and machine learning for cybersecurity (Mohamed 2023). Artificial intelligence and machine learning algorithms not only enhance existing security solutions but also enable the development of proactive security measures such as predictive threat analysis. Artificial intelligence has become one of the most important tools for cybersecurity teams. It improves threat detection and response accuracy, strengthening defenses against a variety of security issues and cyberattacks (Azizan et al. 2021; Alfoudi et al. 2022; Kaur et al. 2023).

In the field of cybersecurity, the deployment of security information and event management systems is essential for organizations to proactively detect and address security threats to protect their business operations. Both network-based (NIDS) and host-based (HIDS), intrusion detection dystems (IDSs) play key roles in ensuring robust cybersecurity. NIDS monitors network traffic for anomalies, while HIDS focuses on individual systems, detecting unusual activity or policy violations, including insider threats. A major problem for IDS is data imbalance, especially when detecting rare attacks such as zero-day attacks. To address this problem, Alfoudi et al. (2022) proposed to improve density-based spatial clustering of applications with noise using a new process based on cluster distance measurements. In addition, Azizan et al. (2021) explored the use of machine learning techniques to improve the performance of NIDS in detecting anomalous network flows. By carefully analyzing system activity and user behavior, HIDS can identify potential security breaches within an organization, including those committed by insiders (Al-Mhiqani et al. 2022).

Identity, authentication, and access control management are responsible for restricting access to assets and related objects to authorized users, processes, or devices, and for performing authorized actions. The use of artificial intelligence or machine learning-based techniques can improve user authentication. They improve physical biometrics, behavioral biometrics, or multifactor authentication (Martín et al. 2021; Siam et al. 2021; Kaur et al. 2023). The paper (Zhang et al. 2022) presented a detailed literature review on the application of artificial intelligence in various areas of cybersecurity, including user access authentication, network situational awareness, hazardous behavior monitoring, and anomalous traffic analysis. Authentication is a predetermined process of confirming a person's identity to authorize or deny their access to a protected system. To authenticate a user, the system must first recognize and identify the user among the other users of the system to determine whether the user is a legitimate member of the user group; otherwise, the user will be identified as an insider who must be blocked. Identification methods can recognize the user either through passwords or through additional information (e.g., biometrics) (Siam et al. 2021). The security system should strengthen user access authentication management, accurately detect all kinds of suspicious behavior, and implement the detection of unauthorized connections. The system should ensure user authentication before operation (Zhang et al. 2022).

Biometrics is a group of certain physiological and behavioral characteristics that uniquely distinguish a person from others. Biometric recognition refers to the use of physical or behavioral human attributes to identify individuals (Neves et al. 2016). Physiological biometrics focuses on the unique physical characteristics of a subject. These attributes, which are unique to each individual, generally do not change significantly over time. Examples of physiological biometrics include fingerprints, iris or retinal patterns, facial features, hand geometry, and DNA. This form of biometrics is considered to be relatively stable and difficult to falsify due to its basis in genetically acquired biological features. Conversely, behavioral biometrics focuses on a person's unique behavioral patterns or habits. These characteristics are based on how individuals interact with systems or perform certain actions. Behavioral biometrics covers various aspects such as keyboard biometrics, gait analysis, signature recognition, voice patterns, and even patterns of user interaction with devices (Abuhamad et al. 2020). Unlike physiological biometrics, behavioral biometrics can be influenced by external factors and context and can, therefore, change over time.

As previously mentioned, keyboard biometrics are behavioral biometrics and can be classified into two main categories: fixed-text (static authentication) and free-text (continuous authentication) analysis. For fixed-text authentication, the user is required to input a specific piece of text, usually a password or passphrase. As the text is contextual, it facilitates the comparison of typing patterns across different sessions. Researchers in the field of biometric authentication have focused on fixed-text scenarios (Giot et al. 2015; Zaidan et al. 2017; Shekhawat and Bhatt 2019). By focusing on specific text inputs, such as predefined phrases or sentences, researchers were able to achieve over 93% multiuser identification accuracy on the well-known Carnegie Mellon University (CMU) dataset (Killourhy and Maxion 2009) using XGBoost (Singh et al. 2020) and 94.7% accuracy with a feed-forward multilayer neural network, implementing resilient backpropagation (Gedikli and Efe 2020). However, the fixed-text method of keyboard biometric authentication may be more vulnerable to cyberattacks. An attacker can learn the specific rhythms associated with the fixed text (Serwadda and Phoha 2013; Stanciu et al. 2016).

Biometric authentication research has increasingly focused on free-text authentication models (Acien et al. 2020; Ayotte et al. 2020; Lu et al. 2020), as they have demonstrated promising results. It aims to reflect the variability and realism of natural language input to develop more robust and scalable authentication systems. Free-text keystroke biometrics allow users to enter any text they want, whether it is emails, documents, or other forms of natural writing. The advantage of this method is that it can be implemented in the background without the direct involvement of the user, making it less intrusive. However, in this case, it is more challenging to compare writing patterns because the text is not consistent across sessions or among different users. Both approaches have strengths and weaknesses, and the choice between them often depends on the specific use case and security requirements.

Static authentication is the process of verifying a person's identity by analyzing their typing style when entering a predefined password. When a person types a password, their keystroke pattern is captured. These data are used to create a unique keystroke pattern profile that authenticates the user. Publicly available datasets have been utilized by researchers for static authentication purposes (Killourhy and Maxion 2009). These datasets are valuable resources for researchers in the evaluation and development of building authentication systems, allowing standardized testing and comparisons between different methodologies. Notably, the CMU dataset has been extensively examined due to its large number of samples per person. The developers of the CMU dataset carried out a system performance assessment and obtained an EER of 0.096 using the Manhattan (scaled) distance function (Killourhy and Maxion 2009). The EER, which measures the trade-off between the false rejection rate (FRR) and the false acceptance rate (FAR) as described and applied for performance measurement in related works (Piugie et al. 2022; Sae-Bae and Memon 2022), is a commonly used metric in biometric security systems to measure the effectiveness of a system in correctly identifying an individual. Numerous researchers have used this dataset with machine learning algorithms (Muliono et al. 2018; Krishna et al. 2019; Liu and Guan 2019), resulting in enhanced accuracy for multiuser identification, reaching approximately 94%. In another study, a single multiclass CNN model was trained on 80% of the samples using a specific data augmentation technique. This approach resulted in an EER of 0.023, while without augmentation, the EER was 0.065 (Çeker and Upadhyaya 2017). An inductive transfer encoder (Monaco and Vindiola 2016) obtained an EER of 0.063.

Table 1 offers a comprehensive comparison of keystroke dynamics methodologies and authentication technologies in cybersecurity. Early research on keystroke dynamics focused on statistical methods and machine learning techniques to analyze typing patterns. For instance, Killourhy and Maxion (2009) introduced a comprehensive benchmark dataset and evaluated several anomaly detection algorithms, highlighting the potential of distance metrics such as the Manhattan distance for user authentication. Subsequent studies, such as Zhong et al. (2012), explored the use of nearest neighbor classifiers and novel distance metrics to improve authentication accuracy. More recently, the integration of deep learning techniques has shown significant promise in enhancing the performance of keystroke dynamics-based authentication systems. CNNs and recurrent neural networks (RNNs) have been employed to capture complex temporal and spatial patterns in typing behavior. For example, Çeker and Upadhyaya (2017) demonstrated the effectiveness of CNNs in keystroke dynamics by achieving notable reductions in EER. Similarly, Lu et al. (2020) utilized RNNs for continuous authentication, emphasizing the importance of temporal dynamics in typing patterns.

Most studies, including (Killourhy and Maxion 2009; Zhong et al. 2012; Monaco and Vindiola 2016), have predominantly used the CMU dataset for their keystroke dynamics authentication studies (see Table 1). The popularity of this dataset emphasizes its relevance and reliability in the field. Techniques applied in this field include anomaly detection algorithms, distance metrics, machine learning, CNNs and deep learning. The focus of these studies is on the fixed-text or static authentication mode. This indicates a strong interest in verifying user identity based on specific, consistent input patterns. The most commonly used accuracy metric is the EER, which is the most important metric in biometric authentication systems. The choice of the EER metric provides a balance between security and reliability in authentication systems. This reduces the number of false acceptances and false rejections. The choice of datasets for experimentation and validation of the results of this

Table 1 A comparative analysis of keystroke dynamics and authentication technologies in cybersecurity in related works

Reference	Methodology used	Dataset	Security field	Auth. mode	Accuracy metric
Killourhy and Maxion (2009)	Anomaly-detection algorithms, distance functions	CMU	Keystroke Dynamics Authentication (KD Auth.)	Static	EER, zero-miss rate
Zhong et al. (2012)	Distance functions	CMU	KD Auth	Static	EER
Monaco and Vindiola (2016)	Inductive transfer encoder	CMU	KD Auth	Static	EER
Çeker and Upadhyaya (2017)	CNN	CMU, GREYC labs	KD Auth	Static	EER
Ivannikova et al. (2017)	k-NN-based	CMU	KD Auth	Static	EER
Muliono et al. (2018)	Machine learning, deep learning	CMU	KD Auth	Static	Accuracy
Krishna et al. (2019)	XGBoost	CMU	KD Auth	Static	Accuracy
Liu and Guan (2019)	CNN	CMU	KD Auth	Static	FAR, Accuracy
Ayotte et al. (2020)	Instance-based algorithms, distance functions	Clarkson II, Buffalo	KD Auth., intrusion detection	Continuous	EER
Abuhamad et al. (2020)	Behavioral biometrics	Various behavioral datasets	Behavioral biometrics	Continuous	EER, FAR, FRR
Lu et al. (2020)	CNN, RNN	Clarkson II, Buffalo	KD Auth	Continuous	EER
Singh et al. (2020)	XGBoost	CMU	KD Auth	Static	Accuracy
Acien et al. (2020)	Siamese RNN	Aalto keystroke data	KD Auth	Continuous	EER
Azizan et al. (2021)	Machine learning, classification algorithms	Network intrusion detection data	Network security	Continuous	Recall
Siam et al. (2021)	CNN	Real-World PPG datasets	Biometric identification	Static	Accuracy
Martín et al. (2021)	Statistical techniques, machine learning	UEBA dataset	Behavior analytics	Static	EER
Al-Mhiqani et al. (2022)	Machine learning	Synthetic datasets, CERT dataset	Insider threat detection	N/A	F-score, TNR, FPR, AUC
Sae-Bae and Memon (2022)	Machine learning, distance fuctions	CMU	KD Auth	Static	FAR, FPR, EER
Piugie et al. (2022)	Deep learning	GREYC-NISLAB	KD Auth	Static	EER
This paper	GAFMAT and SNN	CMU, GREYC-NISLAB	KD Auth	Static	EER, Accuracy

paper, in particular the CMU and GREYC-NISLAB datasets, is in line with established practice in the field, as these datasets are widely known and used in the literature.

Deep learning neural networks are increasingly becoming part of user authentication systems, and SNNs are one such approach (Zhou et al. 2020). SNNs and triplet networks have attracted attention for their effectiveness in comparing and identifying similarities between input samples (Koch et al. 2015; Ondrašovič and Tarábek 2021). This enables tasks such as user authentication and verification. Using the capabilities of SNNs, researchers have aimed to improve the accuracy, reliability, and security of user authentication systems in cybersecurity landscapes (Zhou et al. 2020; Tao et al. 2022). An SNN is commonly composed of several parallel branches that share the same architecture and weighting parameters. By processing input samples in each branch, the network acquires the ability to measure and evaluate the similarity or dissimilarity of these samples (Hadsell et al. 2006).

To exploit the potential of SNNs with CNNs, our study explores the transformation of keystroke dynamics into images for network processing. The most crucial initial step is to explore methods for transforming time series data (keystroke data) into images (visual format). Such conversion allows CNNs to efficiently extract and learn important features from the data, taking advantage of their ability to perform image-oriented tasks such as image classification, object detection, and segmentation. It also allows CNNs to identify patterns and perform mathematical operations on transformed data, which can increase productivity compared to using only tabular or numerical data. The ability of CNNs to derive higher-level abstractions from the visual representation of tabular data provides favorable prospects for improved performance in additional tasks. There are various methods for transforming numerical data or time series into images. Some of these include the Markov Transition Field, the Gramian Angular Summation Field, the Gramian Angular Difference Field, and the Recurrence Plot methods (Estebsari and Rajabi 2020; Medvedev et al. 2023).

3 Methodology

In the emerging field of cybersecurity, especially in times of war (Gofman and Villa 2023), there is an increasing demand for advanced IPSs that leverage behavioral biometrics through deep neural networks. This approach is becoming increasingly vital as individuals face the risk of being forced to disclose passwords or becoming victims of sophisticated phishing attacks. When the threat of conflict approaches, it is essential to consider all possible measures to protect national security. This encompasses protecting critical infrastructure, where a breach leads to catastrophic consequences for national stability. Given that malicious insider threats are identifiable (Azizan et al. 2021), incorporating keystroke biometrics in user authentication serves as an essential first line of defense against the unauthorized use of others' passwords.

The proposed static authentication process is designed to track the keystroke dynamics of the user's entered password. Each keystroke is associated with a timestamp, and relevant features are extracted from time series data, and subsequently transformed into an image representation. This image is then compared against a pre-established database to ascertain the presence of a similar, previously entered password linked with the username. Access to the system is granted if a corresponding match is identified. Conversely, when no matching password is identified, the IDS generates an informational log, prompting the user to re-enter the password.

If the user fails to enter the password correctly after a certain number of attempts defined by group policy, the user's account becomes locked out. This action triggers the intrusion prevention system, which then generates a critical log. Subsequently, specialists at the security operation center are alerted (see Fig. 1). Hence, even in scenarios where a password within the critical infrastructure is compromised or illegally acquired by an unauthorized person, the system is capable of detecting inconsistencies in the input pattern. This process effectively shows that the current user is not the legitimate owner of the password. Such a mechanism significantly increases the system's resilience to potential security breaches.

This approach introduces a significant opportunity to integrate password authentication techniques into critical infrastructure. The challenge lies in discerning the similarity of keystroke dynamics to ascertain whether the password input was executed by a legitimate user or an insider. A technique that transforms password patterns to train SNNs is able to accurately identify the genuine user associated with an entered password.

The SNN (or triplet network) architecture uses three CNN branches and a triplet loss function at the output layer. This setup estimates the distance between images as detailed in (Schroff et al. 2015) (see Fig. 2). Throughout this paper, the term "Siamese neural network" will be used, as the triplet network is an enhancement of the SNN (Bromley et al. 1993; Schroff et al. 2015). More recently, in the context of SNNs, training often involves the use of triplets:

- Anchor—a reference sample against which other items are compared,
- Positive—a sample that is similar or related to the anchor,
- Negative—a sample that is not similar or related to the anchor.

During the network training process, triplets are formed. These triplets consist of an anchor image, a positive image from the same user, and a negative image from another user. After training, the SNN creates corresponding embeddings for all triplets. These embeddings are vectors in a multidimensional latent space that represent the input data or images. The idea

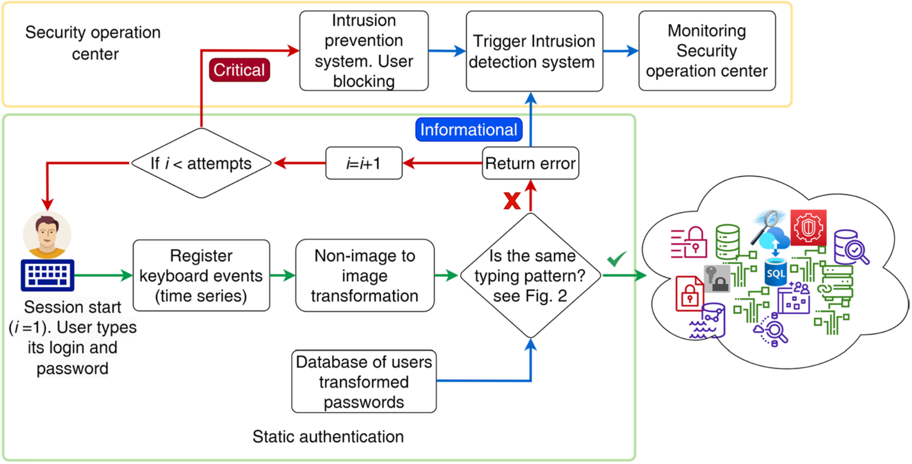

Security operation center
Critical
Intrusion prevention system. User blocking
Trigger Intrusion detection system
Monitoring Security operation center
Informational
If i < attempts
i=i+1
Return error
Session start (i =1). User types its login and password
Register keyboard events (time series)
Non-image to image transformation
Is the same typing pattern? see Fig. 2
Database of users transformed passwords
SQL
Static authentication

Fig. 1 Schematic representation of the user authentication process using an intrusion detection system and an intrusion prevention system based on user typing behavior

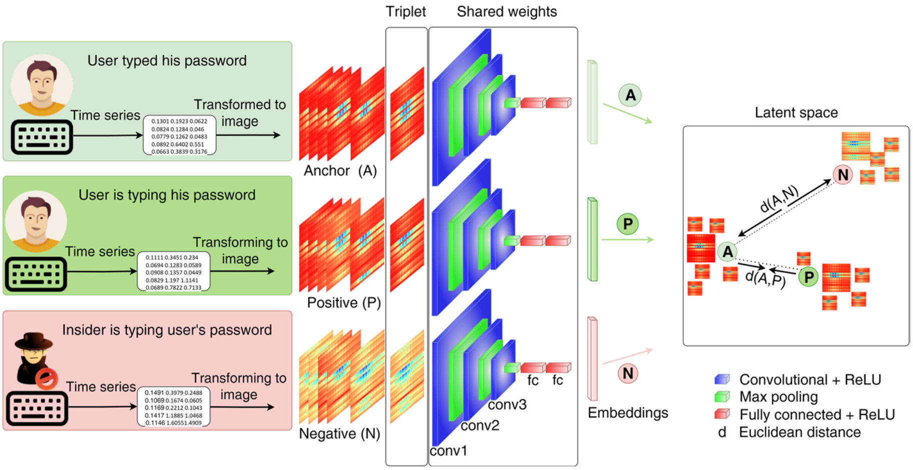
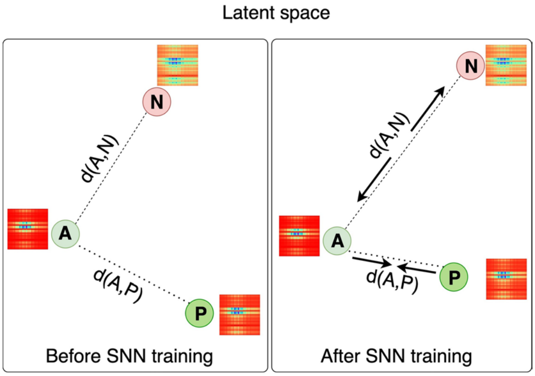


Fig. 2 Schematic representation of the proposed methodology for time series transformation from keystroke biometric data features into images and the training process of the SNN with CNN branches

is that similar images will yield embeddings located close to each other in this space, while dissimilar images result in embeddings that are more distant (see Fig. 3). To determine the similarity between two images, the distance between their embeddings can be computed using metrics such as Euclidean distance or cosine similarity. In the context of triplets, the distance between the anchor and the positive sample in a multidimensional latent space should be small, indicating high similarity. The distance between the anchor and the negative sample should be significant, indicating low similarity. By exploring these distances, decisions can be made about the similarity or dissimilarity of new, unidentified images (or samples) compared to known anchors.

The training process of the SNN involves searching for similarities between the positive and anchor images while promoting dissimilarities between the anchor and negative samples. The triplet loss function (Eq. 1) is designed to minimize the distance between the anchor and the positive image (as they belong to the same class) while maximizing the distance between the anchor and the negative image (as they belong to different classes), depending on the margin size (see Fig. 3).

Fig. 3 Example of a triplet before and after training an SNN: the triplet loss function minimizes and maximizes the corresponding distances during network training



$$L(a,p,n) = \max\left(||f(a) - f(p)||^2 - ||f(a) - f(n)||^2 + m, 0\right), \tag{1}$$

where

- $||f(a) - f(p)||^2$ is the squared Euclidean distance between the embeddings of the anchor and positive samples computed in a multidimensional latent space,
- $||f(a) - f(n)||^2$ is the squared Euclidean distance between the embeddings of the anchor and negative samples computed in a multidimensional latent space,
- f is the embedding function that maps an input to its embedding,
- m is the margin that is enforced between positive and negative pairs.

Numerous researchers have previously used a triplet loss function to train their models, considering it a suitable option for user authentication (Ding et al. 2015; Cheng et al. 2016; Dong and Shen 2018; Yan et al. 2021; Sandhya et al. 2022). The triplet loss function includes a margin that sets the desired separation between positive and negative samples relative to the anchor. The margin within the triplet loss function allows a clear distinction between similar and dissimilar samples, ensuring that the distance or dissimilarity between the anchor and the negative sample is greater than the distance between the anchor and the positive sample by at least a predefined threshold value.

While SNNs with CNN branches perform well in image comparison tasks (Melekhov et al. 2016; William et al. 2019; Valero-Mas et al. 2023), their direct applicability for password keystroke patterns can be challenging due to inherent differences in data structures. By transforming keystroke dynamics into images for CNN training, which is a branch of the SNN, this approach leverages the strengths of these networks (Sharma et al. 2019; Zhu et al. 2021). This transformation enhances the network's ability to distinguish certain behavioral biometric differences between authentic users and insider typing patterns.

The choice of the SNN combined with CNN branches was based on their effectiveness in image recognition tasks, as they are able to learn similarity measures between input data. Traditional classification networks may struggle with significant class imbalance, while SNNs may be more robust in such scenarios. Instead of classifying a large number of classes, they measure similarity to a reference (anchor). SNNs generalize well to new data. Once trained, they can compare any new sample to a known reference without the need for retraining. The specific hyperparameters of the SNN used in this study are summarized in Table 2. In the SNN, the total parameters of each branch of the CNN mainly depend on the image size. A summary of the CNN used for SNNs, with an input image size of 31×31, is provided in Table 3. Each convolutional layer is followed by batch normalization and max pooling operations. This is followed by a flattened layer, the output of which is used as the input to the dense layer. The last layer

Table 2 Hyperparameters of the SNN architecture

Hyperparameters	Options
Convolutional layers	3
Kernel number	128
Kernel size	8, 6, 4
MaxPooling filter size	3, 2, 2
Dense	512, 256
Output activation function	ReLU

Table 3 Summary of the CNN used in the SNN architecture

Layer (type)	Output shape	Number of parameters
InputLayer	(None, 31, 31, 3)	0
Conv2D	(None, 31, 31, 128)	24704
BatchNormalization	(None, 31, 31, 128)	512
MaxPooling2D	(None, 15, 15, 128)	0
Conv2D	(None, 15, 15, 128)	589952
BatchNormalization	(None, 15, 15, 128)	512
MaxPooling2D	(None, 7, 7, 128)	0
Conv2D	(None, 7, 7, 128)	262272
BatchNormalization	(None, 7, 7, 128)	512
MaxPooling2D	(None, 3, 3, 128)	0
Flatten	(None, 1152)	0
Dense	(None, 512)	590336
Dense	(None, 256)	131328
Lambda	(None, 256)	0

of the network has 256 outputs that have been normalized using L2 normalization. The network output can be considered as an embedding of the original input. The network, which has a depth of 12 layers, covers a total of 1,600,128 parameters.

Using the SNN architecture (Bromley et al. 1993; Schroff et al. 2015), the effectiveness of IDS and IPS can be improved by better identifying the user by their unique password input patterns. Consider a given scenario where each password entry is transformed into an image and stored in a database associated with the corresponding username. The system, based on an SNN, is designed to analyze and capture the unique typing characteristics of a user as he or she interacts with the system using a keyboard. This network processes input data to identify and differentiate individual typing patterns. The behavioral data are then aggregated using complex algorithms to create a multidimensional representation in the latent space. This results in individual clusters, each corresponding to a different user. These clusters allow for a detailed study of each user's typing behavior.

Each individual possesses a distinct typing style, making us unique in the way we interact with keyboards. If unauthorized access occurs or credentials are compromised, our proposed methodology can identify and prevent unauthorized individuals from exploiting stolen or purchased passwords to gain access to the system. By leveraging the inherent uniqueness of typing patterns, our approach can effectively detect and mitigate unauthorized login attempts. This increases the security and protection of user credentials in the system.

4 Non-image to image data transformation

Building on the methodology described in Sect. 3, this section introduces feature extraction from keystroke dynamics, discusses existing methods for transforming data into images, and describes a new method proposed in this paper.

4.1 Keystroke dynamics

Keystroke biometric models are developed by recording keystroke timing, which captures the intervals between each key press and release event. Analyzing this timing information enables the extraction of various features, such as Hold time, Release-Press time, Press-Press time, and Release-Release time. These features yield critical insights into users' typing patterns (see Fig. 4). This method of collecting time series data forms the basis for comprehensive analysis. It enables the extraction of features that effectively describe the unique and complex dynamics of individual keystrokes. This process provides valuable information on the unique typing patterns of users.

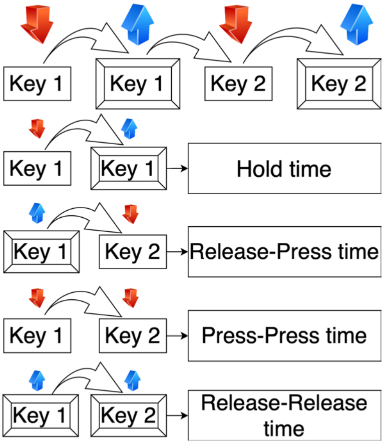
4.2 Image-based time series data

For some numerical data, the ordering of features can be reversed in a two-dimensional space to explicitly represent the relationships among these features. In this way, it becomes possible to transform tabular data into images, from which CNNs can learn. By exploiting these feature relationships, CNNs may enhance prediction or classification performance compared to models trained solely on tabular data (Zhu et al. 2021). In the transformation process, each sample of tabular data is converted into an image. In these images, features and their values are represented by pixels and pixel intensities, respectively.

A number of methods exist for transforming (or encoding) numerical or non-image data into images, such as the gramian angular summation field (GASF), the gramian angular difference field (GADF), the Markov transition field (MTF), and the recurrence plot (RP) methods. These methods are used in various applications, including biometrics for user authentication (Dias et al. 2020; Wang and Oates 2015; Medvedev et al. 2023). The purpose of these transformations is to extract meaningful features from the data, enabling further analysis using deep learning techniques. Each method under review emphasizes specific data characteristics, such as frequency, distribution, similarity, amplitude fluctuations, periodicity, or underlying patterns.

Techniques such as GASF, GADF, MTF, and RP can be used to improve the performance of deep learning algorithms for user authentication from biometric data. GASF and GADF, like MTF, are able to capture important time series features, including periodicity,

Fig. 4 Vizualizing keystroke dynamics capturing model



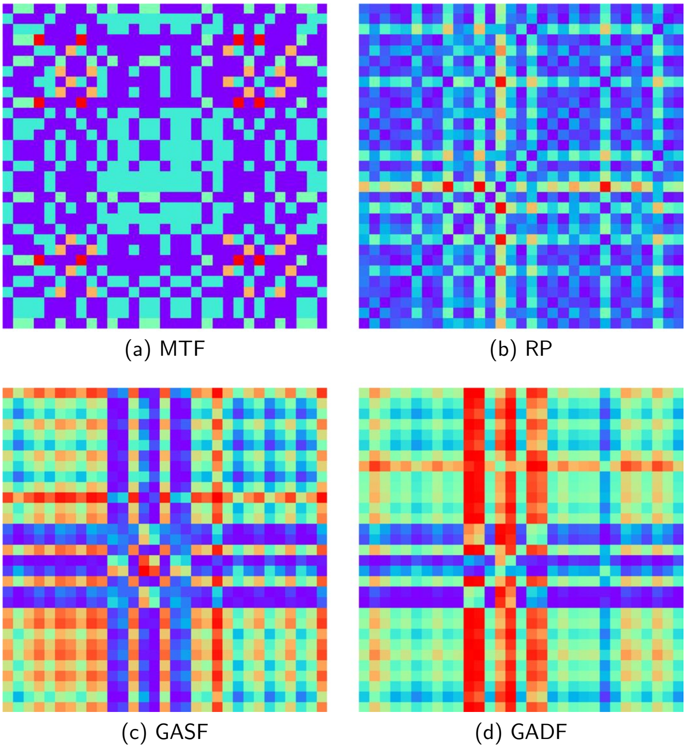
trend, and irregularity. The GASF and GADF methods, which are based on the gramian angular field (GAF) technique, transform time series signals into images by transferring them into polar coordinate space (Wang and Oates 2015). The GASF method considers the sum of the angles, whereas GADF emphasizes the difference, thereby highlighting distinct aspects of the data. RP is a method for analyzing dynamical systems and time series data. It facilitates the uncovering of the overall structure, non-stationarity, and hidden recurrent elements of a time series. Additionally, RP provides a graphical representation of recurrent dynamics. It characterizes the proximity of states in the state space of a dynamical system reconstructed with a time delay (Chen et al. 2018). RP is less effective at encoding very long sequences. For very long sequences, the resulting RP images become so large that their discretization is relatively small (Zhang et al. 2020). In contrast, MTF transforms time series into visual representations. This approach captures significant dynamics and facilitates the use of CNNs to extract and analyze features in various domains (Zhao et al. 2022).

These methods represent only a few of the ways in which time series or non-image data can be transformed into images for analysis using deep learning techniques. They demonstrate diverse approaches for transforming time series or non-image data into images suitable for deep learning analysis. The choice of an appropriate method largely depends on the features of the data and the specific problem to be solved (Medvedev et al. 2023).

The CMU dataset (Killourhy and Maxion 2009) is used as a benchmark for comparing various methods, including the new method detailed in the next section. This comparison aims to identify the most appropriate methods for the problem under analysis. When the password ".tie5roanl" is typed, the keyboard generates 31 different time series, each corresponding to the keystroke dynamics of the password. Employing transformation methods such as GASF, GADF, MTF, and RP, as described in (Estebsari and Rajabi 2020; Medvedev et al. 2023), allows for the transformation of these individual keystroke dynamics into images. Consequently, this process yields four different images for each method, all representing the same password, as illustrated in Fig. 5.

4.3 Gabor filter matrix transformation

Drawing upon insights gained from the analysis of the literature on the transformation of non-image data into images, we have developed a novel method named gabor filter matrix transformation (GAFMAT). This approach is grounded in the principles of the Gabor filter (Kamarainen et al. 2006). Keystroke dynamics, which include timing and rhythm variations, are crucial for identifying individual typing patterns. The Gabor filter has high performance in both frequency and time localization, enabling it to capture these variations effectively. It has the capability to isolate specific frequencies while simultaneously retaining information about the timing of events in the signal. The keystroke dynamics data may contain noise due to variations in typing speed, keyboard differences, or environmental factors. The specificity of the Gabor filter provides a natural robustness to such noise. It filters out irrelevant fluctuations while preserving the essential characteristics of keystroke dynamics (Imamura and Arizumi 2021). The proposed method, GAFMAT, transforms time series data into image representations. This novel approach shows promising potential for improving the analysis and interpretation of keystroke dynamics in authentication systems. The Gabor filter has been chosen for its specific design for feature extraction in two-dimensional images. The process involves adapting and applying the Gabor filter to one-dimensional time series or discrete signals (see Eq. (2)), thereby emphasizing features of keystroke dynamics (see Fig. 6). In the figure, there are two curves: the original discrete



(a) MTF

(b) RP

(c) GASF

(d) GADF

Fig. 5 Example of a typed password of the same user obtained by different methods: **a** Markov transition field, **b** Recurrence plot, **c** Gramian angular summation field, **d** Gramian angular difference field

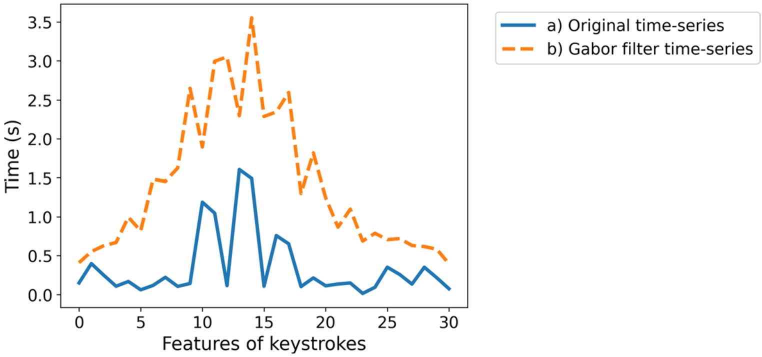


Fig. 6 Emphasizing the time series features of keystroke dynamics using the Gabor filter: blue for the discrete signal and dashed orange for the discrete signal after applying the Gabor filter

signal is depicted as a blue line, while the dashed orange line represents the values of the discrete signal after applying the Gabor filter. It is important to note that the Gabor filter, by its nature, highlights features of the discrete signal. As shown in the figure, the filter particularly emphasizes the peaks of the signal. By using the distinctive properties of the Gabor filter, the goal is to improve the representation and visualization of keystroke dynamics. This improvement facilitates more effective discrimination and analysis of key features within time series data.

$$\begin{aligned} gabor &= \exp\left(-\frac{0.5 \cdot x'^2}{\sigma^2}\right) \cdot \cos\left(2\pi \cdot \frac{x'}{\lambda} + \psi\right), \\ x' &= x \cdot \cos\theta, \end{aligned} \tag{2}$$

where

- σ: Parameter defining the filter width. A larger σ results in a wider filter.
- θ: Orientation of the filter. In the 1D case, it effectively scales the x values.
- λ: The wavelength of the sinusoidal factor, which determines the frequency of the filter. A larger λ results in a lower frequency filter.
- ψ: This is the phase offset of the sinusoidal factor, which can be used to create band-pass or band-reject Gabor filters.

The *GaborFilter* function (see Algorithm 1) is used to apply a Gabor filter to the timestamps generated by password input. This function takes a discrete signal, representing the timestamps of entered passwords, and several parameters, including σ, θ, λ, and ψ, which define the characteristics of the Gabor filter. The function determines the value of the discrete signal and generates a range of values based on the σ parameter. These values are then transformed using the θ parameter. The Gabor filter is calculated by combining the exponential and cosine functions based on the provided parameters. The resulting filter is then normalized. Finally, the signal is convolved with the Gabor filter, and the output is returned as a filtered signal (see Algorithm 1).

Algorithm 1 Gabor filter algorithm

```
function GaborFilter($discrete\_signal, \sigma, \theta, \lambda, \psi$)
    $n \leftarrow$ length of $discrete\_signal$
    Initialize $x$ as an array of size $n$ generating evenly-spaced values in an interval $(-3\sigma, 3\sigma)$
    $x \leftarrow x \cdot \cos\theta$
    Initialize $gabor$ as an empty array of size $n$
    $gabor \leftarrow \exp\left(-0.5 \cdot \left(\frac{x'}{\sigma}\right)^2\right) \cdot \cos\left(2\pi \cdot \frac{x'}{\lambda} + \psi\right)$
    $gabor \leftarrow \frac{gabor}{\sqrt{\sum_{i=0}^{n-1} gabor[i]^2}}$
    $gabor \leftarrow Convolution(discrete\_signal, gabor)$
return $gabor$
end function
```

The GAFMAT algorithm (see Algorithm 2) is specifically designed to create an image representation of a given discrete signal by applying the Gabor filter (see Algorithm 1). The *GaborFilter* function uses a discrete signal and a list of parameters

Table 4 List of parameters used for the GAFMAT algorithm

Parameter	Values
σ	2, 4, 8, 16
θ	$0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$
λ	16, 8, 4, 2
ψ	$0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$

$(\sigma, \theta, \lambda, \psi)$ (see Table 4). The algorithm begins by initializing an empty image array, matching the shape of the input signal. It then iterates through various combinations of the parameter, applying the *GaborFilter* function to the discrete signal with each iteration. Finally, the algorithm returns the resulting image, which represents the combination of multiple Gabor-filtered versions of the original discrete signal (see Algorithm 2). The outer product of two arrays is computed, producing a new array where each element is the product of the corresponding elements from the input arrays. This computation involves all possible pairwise products of the original time series array and the values obtained by the GAFMAT, which are then systematically arranged in a matrix structure (Eq. 3). The resulting matrix *image*2*D* represents the pairwise products of each element in arrays *a* and *b*. The matrix *image*2*D* is finally visually represented as an image offering a comprehensive visual representation. Such a visualization provides a clear and intuitive understanding of the data, enabling efficient interpretation and analysis of key patterns and relationships (see Fig. 7).

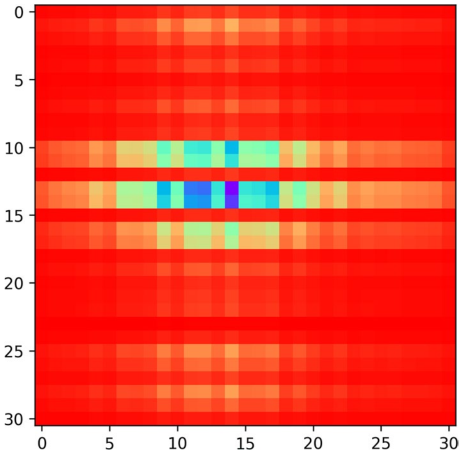


Fig. 7 The result of transforming the time series features of keystroke dynamics into an image using the GAFMAT algorithm

Algorithm 2 GAFMAT algorithm

```
Require: $discrete\_signal, \sigma\_list, \theta\_list, \lambda\_list, \psi\_list$
$n \leftarrow$ length of $discrete\_signal$
$image \leftarrow$ create zero array of size $n$
$combinations \leftarrow CartesianProduct(\sigma\_list, \theta\_list, \lambda\_list, \psi\_list)$
    ▷ The set of all possible pairs (see Table 4)
for each $(\sigma, \theta, \lambda, \psi)$ in $combinations$ do
    $gabortemp \leftarrow gabor(discrete\_signal, \sigma, \theta, \lambda, \psi)$
    ▷ see Algorithm 1
    $gabor \leftarrow transpose(gabortemp)$
    $image2D \leftarrow OuterProduct(image, gabor)$
end for
return $image2D$
```

$$image2D = \begin{bmatrix} a_1b_1 & a_1b_2 & \cdots & a_1b_n \\ a_2b_1 & a_2b_2 & \cdots & a_2b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_nb_1 & a_nb_2 & \cdots & a_nb_n \end{bmatrix} \tag{3}$$

The newly proposed GAFMAT method enhances the user's keystroke dynamics by scaling up significant values. This scaling results in larger numerical values that are accentuated with a variety of colors. Its robustness to common noise and interference also distinguishes it from traditional approaches. The next section compares the novel method with established techniques and provides an explanation of the results, demonstrating the effectiveness and uniqueness of the proposed method.

5 Experiments and results

In this section, the performance of the proposed methodology is evaluated. To demonstrate the distinctive features and effectiveness of the methodology, experiments were conducted using two publicly available fixed-text datasets: the CMU dataset and the GREYC-NISLAB dataset. The experiments for our study were conducted on an Apple MacBook Pro with an M1 Pro chip, featuring a 10-core CPU and a 16-core GPU. Each core is split into 16 execution units (EUs), and each EU consists of 8 arithmetic logic units (ALUs), for a total of 256 EUs and 2,048 ALUs across the GPU. This powerful setup, equipped with 32 GB of unified RAM, ensures the efficient handling of complex computational processes essential for deep learning-based networks, as detailed in Table 5. For this study, TensorFlow (Abadi et al. 2015), a widely used public library for large-scale analysis and machine learning, was chosen because it can use multiple CPUs or GPUs.

Table 5 Experimental platform technical specifications and system configuration

Platform	Details
Model	Apple MacBook Pro 14-inch
Processor model	Apple M1 Pro
CPU	10-core
GPU	16-core
RAM	32 GB unified
Disk space	512 GB SSD
Operating system	macOS Sonoma
Python framework	TensorFlow 2.9.1
	Matplotlib 3.7.0
	Numpy 1.22.4
	Pandas 1.5.3

5.1 Performance metrics

Choosing the right metric is critical to evaluate the performance of SNN-based models. These metrics assess the accuracy of the model in distinguishing between legitimate and unauthorized users, which is very important for ensuring system performance in a dynamic cybersecurity environment. Each metric was computed on a per-batch basis in our analysis. The validation dataset represents 30% of the total dataset. Subsequently, each metric was evaluated for every individual batch, and the average value across all batches was reported as the final result. This approach ensured that the metrics were representative of the overall performance of the validation dataset while considering the inherent variability between batches. A comprehensive set of metrics was employed to assess the performance of the trained models, including the following:

- The EER, the most commonly used accuracy metric in biometric authentication systems (see Table 1)
- Area under the ROC curve (AUC)
- Euclidean distance:
  - Between the embeddings of the anchor and positive samples in a multidimensional latent space (AP_ED)
  - Between the embeddings of the anchor and negative samples in a multidimensional latent space (AN_ED)
- Standard deviation of Euclidean distances:
  - Between the embeddings of the anchor and positive samples in a multidimensional latent space (AP_STD)
  - Between the embeddings of the anchor and negative samples in a multidimensional latent space (AN_STD)
- Cosine similarity:
  - Between the embeddings of the anchor and positive samples in a multidimensional latent space (AP_CS)
  - Between the embeddings of the anchor and negative samples in a multidimensional latent space (AN_CS)

- Accuracy (see Eq. (4))

In evaluating the performance of the proposed methodology, special attention was focused on the EER as the main metric. The EER was chosen due to its wide acceptance and use in biometric authentication systems as a balanced measure of accuracy. The EER is a specific point on the ROC curve. It is a rate at which the FAR and FRR are equal, offering a simple and effective measure of a system's performance in distinguishing between authorized users and impostors. In the experiments conducted, the accuracy metric for both the validation and test datasets is adapted for the classification task using an SNN. This metric determines the fraction of cases where a positive score outperforms a negative score, as shown in Eq. (4). The positive score represents the Euclidean distance between the embeddings of the anchor and positive samples in a multidimensional latent space. The negative score corresponds to the Euclidean distance between the embeddings of the anchor and negative samples in the same multidimensional latent space.

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N} I(\text{pos\_scores}_i < \text{neg\_scores}_i), \tag{4}$$

where

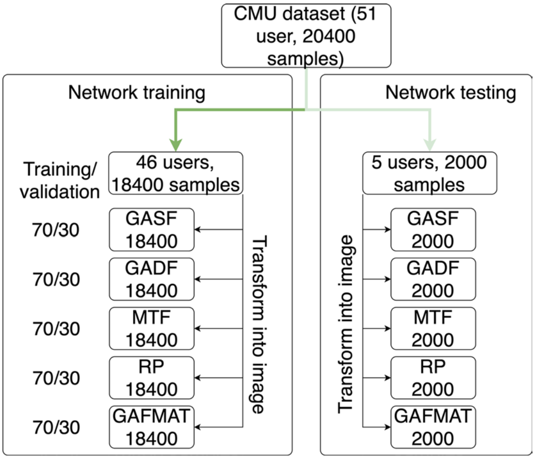
- N is the total number of samples,
- pos_scores_i represents the positive score for the i-th sample,
- neg_scores_i represents the negative score for the i-th sample,
- $I(\text{pos\_scores}_i < \text{neg\_scores}_i)$ is an indicator function that returns 1 if the condition $\text{pos\_scores}_i < \text{neg\_scores}_i$ is true, and 0 otherwise,
- $\sum_{i=1}^{N} I(\text{pos\_scores}_i < \text{neg\_scores}_i)$ counts the number of times the positive score is less than the negative score.

5.2 CMU dataset analysis and results

Fixed-text datasets are chosen to perform the experiment. One is the Carnegie Mellon University dataset (Killourhy and Maxion 2009), which has been extensively examined due to its large number of samples per person and ease of comparison with alternative methodologies and results. A total of 51 individuals were enlisted to participate in a typing task involving the entry of the strong password ".tie5Roanl". Each participant completed eight data collection sessions, typing the password 50 times per session. This results in a dataset of 20,400 password samples.

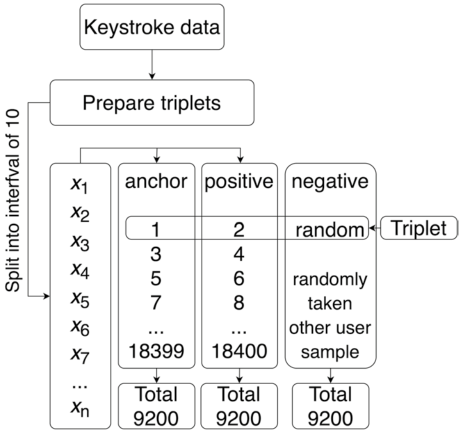
Prior to commencing the data analysis, a random selection was made to exclude the data of five users (see Fig. 8), resulting in a dataset comprising 46 individuals with 18,400 samples. The excluded data from those five users, consisting of 2000 samples, were set aside in a separate folder for testing purposes. This segregation was intended to ensure that the network would not be exposed to any of these data during the network training phase. The password samples from both the training/validation folder (18,400 samples) and the testing folder (five users with 2000 samples) were transformed into image representations. This process yielded five datasets for network training/validation, each processed using different conversion methods (GASF, GADF, MTF, RP, GAFMAT). Additionally, five folders were created, each containing samples (images) of five users for network testing using the same

Fig. 8 The process of preparing CMU data for model training/validation and testing



transformation methods. In each dataset, each user entered the password 400 times, which were then split into two equal parts of 200 attempts each. The system alternated trials to classify the user's password entry behavior. Specifically, every second trial was considered an anchor sample of the user's password behavior, and the trials immediately following it were considered positive samples (see Fig. 9). This division was based on the observation that users who become familiar with the password improve their typing speed over time and develop a more stable typing pattern. Therefore, when comparing anchor samples with positive samples, one should compare those which, over time and with the learning of the password, would not have drifted apart between trials. As a result, each dataset consisted of 9200 positive samples (images) and 9200 anchor samples (images). 70% of created triplets were used for training and the remaining 30% for validation. Additionally, for testing purposes, 1000 positive images and 1000 anchor images were extracted for the test datasets. To create training triplets for the SNN, the anchor and positive samples were taken from the same user, while the negative sample was randomly selected from a different user. This procedure was repeated for each dataset with different conversion methods.

Fig. 9 Splitting CMU data into the anchor and positive samples for each transformed dataset using the GASF, GADF, MTF, RP, and GAFMAT methods for triplet preparation



In the experiment, triplets were fed as input to the SNN, and a margin size of 0.5 ($m = 0.5$, see Eq. 1) was used. This decision was based on our previous research, where an investigation was conducted to determine the optimal margin size for various experimental setups (Budžys et al. 2023; Medvedev et al. 2023).

The SNN was trained using the Adam optimizer over 100 epochs. To prevent overfitting, an early stopping function was enabled, which stopped training if the model's performance on the validation dataset did not improve. Additionally, a batch size of 128 was chosen for efficient computation and optimization. The batch size was determined after a series of experiments to find the balance between computational efficiency and model performance. The validation dataset was used to monitor the performance of the model, and the best weights were saved based on the validation loss. Following training, the optimal weights from each training epoch were saved, resulting in the storage of five sets of different network weights related to GADF, GASF, MTF, RP, and GAFMAT.

Table 6 summarizes the results obtained by each of the different transformation methods applied to the validation dataset. The results are evaluated according to the metrics described in Sect. 5.1. The data in the table indicate that the most accurate methods were GADF, with an accuracy of 0.99077, and GAFMAT, with an accuracy of 0.98935. Using RP and GASF, the values obtained were 0.98331 and 0.98473, respectively. In contrast, the MTF showed a noticeably lower accuracy of 0.94744.

The results in Table 6 suggest that GADF outperforms the other methods in terms of distance metrics, as it yields lower AP_ED values than the other methods. This implies that the positive images are positioned closer to the anchor. However, the higher AN_ED values for GADF indicate that the method struggles to distinguish negative images from the anchor, in contrast to the superior performance of GAFMAT, which achieved an AN_ED value of 1.7637. A higher AN_ED value suggests that the other methods possess the ability to better discriminate negative images relative to the anchor. In summary, although GADF excels in proximity to the anchor with its lower AP_ED, and its comparative weakness in distinguishing negative images is evident from the higher AN_ED values. GADF exhibited the lowest AP_STD value of 0.27487, indicating less variability within the anchor and positive samples. Similarly, GADF had the highest AN_STD value of 0.32888, indicating more variability within the anchor and negative samples. This trend was also observed for the other methods.

Table 6 Results of image transformation methods on keystroke dynamics data from the CMU dataset using the GADF, GASF, RP, MTF, and GAFMAT algorithms: Metrics-based evaluation on validation data

Metrics	Non-image to image transformation methods				
	GADF	GASF	RP	MTF	GAFMAT
Accuracy↑	0.99077	0.98473	0.98331	0.94744	0.98935
EER↓	0.04794	0.05540	0.05327	0.12074	0.04545
AUC↑	0.98612	0.98290	0.98394	0.94862	0.98668
AP_ED↓	0.44127	0.47255	0.43633	0.56487	0.48600
AN_ED↑	1.72784	1.71689	1.68884	1.59469	1.76378
AP_STD↓	0.27487	0.29295	0.28245	0.36906	0.31383
AN_STD↓	0.32888	0.34455	0.34881	0.40005	0.31295
AN_CS↓	0.45772	0.45264	0.46871	0.46011	0.43755
AP_CS↑	0.77936	0.76373	0.78183	0.71756	0.75700

The GADF has the highest AP_CS value of 0.77936, indicating a high cosine similarity between the anchor and positive samples. On the other hand, GADF also had the highest AN_CS value of 0.45772, indicating a relatively high cosine similarity between the anchor and negative samples. The other methods showed similar patterns, where GADF generally had higher AP_CS and AN_CS values. In the context of cosine similarity, a higher value is generally considered better. When the cosine similarity between two vectors (anchor and positive or anchor and negative) is closer to 1, the vectors point in a similar direction and have a higher degree of similarity. This is beneficial in many applications where similarity or correlation between vectors is important, and can be useful in a variety of tasks, such as document similarity, recommender systems, and pattern recognition.

From Table 6, it can be observed that the lowest EER value of 0.04545 was obtained using GAFMAT. This indicates a lower threshold at which the trade-off between the FAR and FRR is achieved. Other methods also showed relatively low EER values, except for the MTF, which had a higher EER of 0.12074. The highest AUC value (0.98668) was obtained using the GAFMAT method. The GADF method yielded results close to those of GAFMAT, with a value of 0.98612. The use of GASF and RP resulted in AUC values of 0.9829 and 0.98394, respectively. The MTF had a slightly lower AUC value of 0.94862.

In a comprehensive evaluation, the use of the GAFMAT and GADF methods showed promising results on a number of metrics, such as accuracy, distance measure, cosine similarity, EER, and AUC. The empirical results highlight the potential effectiveness of GAFMAT and GADF as transformation methods for the analysis of the dataset compared to the other methods considered.

In the following research, a comparative analysis was conducted to evaluate the effectiveness of different transformation methods on a test dataset. The test dataset consists of previously unseen data samples that were processed using the same transformation method. By evaluating the results obtained from each method, we aimed to gain insights into their effectiveness and identify possible variations in performance (see Table 7).

The results of the validation data provided in Table 6 indicate clear variations in the performance of the different methods, with some methods demonstrating better performance than others. However, it is important to highlight that the results obtained on the test data (see Table 7) are significantly lower than those obtained on the validation data. These differences are consistent across the test dataset, indicating that the performance differences observed in the validation dataset are also valid for the test data. The accuracy of

Table 7 Results of image transformation methods on keystroke dynamics data from the CMU dataset using the GADF, GASF, RP, MTF, and GAFMAT algorithms: Metrics-based evaluation on test data

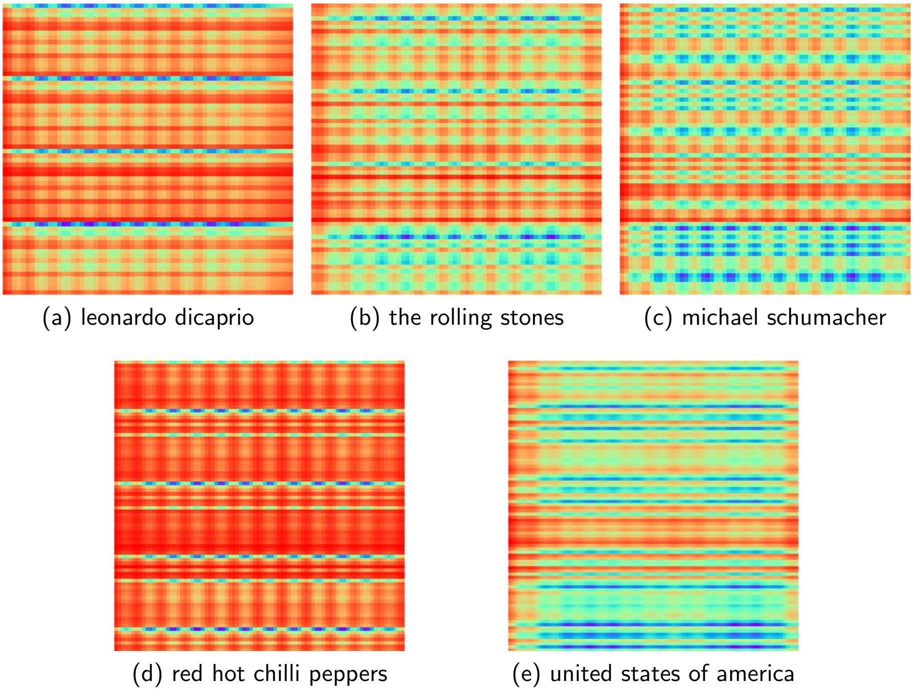
Metrics	Non-image to image transformation methods				
	GADF	GASF	RP	MTF	GAFMAT
Accuracy↑	0.86800	0.8540	0.82900	0.85400	0.86600
EER↓	0.21000	0.24500	0.23900	0.24500	0.21500
AUC↑	0.85928	0.83398	0.83937	0.83398	0.85951
AP_ED↓	0.73164	0.86555	0.84481	0.86555	0.83616
AN_ED↑	1.41323	1.50249	1.50904	1.50249	1.52453
AP_STD↓	0.41727	0.45697	0.47537	0.45697	0.44798
AN_STD↓	0.43871	0.44504	0.44953	0.44504	0.42488
AN_CS↓	0.46378	0.40799	0.41154	0.40799	0.40983
AP_CS↑	0.63418	0.56723	0.57760	0.56723	0.58192

the models decreased by approximately 10% to 0.86, and the EER increased from 0.05 to approximately 0.20. This outcome suggests that the SNN incorrectly classifies one out of every five negative samples as positive, highlighting a significant rate of false positives. Nevertheless, the analysis shows the promise of using the GAFMAT and GADF methods over other methods in analyzing the test dataset.

5.3 GREYC-NISLAB dataset analysis and results

In the initial phase of the experiments using the CMU dataset, it was empirically determined that the proposed GAFMAT method achieved the lowest EER value. Therefore, to further validate the effectiveness of the proposed methodology, the analysis was extended to include an additional dataset of fixed-text passwords. These additional experiments allowed us to evaluate the effectiveness of the method on different datasets and to perform validation comparisons with results reported in related works.

The GREYC-NISLAB dataset described in (Idrus et al. 2013) was collected in 2013 and includes five passwords entered by 110 users. The passwords are as follows: 1. "leonardo dicaprio" 2. "the rolling stones" 3. "michael schumacher" 4. "red hot chilli peppers" 5. "United States of America" (note: the spelling is as provided in the original data file). Each user entered five different passwords ten times with both hands and ten times with one hand, depending on whether the user was left- or right-handed. The dataset of a single password consists of 2200 samples. In total, the dataset comprises 11,000 data samples corresponding to 110 users, with 20 samples per user. Each password has different keystroke



(a) leonardo dicaprio (b) the rolling stones (c) michael schumacher

(d) red hot chilli peppers (e) united states of america

Fig. 10 Image-based representations of distinct passwords of the same user, generated using the GAFMAT algorithm. Password data source: GREYC-NISLAB dataset

patterns, so the number of keystroke dynamics features ranges from 64 to 92. Using the GAFMAT method, each password was transformed into the corresponding graphical representations, resulting in password images (see Fig. 10).

The experiments were carried out according to the procedures described in Sect. 5.2. The last five users from each password set were selected for testing. The final 2100 samples from each password dataset were split at a 70:30 ratio into training and validation sets. The results obtained from the validation data were very similar to those from the CMU dataset and are presented in Table 8. The results were evaluated according to the metrics described in Sect. 5.1.

As shown in Table 8, the network could classify each password with an average accuracy of 0.98. Notably, the highest accuracy was achieved for the passwords "United States of America" and "michael schumacher", both with a high accuracy of 0.99. Using the Euclidean distances between the anchor and the positive sample (AP_ED), as well as between the anchor and the negative sample (AN_ED), the network was able to effectively detect differences between the positive and negative samples with respect to the anchor. As a result, the triplet loss function resulted in a decrease in the distance between the anchor and the positive samples to a range of 0.39–0.45 and an increase in the distance between the anchor and the negative samples to a range of 1.48–1.63. These metrics highlight the crucial difference between positive and negative samples in metric space. These empirical results underscore the effectiveness of our choice of a 0.5 margin (Budžys et al. 2023).

The standard deviation of the distance between the anchor and the positive samples is approximately 0.2, and that between the anchor and the negative samples is approximately 0.39. This indicates that the network tended to admit more positive samples than negative samples, as the positive samples were twice as dispersed compared to the mean.

Another metric for quality evaluation, cosine similarity, was effective in distinguishing between positive and negative samples in relation to the anchor. The cosine similarity indicates that the positive sample is oriented in one direction relative to the anchor, with a value of approximately 0.78. Conversely, the negative samples are oriented in the opposite direction and have a value close to 0.5 relative to the anchor sample.

Table 8 Results using different accuracy metrics for passwords from GREYC-NISLAB on a validation dataset when transforming time series features of keystroke dynamics into an image using the GAFMAT algorithm

Metrics	Passwords (GREYC-NISLAB)				
	Leonardo dicaprio	The rolling stones	Michaell schumacher	Red hot chilli peppers	United States of America
Accuracy↑	0.97656	0.98698	0.99219	0.97778	0.99220
EER↓	0.07552	0.04688	0.0651	0.04444	0.04688
AUC↑	0.97824	0.98667	0.98771	0.98272	0.98847
AP_ED↓	0.44736	0.43986	0.39958	0.45165	0.39566
AN_ED↑	1.55644	1.61202	1.48864	1.63478	1.61275
AP_STD↓	0.24318	0.21992	0.20467	0.21505	0.19676
AN_STD↓	0.40601	0.37381	0.38351	0.38917	0.38013
AN_CS↓	0.49905	0.48703	0.52795	0.47839	0.49790
AP_CS↑	0.77632	0.78007	0.80021	0.77417	0.80217

The most important indicator for validating the proposed GAFMAT method is the EER. For three specific passwords (“the rolling stones”, “red hot chilli peppers”, “United States of America”), the EER value varied by approximately 0.045. Moreover, for the “leonardo dicaprio” and “œmichael schumacher” passwords, the EER values are 0.07552 and 0.0651, respectively. Obviously, in the case of the three passwords, the proposed methodology and approach provided an almost similar EER to the CMU dataset. However, it is important to note that the sample sizes of the datasets are different. The CMU dataset contains 400 instances of the same password for each user, while the GREYC-NISLAB dataset has only 20 samples for each user.

After obtaining the validation results, an experiment was further conducted on the test dataset to determine whether the password length would yield better results on unseen data. Prior to the training phase, a subset of five users was selected from each password dataset, allocating 100 samples for testing each individual password. After the training process, during which the optimal values of the weights were stored, the network was initialized with these parameters. The results of the evaluation using the test unseen data corresponding to the five users mentioned above are summarized in Table 9. The analysis shows that the results for the test data have the same trend as the results for the validation data, although their values have decreased. As shown in Table 9, the accuracy decreased to approximately 0.85. The Euclidean distances between the anchor and the positive samples increased, ranging from 0.67 to 0.87. In contrast, the distances between the anchor and negative samples remained almost the same as those in the validation data (see Table 8). Such observations suggest that even when assessing the quality using test data, the network retains the ability to distinguish between positive and negative samples compared to the anchor. This trend is also observed for the standard deviation. While the AN_STD remains the same as that for the validation dataset, remaining in the range of 0.4–0.49, the AP_STD decreases by almost half compared to the Euclidean anchor-positive distance (AP_ED).

Since the objective in our case is to minimize the EER, we consider this indicator as a baseline, which in the analysis of the GREYC-NISLAB dataset ranges between 0.14 and 0.22 for the test data, as shown in Table 9. The user authentication paradigm of the network is formulated in such a way that it can compare a newly entered password, transformed

Table 9 Results using different accuracy metrics for passwords from GREYC-NISLAB on a test dataset when transforming time series features of keystroke dynamics into an image using the GAFMAT algorithm

Metrics	Passwords (GREYC-NISLAB)				
	Leonardo dicaprio	The rolling stones	Michaell schumacher	Red hot chilli peppers	United States of America
Accuracy↑	0.84000	0.86000	0.86000	0.84000	0.92000
EER↓	0.16000	0.20000	0.22000	0.22000	0.14000
AUC↑	0.90320	0.85920	0.85400	0.86680	0.89240
AP_ED↓	0.78894	0.86642	0.67407	0.87670	0.75085
AN_ED↑	1.55808	1.49985	1.33055	1.55131	1.50073
AP_STD↓	0.41371	0.40861	0.31141	0.44201	0.43587
AN_STD↓	0.40956	0.41111	0.49554	0.40963	0.42794
AN_CS↓	0.41324	0.40843	0.49884	0.39300	0.43711
AP_CS↑	0.60553	0.56679	0.66297	0.56165	0.62458

according to the GAFMAT technique, with previous entries, aiming to achieve an EER close to zero. Currently, an EER of approximately 0.2 is observed, which indicates that improvements are necessary. To summarize, the SNN with a triplet loss function is able to distinguish between positive and negative samples in the test data. However, the obtained accuracy values are definitely lower than those of the validation data.

The observed EER values indicate that the accuracy of this metric is affected by the password length. This is supported by the fact that the CMU dataset contains 31 features, and the GREYC-NISLAB dataset contains 64 to 92 password features. In particular, the EER for the password "United States of America", which is the longest in the set with 92 features, was 0.14 (see Table 9). The EER of the next extended password, "red hot chilli peppers", with 84 features, was 0.22. These observations suggest that the EER is influenced mostly by the password's inherent features rather than its length.

5.4 General results

The following are the empirical results obtained on the CMU and GREYC-NISLAB datasets to evaluate the performance of the proposed GAFMAT algorithm. The comparative analysis with other studies is performed by comparing the EER results of the validation data from the CMU dataset with those reported in the previous literature on multiclass identification (see Table 10). It should be noted that many published papers report results based mainly on validation data. Therefore, our comparative analysis with other studies is performed using the EER results on the validation data from the CMU dataset. The EER results on the test data of the GREYC-NISLAB dataset are used for comparative analysis with a recent study on user authentication (Piugie et al. 2022).

Table 10 presents a focused performance evaluation of different authentication methods (see Table 1) using the CMU dataset. It aims to emphasize advances in EER reduction on CMU data. The method based on the Manhattan distance (scaled) reported by Killourhy and Maxion (2009) showed an EER of 0.096, indicating less efficiency in balancing false acceptances and false rejections. In contrast, the nearest neighbors method with a new distance metric, as explored by Zhong et al. (2012), showed an EER of 0.084 with outlier removal and 0.087 without it. Similarly, the inductive transfer encoder approach, applied by Monaco and Vindiola (2016), resulted in an EER of 0.063, which, although closer to the result of the GAFMAT method, remains less optimal. The CNN used by Çeker and Upadhyaya (2017) achieved

Table 10 Performance evaluation for CMU dataset passwords on validation data: a comparison of results in terms of EER values

References	Method	EER
This Paper	GAFMAT	0.04545
Killourhy and Maxion (2009) (original)	Manhattan distance (scaled)	0.09600
Zhong et al. (2012)	Nearest neighbor (new distance metric) + outlier removal	0.08400
Zhong et al. (2012)	Nearest neighbor (new distance metric)	0.08700
Monaco and Vindiola (2016)	Inductive transfer encoder (Manhattan distance)	0.06300
Çeker and Upadhyaya (2017)	CNN	0.06500
Ivannikova et al. (2017)	Dependence clustering with Manhattan distance	0.07700
Sae-Bae and Memon (2022)	Manhattan distance (scaled with standard deviation)	0.09160

an EER of 0.065, indicating fairly good performance. In addition, methods such as dependency clustering with Manhattan distance (Ivannikova et al. 2017) and Manhattan distance (with standard deviation scaling) (Sae-Bae and Memon 2022) showed EERs of 0.077 and 0.0916, respectively, indicating lower and insufficient authentication accuracy. This comparative analysis clearly indicates that the GAFMAT method significantly outperforms existing methods in terms of authentication accuracy, as evidenced by its significantly lower EER in the context of the CMU dataset. The results highlight the potential of the GAFMAT method for more accurate and reliable user authentication in cybersecurity applications. The performance of the proposed method was specifically compared to that of studies that used the complete sample set of the CMU dataset without excluding outliers, in contrast to a previous study (Monaco and Vindiola 2016) in which outliers were removed, resulting in an EER of 0.047, but an overall EER of 0.063. In the paper (Sae-Bae and Memon 2022), the highest EER of 0.045 was obtained, but these results are only for "good" users. The authors of the paper set the FAR threshold and calculated what the EER would be for "good", "average", and "bad" users. Despite these results, the average EER of 0.0916 of all users was taken and compared with the results obtained by the methods presented in this paper.

The results obtained on the CMU dataset indicate that transforming the numerical values into images using techniques such as GADF, GASF, and RP resulted in EER values of 0.04794, 0.0554, and 0.05327, respectively (see Table 8). These findings highlight the effectiveness of our proposed approach for transforming passwords into images for training the SNN, which improved the performance over previous state-of-the-art methods. Significantly, our proposed method for converting numerical data into images, called GAFMAT, achieved an improved EER value of 0.04545 (see Table 10).

A comparative analysis of the validation and test results of the GREYC-NISLAB dataset was carried out in this study, with particular regard to the evaluation of their performance in terms of EER and accuracy. This choice was made because recent research on this dataset of user authentication tasks has focused on improving accuracy and achieving better EER values (Piugie et al. 2022). Therefore, our study aimed to compare our results with these established benchmarks.

The information in Table 9 allows comparison of our results with those of other authors (Piugie et al. 2022). As indicated in this study, the best results for EER using the GoogleNet model were 0.1843 for "leonardo dicaprio", 0.1423 for "michaell schumacher", and 0.148 for "United States of America". Meanwhile, our proposed methodology with the implemented GAFMAT method achieved EER values of 0.16, 0.22, and 0.14, respectively. Notably, our implementation of a 12-layer CNN, while not as deep as the 22-layer deep neural network (GoogleNet), yielded results on the test dataset that are comparable to or slightly better than the network containing almost twice as many layers.

The results obtained on CMU data clearly demonstrate that the proposed GAFMAT method combined with an SNN achieves significantly lower EERs than do existing methods such as GADF, GASF, MTF and RP. The method achieved an EER of 0.04545 on the CMU dataset. In addition, the method achieved a high level of accuracy for the GREYC-NISLAB dataset, with EERs ranging from 0.04444 to 0.07552. The findings emphasize the remarkable performance of the proposed solution to distinguish genuine users from impostors.

6 Challenges and future research directions

Existing methods for detecting insider threats often rely on a combination of user behavior monitoring, anomaly detection, and traditional authentication methods. Our proposed methodology, which utilizes the GAFMAT method and an SNN with CNN branches, primarily aims to enhance user authentication through the analysis of keystroke dynamics. This approach is more specific than general monitoring of user behavior, as it focuses on the nuances of typing patterns to detect anomalies. Unlike many existing systems that may require additional hardware for biometric authentication, our system utilizes existing keyboard inputs. This not only makes it a cost-effective solution, but also allows for easy integration into the current infrastructure without the need for significant modifications or upgrades.

6.1 Strengths and limitations

The main strength of our methodology is that it uses only a keyboard, making it easy to integrate into existing systems. Empirical studies conducted using the publicly available CMU and GREYC-NISLAB datasets confirm the effectiveness of the approach, as demonstrated by a reduction in EER and an enhancement in user authentication accuracy. This improvement in results was influenced by transforming the keystroke dynamics patterns into images, which are then utilized to train SNNs through supervised learning.

However, the applicability of the proposed methodology to different datasets or real-world scenarios remains to be thoroughly tested. Experimental observations revealed that the EER is influenced by both the length and the complexity of the passwords. It is very important to study how different password lengths and complexities affect the model performance, considering the relevance to real-world applications that operate under varying password policies and within dynamic authentication environments. The performance of the model can also be affected by changes in user behavior over time or variations in typing due to environmental factors (e.g., different keyboards or physical conditions).

6.2 Future work

To address the limitations of relying on supervised learning, to increase model adaptability, and to reduce dependence on labeled datasets, it is appropriate to incorporate unsupervised learning techniques in future enhancements. For example, clustering methods can generate representative triplets regardless of whether the data are labeled or not, potentially leading to more generalizable models. Another key area of research is to extend the scope of the system beyond critical infrastructure to other areas that require secure user authentication.

The system is planned to be integrated into real existing authentication mechanisms in critical infrastructure. For new users, the initial step involves the collection of keystroke dynamics data by capturing timestamps during password entry. These data are then transformed into images using the GAFMAT method and securely stored in a database. When a login attempt is made, the keystroke dynamics of the entered password are immediately transformed into an image using GAFMAT. The transformed image is fed to the trained SNN. The network processes the image to generate an embedding, which is a vector representation of the image in a multidimensional space. The embedding is then compared to the embeddings of previously stored password images. This comparison involves computing the distance between the embeddings in the multidimensional feature space. A critical aspect of the system is determining a threshold to distinguish legitimate attempts from impostor attempts. Initially, a threshold baseline is determined based on historical data of genuine login attempts. This baseline represents the typical range of distances observed in embeddings of authentic password entries. Decisions on the legitimacy of a login attempt are based on the proximity of the input embedding to stored embeddings. If the distance falls within the predefined threshold, the login attempt is considered genuine. To detect insiders, the system must monitor discrepancies between current keystroke dynamics and stored password samples. Significant deviations may trigger additional security measures, such as multifactor authentication or the notification of security personnel.

7 Conclusions

This paper proposes a comprehensive artificial intelligence-based user authentication methodology for insider threat detection based on behavioral biometrics and deep neural networks to improve the effectiveness of intrusion detection systems and intrusion prevention systems. The need for such a system in the modern era, where war and cyberattacks are massive, cannot be overemphasized. Recently, potential threats have been observed where unauthorized individuals gain access to devices and accounts in critical infrastructure, causing significant damage.

From a theoretical point of view, this study extends the understanding of behavioral biometrics in the field of cybersecurity. It highlights the effectiveness of transforming keystroke dynamics into images, which provides a new perspective in the field of biometric analysis. This study presents a static authentication process that uses the unique timing characteristics of a user's keystrokes to generate a password input pattern for further analysis. By comparing these patterns against a database of known legitimate password patterns, it is possible to determine whether the pattern entered matches the pattern of a known user. Thus, even if an unauthorized person learns the user's password, they still need to replicate the user's unique input behavior to gain access, providing an additional layer of security.

A key contribution of this study is the introduction of the GAFMAT method for transforming non-image or numerical data into images, which enhances the versatility and extends the application of data to deep learning tasks. This solution, combined with a Siamese Neural Network (SNN) and triplet loss function, has significant potential for improving the accuracy of user authentication using keystroke biometrics. This method of transforming numerical values into images has proven its effectiveness on both the CMU and GREYC-NISLAB datasets, as evidenced by its competitive and often better Equal Error Rate (EER) values compared to those of other state-of-the-art methods. A lower EER of 0.04545 and an average accuracy of 98.9% were obtained on the CMU dataset. This result highlights the accuracy of the method for distinguishing legitimate users from potential insider threats. Furthermore, the method demonstrated a high level of accuracy, averaging 98% on the GREYC-NISLAB dataset for different passwords and ranges of EER from 0.04444 to 0.07552, further confirming its effectiveness.

The practical application of the user authentication methodology, which combines the novel GAFMAT method and the SNN, addresses real-world cybersecurity problems in a trustworthy and effective way to distinguish legitimate attempts from impostor attempts. This research highlights the potential of integrating behavioral biometrics with deep learning techniques to improve security in today's digital age.

Although the results obtained are promising, they reveal the limitations of the applicability of the solution to different datasets and real-world scenarios. In particular, EER is sensitive to the length and features of the password, while factors such as password complexity, changes in user behavior, and environmental conditions (e.g., different keyboards) can have a significant impact on the performance of the solution. Additionally, our proposed methodology is based on supervised learning, so labeled data are needed to train the model.

To mitigate the limitations of supervised learning, future improvements should include unsupervised learning techniques to increase the applicability of the model. In addition, investigating the integration of this authentication methodology with existing security systems in critical infrastructure can provide valuable insights into the challenges and benefits of its practical implementation.

Author contributions Authors contributed equally to this work.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

Abadi M, Agarwal A, Barham P et al (2015) TensorFlow: large-scale machine learning on heterogeneous systems. https://www.tensorflow.org/softwareavailablefromtensorflow.org

Abdulrahman SA, Alhayani B (2023) A comprehensive survey on the biometric systems based on physiological and behavioural characteristics. Mater Today Proc 80:2642–2646

Abuhamad M, Abusnaina A, Nyang D et al (2020) Sensor-based continuous authentication of smartphones' users using behavioral biometrics: a contemporary survey. IEEE Internet Things J 8(1):65–84. https://doi.org/10.1109/JIOT.2020.3020076

Acien A, Morales A, Vera-Rodriguez R et al (2020) Typenet: scaling up keystroke biometrics. In: 2020 IEEE international joint conference on biometrics (IJCB). IEEE, pp 1–7, https://doi.org/10.1109/IJCB48548.2020.9304908

Alfoudi AS, Aziz MR, Alyasseri ZAA et al (2022) Hyper clustering model for dynamic network intrusion detection. IET Commun 2022:10

Al-Mhiqani MN, Ahmad R, Abidin ZZ et al (2022) A new intelligent multilayer framework for insider threat detection. Comput Electric Eng 97:107597

Ayotte B, Banavar M, Hou D et al (2020) Fast free-text authentication via instance-based keystroke dynamics. IEEE Trans Biometric Behav Identity Sci 2(4):377–387. https://doi.org/10.1109/TBIOM.2020.3003988

Azizan AH, Mostafa SA, Mustapha A et al (2021) A machine learning approach for improving the performance of network intrusion detection systems. Ann Emerg Technol Comput 5(5):201–208

Basit A, Zafar M, Liu X et al (2021) A comprehensive survey of AI-enabled phishing attacks detection techniques. Telecommun Syst 76:139–154

Bicakci K, Salman O, Uzunay Y et al (2020) Analysis and evaluation of keystroke dynamics as a feature of contextual authentication. In: 2020 international conference on information security and cryptology (ISCTURKEY). IEEE, pp 11–17, https://doi.org/10.1109/ISCTURKEY51113.2020.9307967

Bromley J, Guyon I, LeCun Y et al (1993) Signature verification using a "Siamese" time delay neural network. Adv Neural Info Process Syst. https://doi.org/10.1142/s0218001493000339

Budžys A, Kurasova O, Medvedev V (2023) Behavioral biometrics authentication in critical infrastructure using siamese neural networks. In: HCI for cybersecurity, privacy and trust, LNCS. pp 1–14, https://doi.org/10.1007/978-3-031-35822-7_21

Çeker H, Upadhyaya S (2017) Sensitivity analysis in keystroke dynamics using convolutional neural networks. In: 2017 IEEE workshop on information forensics and security (WIFS). IEEE, pp 1–6,https://doi.org/10.1109/WIFS.2017.8267667

Chen CB, Yang H, Kumara S (2018) Recurrence network modeling and analysis of spatial data. Chaos Interdisc J Nonlinear Sci. https://doi.org/10.1063/1.5024917

Cheng D, Gong Y, Zhou S et al (2016) Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In: Proceedings of the iEEE conference on computer vision and pattern recognition, pp 1335–1344, https://doi.org/10.1109/CVPR.2016.149

Dias D, Pinto A, Dias U et al (2020) A multirepresentational fusion of time series for pixelwise classification. IEEE J Select Topics Appl Earth Observ Remote Sens 13:4399–4409. https://doi.org/10.1109/JSTARS.2020.3012117

Ding S, Lin L, Wang G et al (2015) Deep feature learning with relative distance comparison for person re-identification. Pattern Recogn 48(10):2993–3003. https://doi.org/10.1016/j.patcog.2015.04.005

Dong X, Shen J (2018) Triplet loss in siamese network for object tracking. In: Proceedings of the European conference on computer vision (ECCV), pp 459–474, https://doi.org/10.1007/978-3-030-01261-8_28

Elliot K, Graham J, Yassin Y et al (2019) A comparison of machine learning algorithms in keystroke dynamics. In: 2019 international conference on computational science and computational intelligence (CSCI). IEEE, pp 127–132,https://doi.org/10.1109/CSCI49370.2019.00028

Estebsari A, Rajabi R (2020) Single residential load forecasting using deep learning and image encoding techniques. Electronics 9(1):68. https://doi.org/10.3390/electronics9010068

Federal Bureau of Investigation (2023) Internet crime report 2022. https://www.ic3.gov/Media/PDF/AnnualReport/2022_IC3Report.pdf

Gedikli AM, Efe MÖ (2020) A simple authentication method with multilayer feedforward neural network using keystroke dynamics. In: Pattern recognition and artificial intelligence: third Mediterranean conference, MedPRAI 2019, Istanbul, Turkey, December 22–23, 2019, Proceedings 3. Springer, pp 9–23, https://doi.org/10.1007/978-3-030-37548-5_2

Giancardo L, Sánchez-Ferro A, Butterworth I et al (2015) Psychomotor impairment detection via finger interactions with a computer keyboard during natural typing. Sci Rep 5(1):1–8. https://doi.org/10.1038/srep09678

Giot R, Dorizzi B, Rosenberger C (2015) A review on the public benchmark databases for static keystroke dynamics. Comput Secur 55:46–61. https://doi.org/10.1016/j.cose.2015.06.008

Gofman MI, Villa M (2023) Identity and war: the role of biometrics in the Russia-Ukraine crisis. Int J Eng Sci Technol 5(1):2

Hadsell R, Chopra S, LeCun Y (2006) Dimensionality reduction by learning an invariant mapping. In: 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06). IEEE, pp 1735–1742, https://doi.org/10.1109/CVPR.2006.100

Idrus SZS, Cherrier E, Rosenberger C et al (2013) Soft biometrics database: A benchmark for keystroke dynamics biometric systems. In: 2013 international conference of the BIOSIG special interest group (BIOSIG). IEEE, pp 1–8

Imamura A, Arizumi N (2021) Gabor filter incorporated cnn for compression. In: 2021 36th international conference on image and vision computing New Zealand (IVCNZ). IEEE, pp 1–5, https://doi.org/10.1109/IVCNZ54163.2021.9653342

Ivannikova E, David G, Hämäläinen T (2017) Anomaly detection approach to keystroke dynamics based user authentication. In: 2017 IEEE symposium on computers and communications (ISCC). IEEE, pp 885–889, https://doi.org/10.1109/ISCC.2017.8024638

Jain AK, Gupta B (2022) A survey of phishing attack techniques, defence mechanisms and open research challenges. Enterp Info Syst 16(4):527–565

Kamarainen JK, Kyrki V, Kalviainen H (2006) Invariance properties of gabor filter-based features-overview and applications. IEEE Trans Image Process 15(5):1088–1099. https://doi.org/10.1109/TIP.2005.864174

Kaur R, Gabrijelčič D, Klobučar T (2023) Artificial intelligence for cybersecurity: literature review and future research directions. Info Fusion. https://doi.org/10.1016/j.inffus.2023.101804

Killourhy KS, Maxion RA (2009) Comparing anomaly-detection algorithms for keystroke dynamics. In: 2009 IEEE/IFIP international conference on dependable systems & networks. IEEE, pp 125–134, https://doi.org/10.1109/DSN.2009.5270346

Koch G, Zemel R, Salakhutdinov R et al (2015) Siamese neural networks for one-shot image recognition. In: ICML deep learning workshop, Lille

Krishna GJ, Jaiswal H, Teja PSR et al (2019) Keystroke based user identification with XGBoost. In: TENCON 2019-2019 IEEE region 10 conference (TENCON). IEEE, pp 1369–1374, https://doi.org/10.1109/TENCON.2019.8929453

Krishnamoorthy S, Rueda L, Saad S et al (2018) Identification of user behavioral biometrics for authentication using keystroke dynamics and machine learning. In: Proceedings of the 2018 2nd international conference on biometric engineering and applications, pp 50–57, https://doi.org/10.1145/3230820.3230829

Krombholz K, Hobel H, Huber M et al (2015) Advanced social engineering attacks. J Info Secur Appl 22:113–122

Liang Y, Samtani S, Guo B et al (2020) Behavioral biometrics for continuous authentication in the internet-of-things era: an artificial intelligence perspective. IEEE Internet Things J 7(9):9128–9143. https://doi.org/10.1109/JIOT.2020.3004077

Liu M, Guan J (2019) User keystroke authentication based on convolutional neural network. In: Mobile internet security: second international symposium, MobiSec 2017, Jeju Island, Republic of Korea, October 19–22, 2017, Revised Selected Papers 2. Springer, pp 157–168, https://doi.org/10.1007/978-981-13-3732-1_13

Lu X, Zhang S, Hui P et al (2020) Continuous authentication by free-text keystroke based on cnn and rnn. Comput Secur 96:101861. https://doi.org/10.1016/j.cose.2020.101861

Manthiramoorthy C, Khan KMS et al (2024) Comparing several encrypted cloud storage platforms. Int J Math Stat Comput Sci 2:44–62

Martín AG, Beltrán M, Fernández-Isabel A et al (2021) An approach to detect user behaviour anomalies within identity federations. Comput Secur 108:102356. https://doi.org/10.1016/j.cose.2021.102356

Mattera M, Chowdhury MM (2021) Social engineering: the looming threat. In: 2021 IEEE international conference on electro information technology (EIT). IEEE, pp 056–061

Medvedev V, Budžys A, Kurasova O (2023) Enhancing keystroke biometric authentication using deep learning techniques. In: 2023 18th Iberian Conference on Information Systems and Technologies (CISTI). pp 1–6, https://doi.org/10.23919/CISTI58278.2023.10211344

Melekhov I, Kannala J, Rahtu E (2016) Siamese network features for image matching. In: 2016 23rd international conference on pattern recognition (ICPR). IEEE, pp 378–383

Mohamed N (2023) Current trends in AI and ML for cybersecurity: a state-of-the-art survey. Cogent Eng 10(2):2272358. https://doi.org/10.1080/23311916.2023.2272358

Monaco JV, Vindiola MM (2016) Crossing domains with the inductive transfer encoder: Case study in keystroke biometrics. In: 2016 IEEE 8th international conference on biometrics theory, applications and systems (BTAS). IEEE, pp 1–8, https://doi.org/10.1109/BTAS.2016.7791165

Muliono Y, Ham H, Darmawan D (2018) Keystroke dynamic classification using machine learning for password authorization. Proc Comput Sci 135:564–569. https://doi.org/10.1016/j.procs.2018.08.209

Neves J, Narducci F, Barra S et al (2016) Biometric recognition in surveillance scenarios: a survey. Artif Intell Rev 46:515–541. https://doi.org/10.1007/s10462-016-9474-x

Ondrašovič M, Tarábek P (2021) Siamese visual object tracking: a survey. IEEE Access 9:110149–110172. https://doi.org/10.1109/ACCESS.2021.3101988

Piugie YBW, Di Manno J, Rosenberger C et al (2022) Keystroke dynamics based user authentication using deep learning neural networks. In: 2022 international conference on cyberworlds (CW), IEEE, pp 220–227, https://doi.org/10.1109/CW55638.2022.00052

Rajkumar VS, Ştefanov A, Presekal A et al (2023) Cyber attacks on power grids: causes and propagation of cascading failures. IEEE Access 11:103154–103176. https://doi.org/10.1109/ACCESS.2023.3317695

Roy S, Pradhan J, Kumar A et al (2022) A systematic literature review on latest keystroke dynamics based models. IEEE Access. https://doi.org/10.1109/ACCESS.2022.3197756

Sae-Bae N, Memon N (2022) Distinguishability of keystroke dynamic template. PLoS ONE 17(1):e0261291. https://doi.org/10.1371/journal.pone.0261291

Sandhya M, Morampudi MK, Pruthweraaj I et al (2022) Multi-instance cancelable iris authentication system using triplet loss for deep learning models. V Comput. https://doi.org/10.1007/s00371-022-02429-x

Schroff F, Kalenichenko D, Philbin J (2015) Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 815–823, https://doi.org/10.1109/CVPR.2015.7298682

Serwadda A, Phoha VV (2013) Examining a large keystroke biometrics dataset for statistical-attack openings. ACM Trans Info Syst Secu 16(2):1–30. https://doi.org/10.1145/2516960

Sharma A, Vans E, Shigemizu D et al (2019) Deepinsight: a methodology to transform a non-image data to an image for convolution neural network architecture. Sci Rep 9(1):11399

Shekhawat K, Bhatt DP (2019) Recent advances and applications of keystroke dynamics. In: 2019 international conference on computational intelligence and knowledge economy (ICCIKE). IEEE, pp 680–683, https://doi.org/10.1109/ICCIKE47802.2019.9004312

Siam AI, Sedik A, El-Shafai W et al (2021) Biosignal classification for human identification based on convolutional neural networks. Int J Commun Syst 34(7):e4685. https://doi.org/10.1002/dac.4685

Singh S, Inamdar A, Kore A et al (2020) Analysis of algorithms for user authentication using keystroke dynamics. In: 2020 international conference on communication and signal processing (ICCSP). IEEE, pp 0337–0341, https://doi.org/10.1109/ICCSP48568.2020.9182115

Stanciu VD, Spolaor R, Conti M et al (2016) On the effectiveness of sensor-enhanced keystroke dynamics against statistical attacks. In: Proceedings of the sixth ACM conference on data and application security and privacy, pp 105–112, https://doi.org/10.1145/2857705.2857748

Tao X, Zhang D, Ma W et al (2022) Unsupervised anomaly detection for surface defects with dual-Siamese network. IEEE Trans Ind Info 18(11):7707–7717. https://doi.org/10.1109/TII.2022.3142326

Ugwuoke CO, Eze OJ, Ameh SO et al (2021) Armed robbery attacks and everyday life in Nigeria. Int J Crim Justice Sci 16(1):186–200

Valero-Mas JJ, Gallego AJ, Rico-Juan JR (2023) An overview of ensemble and feature learning in few-shot image classification using siamese networks. Multimedia Tools Appl 2023:1–24

Verizon (2022) Data breach investigation report 2022. https://www.verizon.com/business/en-gb/resources/2022-data-breach-investigations-report-dbir.pdf

Wang Z, Oates T (2015) Imaging time-series to improve classification and imputation. In: Proceedings of the 24th international conference on artificial intelligence, pp 3939–3945

William I, Rachmawanto EH, Santoso HA et al (2019) Face recognition using facenet (survey, performance test, and comparison). In: 2019 fourth international conference on informatics and computing (ICIC). IEEE, pp 1–6

Yan C, Pang G, Bai X et al (2021) Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss. IEEE Trans Multimedia 24:1665–1677. https://doi.org/10.1109/TMM.2021.3069562

Zaidan D, Salem A, Swidan A et al (2017) Factors affecting keystroke dynamics for verification data collecting and analysis. 2017 8th international conference on information technology (ICIT). IEEE, New York, pp 392–398. https://doi.org/10.1109/ICITECH.2017.8080032

Zhang Y, Hou Y, Zhou S et al (2020) Encoding time series as multi-scale signed recurrence plots for classification using fully convolutional networks. Sensors 20(14):3818. https://doi.org/10.3390/s20143818

Zhang Z, Ning H, Shi F et al (2022) Artificial intelligence in cyber security: research advances, challenges, and opportunities. Artif Intell Rev. https://doi.org/10.1007/s10462-021-09976-0

Zhao X, Sun H, Lin B et al (2022) Markov transition fields and deep learning-based event-classification and vibration-frequency measurement for φ-otdr. IEEE Sens J 22(4):3348–3357. https://doi.org/10.1109/JSEN.2021.3137006

Zhong Y, Deng Y, Jain AK (2012) Keystroke dynamics for user authentication. 2012 IEEE computer society conference on computer vision and pattern recognition workshops. IEEE, New York, pp 117–123. https://doi.org/10.1109/CVPRW.2012.6239225

Zhou X, Liang W, Shimizu S et al (2020) Siamese neural network based few-shot learning for anomaly detection in industrial cyber-physical systems. IEEE Trans Industr Inf 17(8):5790–5798. https://doi.org/10.1109/TII.2020.3047675

Zhu Y, Brettin T, Xia F et al (2021) Converting tabular data into images for deep learning with convolutional neural networks. Sci Rep 11(1):11325. https://doi.org/10.1038/s41598-021-90923-y

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Arnoldas Budžys[1] · **Olga Kurasova**[1] · **Viktor Medvedev**[1]

✉ Arnoldas Budžys
arnoldas.budzys@mif.stud.vu.lt

Olga Kurasova
olga.kurasova@mif.vu.lt

Viktor Medvedev
viktor.medvedev@mif.vu.lt

[1] Institute of Data Science and Digital Technologies, Vilnius University, Akademijos Str. 4, 08412 Vilnius, Lithuania