IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AUTOMATIC MALWARE SIGNATURE CLASSIFICATION USING DEEP LEARNING

K.Suga Priya[1], Mrs.P.Jasmine Lois Ebenazer[2], Mrs.T.Usha Felshia[3] and Mrs.E.Julie Ruth[4]

Department of Computer Applications, Sarah Tucker College, Thirunelveli-7.

ABSTRACT: Deep learning has advanced to the point that it has surpassed old handmade methodologies and even humans for a variety of jobs in recent years. However, the amount of publicly available data for some tasks, such as the verification of handwritten signatures, is limited, making it difficult to verify the true limits of deep learning. Aside from the absence of publicly available data, evaluating the improvements of novel proposed methodologies is difficult due to the use of various databases and experimental protocols. The study's main contributions are I an in-depth analysis of state-of-the-art deep learning approaches for online signature verification; ii) the presentation and description of the new DeepSignDB online handwritten signature biometric public database; iii) the proposal of a standard experimental protocol and benchmark to be used by the research community to perform a fair comparison of novel approaches with the state-of-the-art and iv) we adopt and analyze Time-Aligned Recurrent Neural Networks (TA-RNNs), a recent deep learning approach for on-line handwritten signature verification. To train more resilient systems against forgeries, this approach combines the promise of Dynamic Time Warping with Recurrent Neural Networks. When considering experienced forging impostors and only one training signature per user, our suggested TA-RNN system beats the current state of the art, attaining outcomes even below 2.0 percent EER.

1. INTRODUCTION

Handwritten signature verification is still a hot topic in academia today. Depending on the type of acquisition, it can be classified as: i) off-line, where the signature is captured using a traditional method of signing with an ink pen on paper and then digitising the image; or ii) on-line, where the signature is captured using electronic devices, resulting in the capture of not only the image of the signature, but also the complete collecting process's signing information (time sequences). In the previous 40 years, online handwritten signature verification has progressed significantly. From the first Wacom devices, which were built expressly to capture handwriting and signatures in office-like circumstances, to today's mobile acquisition scenarios, when signatures can be taken using our own personal smartphone anywhere, we've come a long way.

Despite advancements in acquisition technology, classic methodologies such as Dynamic Time Warping (DTW), Hidden Markov Models (HMM), and Support Vector Machines remain at the foundation of most state-of-the-art signature verification systems (SVM). When compared to other biometric traits like face and fingerprint, where deep learning has beaten traditional approaches by a wide margin, and even tasks more closely related to signature verification like handwriting recognition, writer verification, and handwritten passwords, this aspect appears to be a bit unusual. So, why aren't deep learning algorithms for online signature verification more frequently used? One significant disadvantage could be the time-consuming process of obtaining a large-scale database for training the models, as signatures are not publically available on the internet like other biometric features like the face.

Another key discovery that motivates this effort, in addition to the shortage of data for training deep learning systems, is that: because multiple experimental protocols and conditions are usually considered for different signature databases, there is a lack of a common experimental methodology to be used by the research community in order to undertake a fair comparison of novel techniques to the state of the art. In light of these issues, we introduce the new DeepSignDB handwritten signature biometric database, the world's largest on-line signature database to date, in this paper. The design, acquisition devices, and writing tools considered in the DeepSignDB database are visually summarised in Figure 1. Its applications range from improving signature verification systems using deep learning to a variety of other prospective research areas, such as exploring: i) User-dependent effects in signature biometrics and handwriting recognition in general, as well as the development of user-dependent methodologies ii) signature biometrics and handwriting in general, as well as the neuromotor processes involved. iii) variables influencing the acquisition of representative and clear handwriting and touch interaction signals, iv) Handwriting and touchscreen signals are examples of human-device interaction factors, as well as the creation of improved interaction methods and v) demographic statistics relating to handwriting and touch contact signals, as well as the development of novel ways for recognising and serving certain population groups.

The following are the study's primary contributions:

- A detailed examination of current state-of-the-art deep learning algorithms for on-line signature verification, with emphasis on the many experimental protocol conditions studied.
- DeepSignDB, a new online database for handwritten signatures. This database was created by combining several well-known datasets with a previously unpublished dataset. It has almost 70K signatures collected from a total of 1526 individuals utilising both stylus and finger inputs. With a total of 8 different devices, two acquisition scenarios, office and mobile, are investigated. The number of acquisition sessions and the types of impostors are also taken into account.

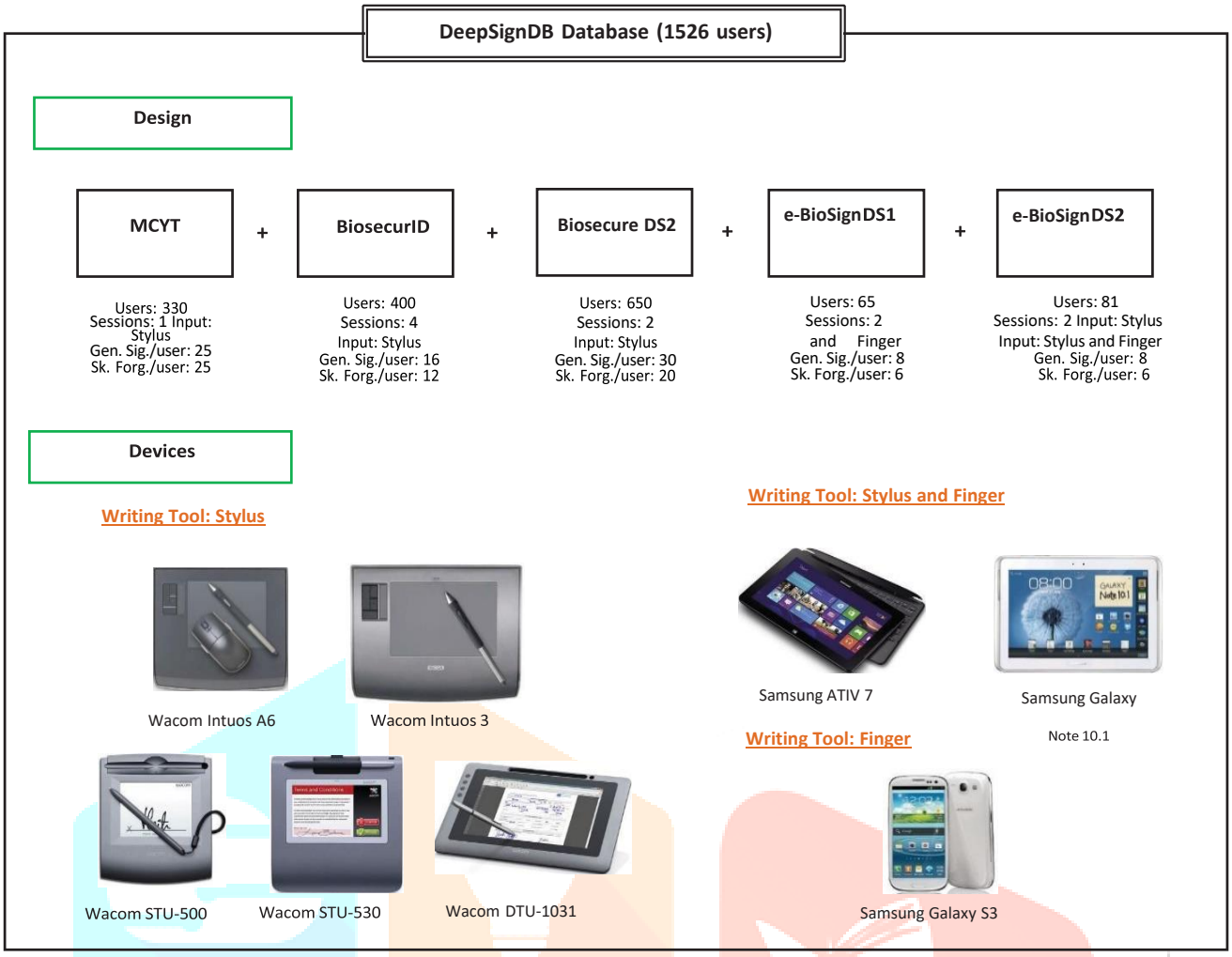

Fig. 1: A description of the new DeepSignDB database's design, acquisition devices, and writing tools. There are 1526 users and 8 distinct recorded devices in all (5 Wacom and 3 Samsung general-purpose devices). Signatures are also gathered using the finger on Samsung devices. Genuine Signatures (Gen. Sig.) and Skilled Forgeries (Sk. Forg.).

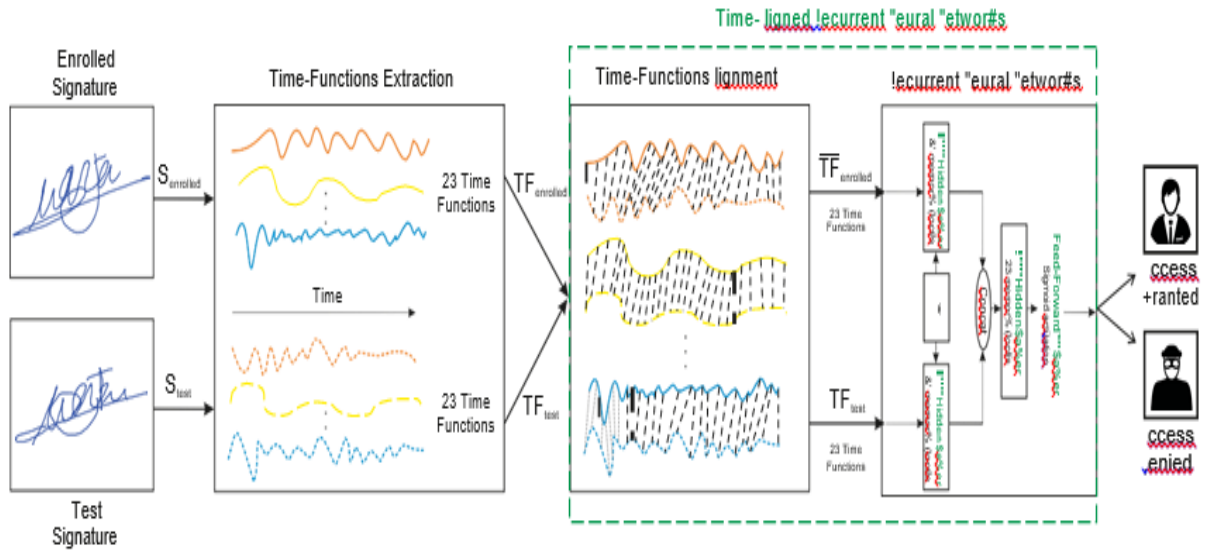

Fig. 2: Time-Aligned Recurrent Neural Networks are used to build our proposed on-line signature verification system. TF and TF signify the original and pre-aligned 23 time functions, respectively, and S denotes one signature sample.

- A standard experimental methodology that is freely available to the research community in order to compare novel approaches to the state of the art. As a result, we additionally make available the files containing all of the signature comparisons performed on the final assessment dataset. As a result, we've created a structure that's simple to replicate.
- An adaptation and evaluation of our latest deep learning system for on-line handwritten signature verification, called Time-Aligned Recurrent Neural Networks (TA-RNNs). This method was first proposed in for fingerprint biometrics on touchscreens. The architecture of our suggested strategy is depicted in Figure 2. It combines the strengths of Dynamic Time Warping and Recurrent Neural Networks (RNNs) to create more resilient systems that can detect forgeries.
- A comparison of DeepSignDB to well-known systems based on DTW, RNNs, and our recently suggested TA-RNNs.

This essay was first published in a paper form. The following aspects of this article have significantly improved: i) We examine state-of-the-art deep learning techniques for on-line signature verification in depth, ii) Our current TA-RNN deep learning technique is adapted and evaluated, iii) We conduct a more thorough evaluation of DeepSignDB, examining system performance for each DeepSignDB scenario and dataset, as well as for DTW, RNNs, and our proposed TA-RNNs, and iv) Our suggested TA-RNN method surpasses existing signature verification methods, emphasising the importance of aligning time-functions.

The remainder of the article is carried out as follows. Sec. II summarises past deep learning-based on-line signature verification research. Sec. III delves into the specifics of our suggested TA-RNN strategy. The DeepSignDB signature database is described in depth in Section IV. Sec. V explains the planned experimental technique as well as the benchmark evaluation that was performed. Sec. VI derives the concluding conclusions and identifies some areas for future research.

2. LITERATURE SURVEY

In this work, we have analyzed how the latest advances in keystroke biometric recognition can help to link behavioral typing patterns in experiments involving 100,000 users and more than 1 million typed sequences. Our proposed system is based on Recurrent Neural Networks adapted to the context of content de-anonymization. Assuming the challenge to link the typed content of a target user in a pool of candidate profiles, our results show that keystroke recognition can be used to reduce the list of candidate profiles by more than 90%. In addition, when keystroke is combined with auxiliary data (such as location), our system achieves a Rank-1 identification performance equal to 52.6% and 10.9% for a background candidate list composed of 1K and 100K profiles, respectively.

This work enhances traditional authentication systems based on Personal Identification Numbers (PIN) and One-Time Passwords (OTP) through the incorporation of biometric information as a second level of user authentication. In our proposed approach, users draw each digit of the password on the touchscreen of the device instead of typing them as usual. A complete analysis of our proposed biometric system is carried out regarding the discriminative power of each handwritten digit and the robustness when increasing the length of the password and the number of enrolment samples. The new e-BioDigit database, which comprises on-line handwritten digits from 0 to 9, has been acquired using the finger as input on a mobile device. This database is used in the experiments reported in this work and it is available together with benchmark results in GitHub.

The Kinematic Theory of rapid movements and its associated Sigma-Lognormal model have been extensively used in a large variety of applications. While the physical and biological meaning of the model have been widely tested and validated for rapid movements, some shortcomings have been detected when it is used with continuous long and complex movements. To alleviate such drawbacks, and inspired by the motor equivalence theory and a conceivable visual feedback, this paper proposes a novel framework to extract the Sigma-Lognormal parameters, namely iDeLog. Specifically, iDeLog consists of two steps. The first one, influenced by the motor equivalence model, separately derives an initial action plan defined by a set of virtual points and angles from the trajectory and a sequence of lognormals from the velocity. In the second step, based on a hypothetical visual feedback compatible with an open-loop motor control, the virtual target points of the action plan are iteratively moved to improve the matching between the observed and reconstructed trajectory and velocity. During experiments conducted with handwritten signatures, iDeLog obtained promising results as compared to the previous development of the Sigma-Lognormal.

This paper proposes a novel approach for on-line signature complexity detection based on Recurrent Neural Networks (RNNs). Complexity of handwritten signatures can vary from very simple ones (just a simple flourish) to very complex signatures (including the handwritten full name and complex flourish). Three different complexity levels are proposed: low, medium, and high. Time functions are extracted from the on-line signatures and a system based on RNNs (BLSTM in particular) is trained to classify the three levels of complexity over a ground truth manually labelled database (BiosecurID with 400 subjects). This initial model is used to automatically label a very large database (DeepSignDB) containing over 1500 subjects, which is then used to train the proposed RNN for signature complexity detection. Promising results ca. 85% of accuracy are achieved. This complexity detector could be used as a first stage in a signature verification system in order to train a specific biometric system per signature complexity level and improve the overall system performance.

In this paper, Legendre polynomials coefficients are used as features to model the signatures. The classifier used in this paper is deep feedforward neural network and the deep learning algorithm is stochastic gradient descent with momentum. The experimental results show better Equal Error Rule reduction and accuracy enhancement on SigComp2011 Dataset presented within ICDAR 2011 in comparison with state-of-the-art methods.

This framework optimizes a Siamese network with a local embedding loss, and learns a feature space that preserves the temporal location-wise distances between time series. To achieve invariance to non-linear temporal distortion, we propose building a dynamic time warping block on top of the Siamese network, which will greatly improve the accuracy for local correspondences across intra-personal variability. Validation with respect to online signature verification demonstrates the advantage of our framework over existing techniques that use either handcrafted or learned feature representations.

In this paper, we propose a novel stroke-based bidirectional RNN architecture. The main idea is to split the signature into multiple patches using strokes. Concatenation of query and reference signature pairs are used as input. The proposed method uses two LSTM RNN networks to extract different features. The first one extracts the features of the strokes and the latter extracts the global features of the whole signatures. The results on the BiosecureID dataset demonstrate that our proposed method can reduce the EER by 33.05%, from 5.6% to 3.75% with fewer features and less training samples. Besides, we find that the proposed stroke based RNN network is 5x faster in training and testing time than Non stroke-based RNN network.

In this work, we have analyzed how the latest advances in keystroke biometric recognition can help to link behavioral typing patterns in experiments involving 100,000 users and more than 1 million typed sequences. Our proposed system is based on Recurrent Neural Networks adapted to the context of content de-anonymization. Assuming the challenge to link the typed content of a target user in a pool of candidate profiles, our results show that keystroke recognition can be used to reduce the list of candidate profiles by more than 90%. In addition, when keystroke is combined with auxiliary data (such as location), our system achieves a Rank-1 identification performance equal to 52.6% and 10.9% for a background candidate list composed of 1K and 100K profiles, respectively.

Data have become one of the most valuable things in this new era where deep learning technology seems to overcome traditional approaches. However, in some tasks, such as the verification of handwritten signatures, the amount of publicly available data is scarce, what makes difficult to test the real limits of deep learning. In addition to the lack of public data, it is not easy to evaluate the improvements of novel approaches compared with the state of the art as different experimental protocols and conditions are usually considered for different signature databases. To tackle all these mentioned problems, the main contribution of this study is twofold: i) we present and describe the new DeepSignDB on-line handwritten signature biometric public database, and ii) we propose a standard experimental protocol and benchmark to be used for the research community in order to perform a fair comparison of novel approaches with the state of the art. The DeepSignDB database is obtained through the combination of some of the most popular on-line signature databases, and a novel dataset not presented yet. It comprises more than 70K signatures acquired using both stylus and finger inputs from a total of 1526 users. Two acquisition scenarios are considered, office and mobile, with a total of 8 different devices. Additionally, different types of impostors and number of acquisition sessions are considered along the database. The DeepSignDB and benchmark results are available in GitHub.

In this paper, we propose a novel writer-independent on-line signature verification systems based on Recurrent Neural Networks (RNNs) with a Siamese architecture whose goal is to learn a dissimilarity metric from the pairs of signatures. To the best of our knowledge, this is the first time these recurrent Siamese networks are applied to the field of on-line signature verification, which provides our main motivation. We propose both Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) systems with a Siamese architecture. In addition, a bidirectional scheme (which is able to access both past and future context) is considered for both LSTMand GRU-based systems. An exhaustive analysis of the system performance and also the time consumed during the training process for each recurrent Siamese network is carried out in order to compare the advantages and disadvantages for practical applications.

3. PROPOSED METHODOLOGY

DEEP LEARNING SIGNATURE VERIFICATION ON-LINE

Despite the absence of publicly available data, some researchers have conducted early evaluations of various deep learning architectures for on-line signature verification. It includes a comparison of various deep learning algorithms, as well as the accompanying database, experimental protocol, and performance results. To begin, we'd like to point out that it's impossible to provide a fair comparison across methodologies because each study used distinct datasets and experimental protocol circumstances. Inter-session variability, the quantity of training signatures available per user, and the complexity of the signatures all have a big impact on how well the system works. This issue affects not only deep learning systems, but the entire area of handwritten signature verification.

One of the first research to look into the capabilities of present deep learning algorithms for online signature verification. Otte et al. used a total of 20 users and 12 valid signatures per user to train their Long Short-Term Memory (LSTM) RNNs in that study. We looked at three potential scenarios: i) training a broad network to spot forgeries and real signatures, ii) For each writer, a separate network is being trained, and iii) Only valid signatures are used to train the network. All of the experiments, however, failed, with the best network configuration achieving a final EER of 23.8 percent, far below the state of the art, implying that LSTM RNN systems trained with standard mechanisms were not appropriate for the task of signature verification because the amount of available data for this task is limited compared to others, such as handwriting recognition.

More recently, with the creation of novel architectures, several researchers have demonstrated the promise of deep learning for the task of on-line signature verification. The authors presented a Siamese architecture-based end-to-end writer-independent RNN signature verification system. In order to have access to both past and future context, both LSTM and Gated Recurrent Unit (GRU) methods were examined, employing both conventional and bidirectional configurations (i.e., BLSTM and BGRU). For skilled forgeries, the suggested system outperformed a state-of-the-art signature verification system based on DTW and feature selection approaches. In the case of random forgeries, however, it failed to outperform DTW.

To extract robust features, Lai and Jin proposed using Gated Auto Regresive Units (GARU) in conjunction with a novel descriptor called Length-Normalized Path Signature (LNPS). The ultimate classification was based on DTW. Experiments were conducted with various databases and experimental techniques, yielding positive results, particularly against random forgeries. It's vital to note the results obtained with the finger as a writing tool while using the Mobisig database. Their proposed method yielded a final EER of 10.9 percent for competent forgeries, which is significantly lower than the MCYT database's result. This finding emphasises the difficult finger input scenario for signature verification.

The authors suggested a system based on an LSTM autoencoder for modelling each signature into a fixed-length feature latent space and a Siamese network for final classification in this research line. The authors tested their method on the SigWiComp2013 dataset and found that expert forgeries yielded an EER of roughly 8.7%.

Multilayer Perceptron (MLP)-based techniques were suggested as simpler alternatives. To model the signatures, Hefny and Moustafa used Legendre polynomials coefficients. Their proposed method was tested using the Sig- Comp2011 (Dutch dataset) and yielded an EER of 0.5%.

ICDAR 2019[3] has just published a number of new approaches offered by various writers. Convolutional Neural Networks (CNNs) and Deep Time Warping (DTW)-based approaches were demonstrated. Only competent forgeries were used to test their suggested approach on the MCYT database, demonstrating how the amount of training data has a significant impact on system performance.

A stroke-based LSTM system was also proposed by Li et al. Their proposed method seems to outperform the findings obtained in the BiosecurID database. However, the findings obtained in other databases were substantially worse, with EERs above 10%, demonstrating the network's weak generalisation capacity.

Sekhar et al. proposed a Siamese CNN architecture in ICDAR 2019 that was similar to the method described. The accuracy of their proposed approach was tested on the MCYT and SVC databases, with drastically different results for each database.

Using fixed-length representations from signatures of variable length, an intriguing study using a lightweight one-dimensional CNN signature verification method was recently proposed. They also looked into the possibility of using synthesis techniques to replace skilled forgeries during training. Their suggested method was tested against professional forgeries utilising the MCYT and SVC databases, with promising results.

On the basis of BLSTM/BGRU, Nathwani presented an on-line signature verification system. The report does not provide much information about the system, architecture, or training technique. On SVC, the best result was an Average Error (AE) of 8.8 percent.

Therefore, we present the results obtained with our proposed TA-RNN system over the new DeepSignDB database in Table I. Due to the aforementioned limitations, we propose and distribute to the research community a standard experimental technique for on-line signature verification in this study, with the goal of allowing future comparative investigation of new proposed architectures.

4. EXPERIMENTAL RESULTS

In our proposed standard experimental technique, we test two different scenarios. First, an office-like situation in which users sign documents using a stylus, and secondly a mobile scenario in which users sign documents with their fingers on mobile general-purpose devices. It's worth noting that the DeepSignDB results are derived by combining all of the signature comparisons of the related databases, rather than by averaging the EERs of the corresponding databases. We investigate a single system threshold in this fashion, modelling real-world events.

5.CONCLUSION

The DeepSignDB on-line hand-written signature database, the largest on-line signature database to date, was introduced in this article. This database has over 70K signatures collected from 1526 people using both stylus and finger inputs. Two acquisition situations (office and mobile) are studied, with a total of eight different devices. Along the database, different forms of impostors and the amount of acquisition sessions are also considered.

REFERENCES

1. A.Morales *et al.*, "Keystroke Biometrics in Response to Fake News Propagation in a Global Pandemic," *2020 IEEE 44th Annual Computers, Software, and Applications* Conference (COMPSAC), 2020, pp. 1604-1609, doi: 10.1109/COMPSAC48688.2020.00-26.
2. R. Tolosana, R. Vera-Rodriguez and J. Fierrez, "BioTouchPass: Handwritten Passwords for Touchscreen Biometrics," in *IEEE Transactions on Mobile Computing*, vol. 19, no. 7, pp. 1532-1543, 1 July 2020, doi: 10.1109/TMC.2019.2911506.
3. M. A. Ferrer, M. Diaz, C. Carmona-Duarte and R. Plamondon, "iDeLog: Iterative Dual Spatial and Kinematic Extraction of Sigma-Lognormal Parameters," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 1, pp. 114-125, 1 Jan. 2020, doi: 10.1109/TPAMI.2018.2879312.
4. R. Vera-Rodriguez et al., "DeepSignCX: Signature Complexity Detection using Recurrent Neural Networks," 2019 International Conference on Document Analysis and Recognition (ICDAR), 2019, pp. 1326-1331, doi: 10.1109/ICDAR.2019.00214.
5. Deep Learning and Feature Representation Using Legendre Polynomial Coefficients},author={Amr Hefny and M Mohamed Yassen Moustafa},booktitle={AMLTA},year={2019}

6. X. Wu, A. Kimura, B. K. Iwana, S. Uchida and K. Kashino, "Deep Dynamic Time Warping: End-to-End Local Representation Learning for Online Signature Verification," 2019 International Conference on Document Analysis and Recognition (ICDAR), 2019, pp. 1103-1110, doi: 10.1109/ICDAR.2019.00179.
7. C. Li et al., "A Stroke-Based RNN for Writer-Independent Online Signature Verification," 2019 International Conference on Document Analysis and Recognition (ICDAR), 2019, pp. 526-532, doi: 10.1109/ICDAR.2019.00090.
8. Chandra, S., Kumar, V. A novel approach to validate online signature using dynamic features based on locally weighted learning. *Multimed Tools Appl* (2022). https://doi.org/10.1007/s11042-022-13159-6
9. R. Tolosana, R. Vera-Rodriguez, J. Fierrez and J. Ortega-Garcia, "Exploring Recurrent Neural Networks for On-Line Handwritten Signature Biometrics," in IEEE Access, vol. 6, pp. 5128-5138, 2018, doi: 10.1109/ACCESS.2018.2793966.
10. R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales and J. Ortega-Garcia, "Do You Need More Data? The DeepSignDB On-Line Handwritten Signature Biometric Database," 2019 International Conference on Document Analysis and Recognition (ICDAR), 2019, pp. 1143-1148, doi: 10.1109/ICDAR.2019.00185.