Online authenticity verification of a biometric signature using dynamic time warping method and neural networks.

1st Krzysztof Walentukiewicz
Multimedia Systems Department
Gdansk University of Technology, ETI
Gdansk, Poland
s175478@student.pg.edu.pl

2nd Albert Masiak
Department of Biomedical Engineering
Gdansk University of Technology, ETI
Gdansk, Poland
s176075@student.pg.edu.pl

3rd Aleksandra Gałka
Multimedia Systems Department
Gdansk University of Technology, ETI
Gdansk, Poland
s175975@student.pg.edu.pl

4th Justyna Jelińska
Multimedia Systems Department
Gdansk University of Technology, ETI
Gdansk, Poland
s175505@student.pg.edu.pl

***Abstract*—To ensure proper authentication, e.g. in banking systems, multimodal verification are becoming more prevelant. An offline signature is a well known however not the safest way to verify identity. In this paper the online signature analysis based on dynamic time warping (DTW) coupled with neural networks has been proposed. The goal of this resarch was to test a hypothesis, that by using neural networks with DTW improves the effectiveness of verification of a handwritten signature, comparing to a clasifier based on fixed thresholds. The DTW algorithm was used as a feature extraction method and a similarity measure. On top of dynamic time warping, as the first and second model a multilayer perceptron (MLP) was proposed. Thirdly a convolutional neural network (CNN) has been developed.**

A dataset has been created, containing model, verification and forged signatures gathered from a research group using a biometric pen. Each individual signature consists of three dimensional signals of accelerometer, gyroscope, inclinometer (both angle and acceleration) and of pressure of the pen. An independent DTW method has been conducted on the multidimensional signals. The forged and verification signatures has been compared with model ones. First MLP model used average and standard deviation of distances between corresponding samples of compared signals and the second one used the average and area under DTW curve. The CNN model, had used the DTW matrix as an input. To evaluate system efficiency the results has been compared with DTW model based on constant thresholds. The research has proved that the DTW coupled with neural networks perform significantly better than the baseline method. The results are presented and discussed in this paper.

***Index Terms*—dynamic time warping, handwritten signature verification, feature-based recognition, neural networks**

I. Introduction

In recent years, there has been an increasing need for secure and reliable methods of authentication. Handwritten signatures are a flawed and not reliable way of authenticating oneself, as they can easily be forged. Due to the development of technologies allowing the analysis of biological traits, which are unique identifying characteristics, such as fingerprints or a scan of a retina, there has been an increasing number of biological based authentication services used throughout the industry.

On-line signature authentication using a biometric pen allows for a less vulnerable to fraud way of analysing signatures, by dynamically collecting samples from various sensors during the signature process, such as pen pressure and acceleration, which can later be used to measure similarities of a model signature and currently examined one.

One such method is the use of Dynamic Time Warping (DTW), which is an algorithm used to measure simmilarities between two signals that can be of different lengths. More detailed descripion of this method is included in a later part of this paper. Previous research has proposed a verification method based on fixed DTW thresholds determined experimentally on a training set [1]. Our goal was to test a hypothesis, that using neural networks to verify the authenticity of a handwritten signature, parameterized using the Dynamic Time Warping method (DTW), improves the effectiveness of verification compared to a classifier based on fixed thresholds.

In this research article, we propose a method for verifying the authenticity of a biometric signature using a combination of dynamic time warping (DTW) and neural networks. The DTW method is used to compare the similarity of two biometric signatures and parametrize them, while the neural network is trained to classify the signatures as belonging to a person trying to authenticate themselves or a forgery attempt based on the DTW similarity score. We also perform a statistical analysis of the models performance to check if there is a statistical difference between them.

II. Related works

A Systematic Literature Review (SLR) has been conducted, which revealed a research gap in the topic of the combination of neural networks and DTW. The research of the related works showed a wide range of different approaches used recently to solve the online signature verification problem. However, the solution proposed in this article - combining DTW algorithm with neural networks turned out to be less common than it was presumed. Only about 20% of all methods used in this field used a combination of DTW and Machine Learning (ML) algorithms. The most popular ML techniques involved Support Vector Machine and Recurrent Neural Networks, but Feed-forward Neural Networks were also among the most explored ways of dealing with the authentication problem.

During the SLR process it has been discovered that over 46% of the articles found used only the DTW method for signature verification. It was expected that many would use this method, because it was widely used in a simmilar field - word recognition in the late 1970's [2] and early 1980's [3]. Despite that one of the first usage of a DTW method for curve matching and signature verification was presented in 1999 [4].

A threshold based DTW signature identification method using signature envelope was presented in [5]. The scheme used basic features such as X, Y coordinates of given signature and was tested on a Japanese handwritten signature dataset. In this approach, the authors developed personalized models, that created a decision boundary based on the maximum and minimum variations of the X and Y signals after DTW method was applied. Although the approach was not complicated it managed to outperform previously proposed approaches by achieving an accuracy score of almost 80% and a False Acceptance Rate (FAR) and False Rejection Rate (FRR) scores of 27.35% and 15.18% respectively.

A method that used the whole DTW matrix in combination with the DTW scores derived from comparing two signatures was proposed in [6]. Until the work of Sharma et. al. prior works utilized only the DTW scores to authenticate a test signature. It has been shown in this paper, that using the fusion of the DTW score and the whole DTW matrix can improve the performance of given model.

Different feature extraction methods have been used in combination with neural networks. A discrete wavelet transform (DWT) was one of the successful ones [7]. From the X and Y coordinates, the features of pen movement angles were calculated. Afterwards every signal was independently transformed by the DWT and combined in a signature feature vector, which was then either matched or rejected, by a neural network. It has been shown, that using the DWT with neural networks can lead to 90% success rate, which shows that such combinations can be a very powerful tool.

The concept of connecting DTW with neural networks was proposed among others in [8]. A Deep Dynamic Time Warping has been introduced, by combining a Siamese Network that extracts a feature sequence from each of the signature signals and a DTW block that aligns the sequences of two inputs. It has been shown that such approach can achieve lower Equal Error Rate (EER) than using only DTW or only a Siamese network.

As an improvement of the previous work the authors took advantage of features that DTW extracts and added Siamese Network to it [9]. The Siamese network was incorporated directly into the DTW algorithm, leading to a novel method called Prewarping Siamese Network. The optimization was done using a local embedding loss. For training of this model four datasets were used: MCYT-100 [10], BiosecurID SONOF [11], and SUSIG vi-sual and blind sub-corpuses [12]. This novel approach resulted in the EER value at around 2.11%.

In [13] the authors decided to fuse the scores of 3 classifiers - Deep BiLSTM, SVM with DTW and SVM with different comparator, proposed in the paper. The signature was recognized as genuine when the sum of the scores of 3 classifiers for genuine signature was higher than for forged. Whereas the separated scores of those classifiers were rather weak, the fusion of them results in EER lower than 1% on both SVC2004 [14] and MCYT-100 datasets.

Around 43% of the articles found during the SLR process did not include DTW at all. One of such approaches was a method described in [15]. The authors were inspired by the latest progress on Recurrent Neural Networks (RNN) and tried to implement it into the problem of signature authentication. However, there are many drawbacks when it comes to using RNN. It requires a relatively large training set and significant amount of computational power. On the other hand the results are promising, getting EER at around 2.37% on SVC-2004 dataset.

Another approach that involved RNN was [16]. Authors tried to combine RNN with a Siamese architecture trained on the BiosecurID dataset. Different training scenarios of authentication problem were considered: skilled forgeries, random forgeries and combination of skilled plus random forgeries. The results were as follows: 5.50% EER for skilled forgeries and 3.00% for random forgeries.

III. Dynamic Time Warping

For the purpose of comparing two signatures made using a biometric pen it was necessary to use an algorithm that could recognize them as belonging to the same person even if they varied in length. It could have been the case of signing with a different speed and size of font.

Dynamic Time Warping (DTW) compares two signals which may be of different lengths - it seeks for the temporal alignment that minimizes a certain distance metric between aligned series. A temporal alignment is a matching between time indexes of the two signals. The algorithm tries to find the best match between samples with regards to their surroundings. The result of running a DTW algorithm is a matrix with minimal alignment costs between samples. In this matrix an optimal path can be created by following the minimal costs of the surrounding samples starting from the comparison of the last sample pair up to the first pair. In

an ideal situation a path created by comparing two signals would be a diagonal of the matrix, meaning that the signals are perfectly matching. The path divergence from the diagonal with its costs is distinctive for a given case and because of that it can be used to differentiate between a forgery attempt and authentic signature. The DTW algorithm formula is described below.

Let's assume two signatures F and G:

$$F = f_1, f_2, f_3, ..., f_n \tag{1}$$

$$G = g_1, g_2, g_3, ..., g_m \tag{2}$$

The distance between them can be described as follows:

$$d(f_i, g_j) = |f_i - g_j| \tag{3}$$

The cells in matrix are computed as a cost function:

$$\gamma_{i,j} = d(f_i, g_j) + min(\gamma_{i-1,j-1}; \gamma_{i-1,j}; \gamma_{i,j-1}) \tag{4}$$

The matrix as a whole can be used as a representation of the comparison between two signals, however it is possible to get similar information from a more compact way, by trying to find the path which represents the optimal cost path in a matrix. To get the path the first thing is to find the last element in matrix and move back to the first element using equations below:

$$w^{'} = \{w_k, w_{k-1}, ..., w(0)\} \qquad max(m;n) < m+n+1 \tag{5}$$

$$w_l^{'} = \begin{cases} (i-1, j-1) & \gamma_{i+1,j+1} = min(\gamma_{i-1,j-1}; \gamma_{i-1,j}; \gamma_{i,j-1}) \\ (i-1, j) & \gamma_{i+1,j} = min(\gamma_{i-1,j-1}; \gamma_{i-1,j}; \gamma_{i,j-1}) \\ (i, j-1) & \gamma_{i,j+1} = min(\gamma_{i-1,j-1}; \gamma_{i-1,j}; \gamma_{i,j-1}) \end{cases} \tag{6}$$

One of the methods to find out if both signatures are similar, proposed in previous research [1] uses the comparison of the result from DTW algorithm $p_s{}'$ with given threshold p_{THR}. Thanks to this measurement it is possible to get a degree of similarity in range 0 to 1. It may be done using the equation below:

$$p = \begin{cases} 1 & p_s^{'} < p_{THR} \\ \frac{p_{THR}}{p_s^{'}} & p_s^{'} > p_{THR} \end{cases} \tag{7}$$

IV. Dataset

The first part of our research was to collect proper data. To do so, 33 people were gathered. The participants were 16-30 years old, representing both genders and right and left-handed writers. In the process of gathering data the study group was divided into smaller subsets of 2 to 5 people. Each person from the sub-group signed 5 model and 5 verification signatures. Then, every other member of the same sub-group tried to forge the person signature 5 times, firstly just after they saw the model signature (random forgery) and secondly after practicing singing for someone else (skilled forgery).

Each such signature consists of 13 different signals which are:

- Accelerometer, with acceleration measured in all three dimensions
- Gyroscope (three dimensions)
- Pressure of the pen
- Inclinometer, measuring both angle and acceleration in three dimensions

Signals of position in time have been neglected due to the ethical concerns and noisy data. Moreover, some signatures were removed and not used in the dataset due to biometric pen defect that pause recording signals in the middle of signature. Finally the data consist of 322 model signatures, 328 verification signatures and 537 forged signatures.

A. Data processing

Each verification and forged signature was compared to corresponding model signature using DTW method. For such pairs of signatures *fastdtw* library [17] was used to calculate accumulated distance and DTW path. We decided to use the independent DTW algorithim [18], therefore calculations were made for each of 13 signals separately.

Instead of calculating the whole matrix at once the *fastdtw* library uses the divide and conquer method to make approximation about the DTW matrix and then goes into details in the smaller parts. This implementation of DTW method was compared to one made by authors and another from *dtw* library [19]. It proved to be the fastest of them in terms of time of execution, while producing fairly accurate results.

Based on the samples of the DTW path we calculated distances between model and compared signal samples creating a new feature used in the dataset called *pairwise cost*. Therefore the basic dataset consists of *accumulated DTW distance, DTW path and pairwise cost* for every signal in every model and compared signature pair. All of the features were saved to the .json files.

From that data a labeled dataset has been created. The processed output of DTW method made from verification and model signatures were labeled as 1 whereas output from forged and model as 0. Subsequently the training set contains 1071 elements of class 0 and 640 elements of class 1.

B. Test set

For each individual one randomly selected verification signature was set aside into test set. Then we generated DTW output of that verification signature combined with every of 5 model individual's signature. Finally, there were five true samples in the test set for each of the 33 test subjects.

The number of falsified samples depended on the type of forgery carried out for the individual. For each person, there was at least one forger (test groups ranged from two to four people) who performed five random forgery attempts, or five random forgeries and five skilled forgery attempts. Therefore, in the test set for some examinees there are five samples of forgeries, and for some there are ten. In the end, the test set contains 383 samples, 223 of class 0 and 160 of class 1.

C. MLP models datasets

Using the pairwise cost data the dataset for MLP models has been created. For the first MLP model, average and standard deviation of pairwise cost have been calculated, whereas the average of pairwise cost and the averaged area under the DTW path has been used in the second one.

D. CNN model dataset

The input of the CNN model is the DTW cost matrix calculated between corresponding sensors from model signatures. All the matrices has been resized to the their average size in dataset. Resizing was necessary due to fixed size of model input to standarize DTW matrices, which were of various sizes, depending on length of signals gathered. We have decided to use linear interpolation to upscale an image or area interpolation to downscale an image. In order to input this to model all matrices are merged into tensor of size 13x270x300.

V. Models

When creating and training the models the main concern was relative small size of the dataset. Resulting, architectures were rather shallow and simple to train properly and achieve satisfactory results. In total three models were created: two MLP models with different datasets and a CNN model.

A. MLP models

The first model architecture was based on Multilayer Perceptron (MLP). It consisted of two hidden layers: 64 and 32 neurons respectively with the addition of dropout with probability 30% to prevent overfitting. As an activation function ReLU function was chosen. Model was trained for 100 epochs using Binary Cross Entropy as a loss function and batch size of 64. For optimization ADAM was chosen with the following parameters:

$$lr = 0.001, \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1 * 10^{-7}$$

The training process was stopped earlier when in 10 last epochs loss haven't improved. This architecture was trained with two different datasets resulting in two distinct models. When it comes to the first model, dataset with mean and standard deviation of optimal path in DTW matrix, for the second the mean and area under optimal path.

B. CNN model

The second developed model architecture was based on Convolutional Neural Networks. The idea was to treat cost matrices produced by DTW as images and input them to the network to predict if given signature is forged or not. Convolutional stage of the model consists of three convolutional layers with 2 dimensional filters in number 16, 32, 64. For all of them padding was set to *same*, kernel size $3x3$ and activation function was *ReLU*. After each convolutional layer were pooling layer, performing max pooling operation in 2 dimensions with window size 2x2 and padding same. Next stage is consisting of 2 dense layers with sizes 128 and 32. For both of them ReLU was used as an activation function and dropout with probability 20%. The training was performed for 100 epochs with a loss function *Binary Cross Entropy*. Batch size was 32 and the optimizer was chosen to be ADAM with parameters :

$$lr = 0.001, \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1 * 10^{-7}$$

During training early stopping was used after 10 epochs without improvement in loss value.

VI. Experiments and statistical analysis

Given the biometric nature of the solutions proposed in this and prior works [1], and their potential applications in industries such as banking, more informative metrics were elected. Specifically, we evaluated the False Acceptance Rate (FAR), defined as the ratio of the number of forged signatures accepted by the system to the total number of forgeries, and the False Rejection Rate (FRR), defined as the ratio of the number of authentic signatures rejected by the system to the total number of authentic signatures.

Ideally, both of these metrics would be equal to zero, but they are dependent on one another - as one metric's value gets lower the second's tends to get higher. During the evaluation of the results, we have given more attention to the FAR metric, as in our opinion, it carries more weight in applications such as the banking industry, where the goal is to minimize the ratio of forgeries accepted by the system, even at the expense of a higher number of incorrectly rejected authentic samples.

A. Model evaluation and comparison

The proposed models have been implemented and evaluated on a test set described in paragraph IV-B. The results of the evaluation, along with the specified metrics, are shown in table I.

TABLE I
METRICS FOR THE MODELS

	constant threshold	CNN	MLP (avg,std)	MLP (area)
FAR	25.07%	**1.57%**	6.01%	4.96%
FRR	18.80%	**13.58%**	27.15%	24.54%

The results of the evaluation indicate that each of the models proposed in this study significantly outperforms the model based on fixed thresholds in terms of the False Acceptance Rate (FAR) metric. It is noteworthy that the convolutional model, achieves the lowest FAR value of1.57%. This represents a notable improvement compared to the baseline model. On the other hand, the Multilayer Perceptron (MLP) models, while achieving lower FAR values, exhibit higher False Rejection Rate (FRR) values, which is an undesirable behavior.

In addition to evaluation on the whole test set, the performance of the proposed models was assessed on only random forgery samples and only skilled forgery samples, contained in the test set. The results of this comparison are presented in table II:

TABLE II
FAR VALUES FOR RANDOM AND SKILLED FORGERIES

	constant threshold	CNN	MLP (avg,std)	MLP (area)
Random forgeries	20.00%	**0.00%**	0.83%	0.83%
Skilled forgeries	29.17%	**1.67%**	8.33%	4.17%

The assesment results of FAR values indicate that all proposed models outperform the baseline significantly. For random forgeries all FAR values are below 1%, which indicates that all models are secure, reliable and resilient to a forger, who has only seen given signature for a brief moment. As for the skilled forgeries the proposed models still outperform the baseline, but a greater variation can be observed. A noteworthy result is that of a CNN model which FAR value is below 2% even for skilled forgeries.

B. Statistical analysis of the results

To evaluate the statistical significance of the results of this study, the Cochran's Q test was employed. This non-parametric statistical test is used to determine whether k treatments have identical effects [20]. In this case, the treatments were the performance of the proposed models.

The following null and alternative hypotheses were used:

- null hypothesis (H_0) : The performance of all the models is equally effective - the proportion of correct predictions is the same between all models.
- alternative hypothesis (H_1) : There is a difference in performance between the models - the proportion of correct predictions in at least one of the models is different.

The Cochran's Q test statistic has been calculated as:

$$T = k(k-1)\frac{\sum_{j=1}^{k}(X_{\bullet j} - \frac{N}{k})^2}{\sum_{i=1}^{b} X_{i\bullet}(k - X_{i\bullet})} \tag{8}$$

where: k is the number of models, $X_{\bullet j}$ is the column total for the j^{th} model, b is the number of test samples, $X_{i\bullet)}$ is the row total for the i^{th} sample, N is the grand total. The test statistic T follows a χ^2 distribution with $k-1$ degrees of freedom.In the case of this study $k = 4$, so the distribution has 3 degrees of freedom. If the p-value associated with the test statistic is less than a certain significance level (for the purpose of this comparison $\alpha = 0.05$ has been chosen), the null hypothesis can be rejected and it can be concluded that there is sufficient evidence to say the proportion of correct predictions is different for at least one of the models. Cochran's Q test statistic and p-value has been calculated and presented in table III.

TABLE III
COCHRAN'S Q TEST RESULTS

χ^2	88.32
p-value	$5.03 * 10^{-19}$

There is sufficient evidence to reject the null hypothesis and conclude that there is a difference in performance between the models.

Furthermore, the McNemar test has also been conducted, to examine the statistical significance of differences between pairs of models. The McNemar test is a well-known statistical test for analyzing the statistical significance of differences in classifier performance [21]. This test is also a χ^2 test for goodness of fit that compares the distribution of counts expected under the null hypothesis with the observed counts. The following null and alternative hypotheses were used:

- null hypothesis (H_0) : The performance of the two analysed models is equally effective - the proportion of correct predictions is the same.
- alternative hypothesis (H_1) : There is a difference in performance between the models - the proportion of correct predictions is different.

The McNemmar's test statistic has been calculated as:

$$\chi^2 = \frac{(b-c)^2}{b+c} \tag{9}$$

where: b is the number of times that the second model has predicted wrongly and the first has predicted correctly and c is the number of times that the first model has predicted wrongly and the second has predicted correctly.

In the same way as in the Cochran's Q test, if the p-value associated with the test statistic is less than a certain significance level (for the purpose of this comparison $\alpha = 0.05$ has been chosen), the null hypothesis can be rejected and it can be concluded that there is sufficient evidence to say the proportion of correct predictions is different for the models.

The McNemar test statistic (χ^2) and p-value have been calculated and the results are presented in table IV.

TABLE IV
MCNEMMAR'S TESTS VERSUS BASELINE

	vs. CNN	vs. MLP (avg,std)	vs. MLP (area)
χ^2	69.54	8.28	15.67
p-value	$7.48 * 10^{-17}$	$4 * 10^{-3}$	$7.52 * 10^{-5}$

There is sufficient evidence to reject the null hypothesis for the comparison between the baseline model and the models proposed in this paper. It can be concluded that there is a difference in performance between the models.

TABLE V
MCNEMMAR'S TESTS WITHIN MODELS

	CNN vs. MLP (avg,std)	CNN vs. MLP (area)	MLP vs. MLP (area) vs (avg,std)
χ^2	44.50	30.56	2.97
p-value	$2.54 * 10^{-11}$	3.24^{-8}	0.08

As can be seen in the table V, the only pair of models for which the null hypothesis can not be rejected is the pair of MLP models, so it cannot be concluded with sufficient confidence, that there is a difference between the performance of these models.

VII. Conclusion

Combination of neural networks with DTW algorithm can be effective in biometric signature verification and significantly outperforms model build with fixed thresholds. Statistic tests showed significant differences between proposed models and the one based on fixed thresholds. It is worth noting, that there is no statistically significant difference between two approaches to MLP model, yet area MLP was slightly better in performance. From three developed models the CNN model displayed the highest accuracy as well as low FAR which is crucial in authentication systems. An increase in performance will require significantly more data to train the model, which will result in better generalization and more robust model. Further works could additionally include gathering better trained forgeries and training the models with them to increase the security level.

References

[1] M. Lech and A. Czyżewski, "Handwritten signature verification system employing wireless biometric pen," in *Intelligent Methods and Big Data in Industrial Applications*. Springer, 2019, pp. 307–319.

[2] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE transactions on acoustics, speech, and signal processing*, vol. 26, no. 1, pp. 43–49, 1978.

[3] C. Myers, L. Rabiner, and A. Rosenberg, "Performance tradeoffs in dynamic time warping algorithms for isolated word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 6, pp. 623–635, 1980.

[4] M. Munich and P. Perona, "Continuous dynamic time warping for translation-invariant curve alignment with applications to signature verification," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1, 1999, pp. 108–115 vol.1.

[5] M. Y. Durrani, S. Khan, and S. Khalid, "Versig: a new approach for online signature verification," *Cluster Computing*, vol. 22, pp. 7229–7239, 2019.

[6] A. Sharma and S. Sundaram, "On the exploration of information from the dtw cost matrix for online signature verification," *IEEE Transactions on Cybernetics*, vol. 48, no. 2, pp. 611–624, 2018.

[7] M. M. Fahmy, "Online handwritten signature verification system based on dwt features extraction and neural network classification," *Ain Shams Engineering Journal*, vol. 1, no. 1, pp. 59–70, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2090447910000080

[8] 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," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*, 2019, pp. 1103–1110.

[9] X. Wu, A. Kimura, S. Uchida, and K. Kashino, "Prewarping siamese network: Learning local representations for online signature verification," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 2467–2471.

[10] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho *et al.*, "Mcyt baseline corpus: a bimodal biometric database," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 150, no. 6, pp. 395–401, 2003.

[11] J. Galbally, M. Diaz-Cabrera, M. A. Ferrer, M. Gomez-Barrero, A. Morales, and J. Fierrez, "On-line signature recognition through the combination of real dynamic data and synthetically generated static data," *Pattern Recognition*, vol. 48, no. 9, pp. 2921–2934, 2015.

[12] A. Kholmatov and B. Yanikoglu, "Susig: an on-line signature database, associated protocols and benchmark results," *Pattern Analysis and Applications*, vol. 12, pp. 227–236, 2009.

[13] T. Dhieb, H. Boubaker, S. Njah, M. Ben Ayed, and A. M. Alimi, "A novel biometric system for signature verification based on score level fusion approach," *Multimedia Tools and Applications*, vol. 81, no. 6, pp. 7817–7845, 2022.

[14] D.-Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll, "Svc2004: First international signature verification competition," in *Biometric Authentication: First International Conference, ICBA 2004, Hong Kong, China, July 15-17, 2004. Proceedings*. Springer, 2004, pp. 16–22.

[15] S. Lai, L. Jin, and W. Yang, "Online signature verification using recurrent neural network and length-normalized path signature descriptor," in *2017 14th IAPR international conference on document analysis and recognition (ICDAR)*, vol. 1. IEEE, 2017, pp. 400–405.

[16] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Exploring recurrent neural networks for on-line handwritten signature biometrics," *Ieee Access*, vol. 6, pp. 5128–5138, 2018.

[17] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intelligent Data Analysis*, vol. 11, no. 5, pp. 561–580, 2007.

[18] M. Shokoohi-Yekta, B. Hu, H. Jin, J. Wang, and E. Keogh, "Generalizing dynamic time warping to the multi-dimensional case requires an adaptive approach." Citeseer, 2015.

[19] T. Giorgino, "Computing and visualizing dynamic time warping alignments in r: the dtw package," *Journal of statistical Software*, vol. 31, pp. 1–24, 2009.

[20] W. G. COCHRAN, "THE COMPARISON OF PERCENTAGES IN MATCHED SAMPLES," *Biometrika*, vol. 37, no. 3-4, pp. 256–266, 12 1950. [Online]. Available: https://doi.org/10.1093/biomet/37.3-4.256

[21] T. Kavzoglu, "Chapter 33 - object-oriented random forest for high resolution land cover mapping using quickbird-2 imagery," in *Handbook of Neural Computation*, P. Samui, S. Sekhar, and V. E. Balas, Eds. Academic Press, 2017, pp. 607–619. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128113189000338