Budapest University of Technology and Economics

IMPROVED PREPROCESSING AND CLASSIFICATION ALGORITHMS FOR ONLINE SIGNATURE VERIFICATION

ELŐFELDOLGOZÁSI ÉS OSZTÁLYOZÁSI ALGORITMUSOK TELJESÍTMÉNYÉNEK JAVÍTÁSA AZ ONLINE ALÁÍRÁSHITELESÍTÉSBEN

Ph.D. Thesis Booklet

Mohammad Majed Ahmad Saleem

Advisor:
Dr. Bence Kovari

BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
DEPARTMENT OF AUTOMATION AND APPLIED INFORMATICS

Budapest
2021

Ph.D. Thesis Booklet

Mohammad Majed Ahmad Saleem

Budapest University of Technology and Economics
Faculty of Electrical Engineering and Informatics
Department of Automation and Applied Informatics

1111 Budapest, Magyar tudósok krt. 2/B.

e-mail: msaleem@aut.bme.hu
tel: +36-702890916

Advisor:
Dr. Bence Kovari
Associate Professor

1 Introduction and motivation

Biometrics are representations of the body and measurements related to human features. Biometric identification and authentication are used to identify an individual on the basis of its distinctive features and access control in computer science. They are challenging tasks when it comes to an automated system with high accuracy.

There are various behavioral biometrics that exist, handwriting signatures occupy an exceptional place in this field [Fairhurst, 1997]. The term signature generally means the signing of a written document with one's hand. A signature aims to authenticate a written document or provide notice of its source and oblige the person who signs a written document to comply with the document's requirements. Signatures are the most widely accepted biometric trait by law for verification purposes [Jain et al., 2020].

In computer-aided signature verification, there are two types of signatures distinguished by the input method. The signature is captured using regular pen and paper in offline or static signatures, then scanned as a file image. While in online or dynamic signatures, the whole signing process is acquired and recorded using special digital devices, e.g., a tablet or digital pens. These devices can capture several features like position, pressure, and velocity as a function of time, thereby adding valuable extra information for the verification process.

The first use of the signature verification concept goes back to 439 AD where it was used for document authentication in the Roman Empire [Bibi et al., 2020]. However, the first signature verification started developing from the 1960s [Fauziyah et al., 2009].

The first online and offline verification system was presented in 1973 [Nagel and Rosenfeld, 1973]. After that, thousands of research papers have been published in the field. Any signature verification system typically consists of four main steps: data acquisition, preprocessing, feature extraction, and verification.

In the verification step, some verification and similarity measurement methods are applied to decide whether a signature is genuine or forged. Several approaches are used for this purpose, such as dynamic time warping (DTW), neural networks, or hidden Markov models (HMMs). Among these methods, DTW has shown the most promising results [Malik et al., 2015].

The performance of signature verification systems is measured using the error rate evaluation methods. Both false acceptance rate (FAR) and false rejection rate (FRR) are considered. The most commonly used method is the equal error rate (EER), where FAR and FRR values crossed.

Several signature verification competitions were held to compare the verification systems. In 2004, the first competition was held in the Hong Kong university [Yeung et al., 2004]. These competitions aimed to evaluate the verification systems under similar conditions. There are no two identical genuine signatures that can be provided by an individual. Signatures vary due to age, signing space, or mood. In addition, skilled forgers can provide very similar signatures after practicing the genuine one. Thus, it is hard to build an accurate signature verification system that can fully distinguish genuine and forged signatures. Therefore, the aim of a verification system is to achieve the minimum possible error rate (considering both types).

Although the results of individual methods tested on individual databases are available, these are usually not considered representative results. In addition, the corresponding implementation and additional data are rarely made available. To my knowledge, such research has never been published in the field of signature verification.

This dissertation is intended to evaluate the algorithms used in the field and improve the accuracy of online signature verification.

Based on these, the objectives of my research were the following:

- *Experimentally evaluate and investigate the algorithms used in each step of verification systems.*
- *Present competitive online signature verification with high accuracy to detect skilled forgeries.*
- *Test and evaluate the proposed system on several databases and using different preprocessing methods to eliminate the effect of other factors during the verification steps.*
- *Provide online signature verification that is reliable for real-life scenarios where only a few genuine signatures are available as references, and no forgeries data are available.*

2 Methodological summary

As long as human experts can outperform automatic signature verifiers, there is a need for improvement. Therefore, the objective of my research is to establish a signature model and evaluation methodologies that are suitable for the classification of signatures. I started by analyzing and summarizing the previous work and results in the field. These results are concluded, evaluated and compared in a way that is helpful to the researchers in the field. Further, the existing algorithms and methods are combined to form many verification scenarios to analyze and evaluate the most useful ones and form a more robust verification system that provides better results. Followed by a novel method of using the sampling frequency as an important factor in verification, leading to a new competitive verification system. In addition, I proposed an optimized jk-nearest neighbor algorithm and a combination of the k-nearest neighbor and the DTW algorithms for online signature verification.

These results were analyzed and published in several journals and conferences. Additionally, the project used in this thesis is publicly available for the researchers for future work.

My results can be organized into the following main categories:

- *Review of most relevant journal and conference papers from the last 30 years in the field of signature verification.*
- *Conduct an extensive survey of possible preprocessing and verification methods.*
- *Examine and – when necessary – improve existing methods used in each step of the verification process.*
- *Apply methods to signature databases and demonstrate the feasibility of their practical application.*
- *Apply a large-scale evaluation of signature verification steps.*

- *Presents a new verification system using the analysis of the most effective methods used in each verification step.*

- *Build a new verification system using the sampling frequency effect on the verification accuracy.*

- *Presents a verification system using an enhanced and optimized classification method.*

- *Proposed a novel combination of the k-nearest neighbour and DTW for online signature verification.*

- *Participate in common competitions in the filed.*

- *Apply the methods to common signature databases and demonstrate the feasibility of practical application.*

3 New theoretical results

My results are summarized in three theses, presented briefly in the following.

Thesis I: Systematic evaluation of preprocessing approaches in online signature verification

Publications related to this thesis: **[1] [2] [6] [7] [8] [9] [11] [18]**.

In this thesis, I investigated and analyzed the existing algorithms in the field, presented a systematic evaluation of preprocessing approaches in online signature verification, and built a verification system that combines the most significant approaches that minimized the error rate. I conducted extensive evaluations of some of the most popular design choices. A signature verifier was created for all linear combinations of the introduced methods, including some additional choices such as feature selection or the distance function used in conjunction with DTW. The resulting 42 336 verifier configurations were tested on five different publicly available databases. Supported by over 211 680 experiments, the results of this thesis are the following:

Sub-thesis I.1.: Effects of training set selection and resampling

In this sub-thesis, I investigated the effect of selecting the training set and the resampling on the accuracy of the verification system. To show the degree of bias, I introduced three additional new deterministic approaches for training set selection. The following four approaches have been used in our experiments to select the test set from an ordered list of genuine signatures of a given signee: first 10, last 10, odd 10, and even 10.

For resampling, my approach calculates a consistent time slot length for the signature based on the expected number of samples and the original length of the signature and estimates the new sample values based on the original data using an interpolation technique. In addition to the original sampling rate, four different sample counts (50, 100, 500, and 1000) were used to achieve both upsampling and downsampling. To evaluate the results, I compared the results of individual signature verifier configurations that differed only in their sample counts.

Let $T = \{t_1, t_2, t_3, \ldots, t_n\}$ represent the timestamps corresponding to the original sample points, and n^* be the expected number of samples. The new time slot length can be calculated as

$$t = \frac{t_n - t_1}{n^* - 1}. \tag{3.1}$$

The original timestamps may be replaced with new ones with respect to $t_i^{'} = t_{i-1}^{'} + t$, $t_1^{'} = t_1$. The values of the other features are calculated at the new timestamps by interpolation. Two interpolation types were used in our experiments: cubic and linear.

The results of the sub-thesis suggest that:

- Using the common method of selecting the first ten genuine signatures as references is highly biased and should be avoided, especially when the signatures are sorted by their acquisition time, see Figure 1.

- The most significant difference (8.85%) can be seen in the database SVC2004. Here, the best average ERR (18.07%) was achieved using the first signatures with odd indices in the training set, while the worst results (26.92%) were achieved using the first ten signatures for training.

- Resampling signatures to a fixed number of samples with linear or cubic interpolation only improved the result in a negligible number of cases while all other cases had a negative impact; the EER increased in four out of five databases. In the case of the fifth database (MCYT100), the error rate also increased in 99% of all the cases, compared to the verification results of the original sampling rates.

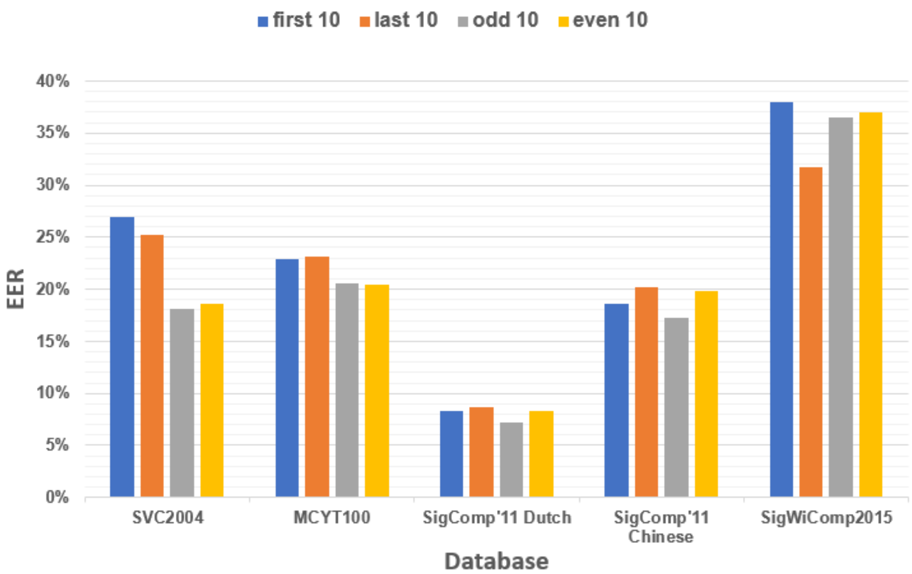

Figure 1: Effect of training set selection strategy on the average EER.

Sub-thesis I.2: Handling of pen-up strokes

In this sub-thesis, I investigated the effect of pun-up stroke handling methods. Both approaches of removing the pen-up strokes and filling in the missing data with interpolated points were tested. The values for the features at the new timestamps were calculated either by linear or cubic spline interpolation. To assess the effects of this operation, I compared pairwise verification results of the non-filtered and filtered configurations.

I also compared three approaches for handling pen-up strokes, namely using the original pen-up strokes, using only pen-down strokes, and interpolation-generated pen-up strokes. Tables 1 and 2 summarize all the results.

The results of the sub-thesis suggest that:

- Removing pen-ups is always beneficial for the SigComp'11 Dutch and Chinese data sets, and it also has a positive effect in 78% of all the cases for the MCYT100 data-set.
- Existing pen-up strokes did not significantly contribute to improving the classification results in the examined databases. Filling in pen-up strokes with synthetic interpolated data did mostly negatively affect the results.
- Using only pen-down strokes may improve the verification accuracy in most cases.

Table 1: Percentage of all Configurations, Where Removing the Pen-up Strokes Resulted in a Lower EER

Database	**Filtering improved EER (all configs)**	**Filtering improved EER (top 5%)**
MCYT-100	13.27%	77.78%
SigComp'11 (Dutch)	29.59%	100.00%
SigComp'11 (Chinese)	38.78%	100.00%

Table 2: Percentage of all Configurations, Where Inserting Synthetic Pen-up Strokes Resulted in a Lower EER

Database	**FillGap linear**	**FillGap cubic**
SVC2004	0%	6.7%
MCYT-100	12.5%	12.5%
SigComp'11 (Dutch)	0%	0%
SigComp'11 (Chinese)	0%	0%
SigWiComp2015	18.8%	18.8%

Sub-thesis I.3: Distance function within DTW

In this sub-thesis, I evaluated the Manhattan and Euclidean distance measurement methods of Manhattan for DTW_D similarity measurements in online signature verification. In Table 3, the effectiveness of the distance functions is compared in terms of two aspects. The first is the percentage of cases in which a distance function improves the equal error rate compared to the other. The second aspect is the quantity of this improvement. The results show that the ratio of cases when the usage of Manhattan distance improved the equal error rate compared to the usage of Euclidean distance was higher than its inverse.

The results of the sub-thesis suggest that:

- The ratio of cases when the usage of Manhattan distance improved the equal error rate compared to Euclidean distance usage is higher than the reverse.
- The usage of Manhattan distance resulted in larger improvements.

Table 3: Comparison of Euclidean and Manhattan Distance Based Verifier Configurations

Database	**Euclidean average EER improvement**	**Euclidean better**	**Equal**	**Manhattan better**	**Manhattan average EER improvement**
SVC2004	0.07%	8.16%	48.98%	42.86%	0.48%
MCYT-100	0.02%	5.10%	47.96%	46.94%	0.38%
SigComp'11 (Dutch)	0.07%	11.22%	51.02%	37.76%	0.14%
SigComp'11 (Chinese)	0.10%	16.33%	44.90%	38.78%	0.28%
SigWiComp2015	0.39%	15.31%	46.94%	37.76%	0.76%

Sub-thesis I.4: Effects of translation and scaling

In this sub-thesis, I investigated the translation and scaling methods and proposed a novel and a competitive combination of both preprocessing methods.

Let us define the preprocessing step as

$$f_i^{'} = f_i - o_f, \tag{3.2}$$

where $f_i^{'}$ is the translated feature value of the i^{th} point of the signature, f_i is the original feature value of this point, and o_f is the value of the new origin in the original coordinate system for the given feature.

This study deals with two main approaches that were implemented and applied during the experiments.

The first approach uses the center of gravity of the features as the new origin. The new origin is calculated as

$$o_f = \frac{\sum_i f_i}{n}. \tag{3.3}$$

The second approach uses the minimum values of the features. In two dimensions, the new origin will be the bottom left corner of the signature-enclosing rectangle. The new origin is calculated as

$$o_f = \min_i f_i. \tag{3.4}$$

I introduce two approaches to scaling. The first scales the values into a fixed interval, while the other approach scales the values based on their standard deviation. In both cases, scaling is executed on the spot such that the minimum value of the scaled features is always fixed.

Let us define "min-max scaling" of a feature F into the interval $[f_{min}, f_{min} + 1]$ as:

$$f_{min} = \min_i f_i, \tag{3.5}$$

$$f_{max} = \max_i (f_i - f_{min}), \tag{3.6}$$

$$f_i' = f_{min} + \frac{f_i - f_{min}}{f_{max}}. \tag{3.7}$$

Let us define the standard deviation-based scaling of a feature as:

$$f_i' = f_{min} + \frac{f_i - f_{min}}{\text{stdev}(f)}, \tag{3.8}$$

where f_i' is the newly scaled value of f_i.

The two most commonly used preprocessing approaches that combine scaling and translation are z-normalization and scaling into the [0,1] range (min–max normalization). These approaches are included in this study through a combination of the previously introduced translation and scaling methods. In addition, the distinction between the two allows me to introduce a new combination, where values are first scaled into the $[f_{min}, f_{min}+1]$ range and then aligned to their center of gravity. Tables 4 and 5 summarize all the results.

The results of the sub-thesis suggest that:

- The best approaches take advantage of both types of translation and scaling algorithms.
- z-normalization results are superior to traditional min-max normalization.
- The proposed novel combination called centered min-max normalization, where values are first scaled into the $[f_{min}, f_{min}+1]$ range and then aligned to their center of gravity yielded competitive results to that of z-normalization.
- Using these approaches in the preprocessing step highly affects the accuracy of verification compared to other algorithms.

Table 4: Average EER of Configurations Using the Given Translation and Scaling Methods for Preprocessing

Preprocessing	**Translation**	**None**	**None**	**To Zero**	**to COG**	**None**	**To Zero**	**To COG**
	Scaling	**None**	**Normalized to 1**	**Normalized to 1**	**Normalized to 1**	**Normalized by stdev**	**Normalized by stdev**	**Normalized by stdev**
Database	SVC2004	19.06%	24.79%	11.61%	8.54%	26.03%	16.36%	8.79%
	MCYT100	31.65%	31.20%	5.97%	3.75%	32.07%	9.45%	3.49%
	SigComp'11 (Dutch)	3.79%	3.51%	3.85%	2.38%	3.71%	4.65%	2.34%
	SigComp'11 (Chinese)	14.65%	16.02%	3.85%	2.80%	16.48%	5.98%	2.30%
	SigWiComp2015	44.00%	42.91%	11.33%	10.59%	44.16%	25.14%	9.62%

Table 5: Comparison of Three Normalization Approaches for the Remaining Configuration Triplets

Database	centered min-max better than min-max	*z*-normalization better than min-max	*z*-normalization better than centered min-max
SVC2004	100.00%	100.00%	42.86%
MCYT-100	100.00%	100.00%	71.43%
SigComp'11 (Dutch)	100.00%	100.00%	35.71%
SigComp'11 (Chinese)	100.00%	100.00%	85.71%
SigWiComp2015	64.29%	78.57%	71.43%

Sub-thesis I.5: The effects of rotation normalization

In this sub-thesis, I investigated the rotation normalization effect on the verification accuracy. The rotation approach taken here is based on that of Xia et al. [Xia et al., 2017]. This method uses the X and Y feature vectors of a signature. The original values x_i and y_i rotated by α are $x_i^{'}$ and $y_i^{'}$, which can be calculated as follows:

$$x_i^{'} = (x_i \times \cos\alpha) - (y_i \times \sin\alpha), \tag{3.9}$$

$$y_i^{'} = (x_i \times \sin\alpha) + (y_i \times \cos\alpha). \tag{3.10}$$

The rotation angle is calculated as

$$\alpha = \frac{1}{2}\arctan\left(\frac{2I_{xy\_centroid}}{I_{y\_centroid} - I_{x\_centroid}}\right), \tag{3.11}$$

Where $I_{xy\_centroid}$, $I_{x\_centroid}$, and $I_{y\_centroid}$ are the moments of inertia referred to as the reference centroids.

The rotation methods were tested on five different databases.

The results of the sub-thesis suggest that:

- Rotation normalization worsened the classification results in over 80% of the total cases and could not produce any performance improvement in three data-sets.
- The data suggests its usage to be more counterproductive in cutting-edge classifiers.

Sub-thesis I.6.: Feature selection

In this sub-thesis, I investigated the effect of selecting the features on online signature verification systems. All possible combinations of the horizontal position, vertical position, and pressure features, namely {X}, {Y}, {P}, {X, Y}, {X, P}, {Y, P}, and {X, Y, P} have been evaluated.

Because of the differences between the databases, one cannot predict which individual feature has the best discriminative power; however, several single features can be grouped. Figure 2. shows the reached average EER values using different feature sets grouped by databases. The results of the sub-thesis suggest that:

- DTW is beneficial in utilizing multiple features; therefore, all three features (X, Y, P) together yield the highest accuracy compared to other feature sets in all the databases.

- I recommend using X, Y, and P together to obtain the best results. Although the average EER of the remaining configurations may be lower for other feature combinations, the best EER was always achieved when the XYP feature combination was used.

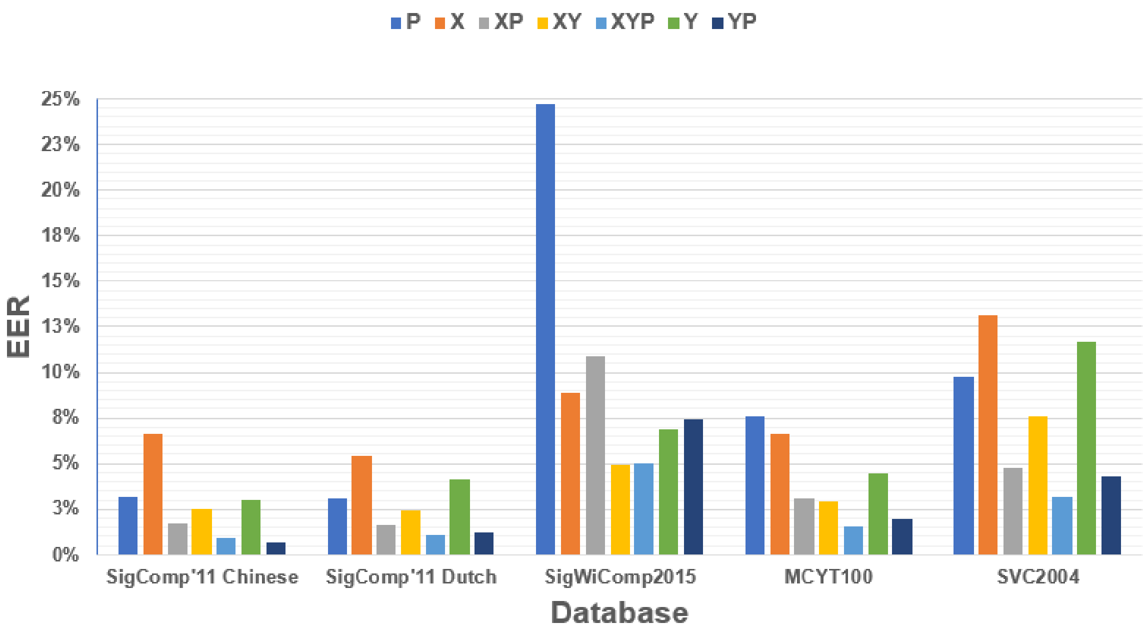

Figure 2: The effect of features selection.

Table 6: Comparison of Results on SVC2004

Reference	Year	EER
[Gruber et al., 2009]	2009	6.84%
[Wang et al., 2011]	2011	6.65%
[Barkoula et al., 2013]	2013	5.33%
min-max		**4.45%**
[Rashidi et al., 2012]	2012	3.61%
[Yeung et al., 2004]	2004	2.84%
***z*-normalization**		**2.83%**
[Chandra et al., 2021]	2021	2.62%
[Hu et al., 2019]	2019	2.5%
[Jia et al., 2019]	2019	2.39%
[Lai et al., 2017]	2017	2.37%
centered min-max		**2.33%**

Table 7: Comparison of Results on MCYT100

Reference	Year	EER
[Kholmatov and Yanikoglu, 2009]	2009	7.8%
[Nanni and Lumini, 2008]	2008	5.2%
[Sae-Bae and Memon, 2014]	2014	4.02%
[Rua and Castro, 2012]	2012	2.85%
[Tang et al., 2017]	2020	2.25%
min-max		**1.92%**
[Sharma and Sundaram, 2016a]	2016	1.55%
z-normalization		**1.35%**
centered min-max		**1.35%**
[Ibrahim et al., 2010]	2010	1.09%

Table 8: Comparison of Results on the SigComp'11 and SigWiComp2015 Databases

Database	Reference/(ID)	EER
SigComp11(Dutch)	[Liwicki et al., 2011]/(7)	7.07%
SigComp11(Dutch)	[Liwicki et al., 2011]/(1)	6.51%
SigComp11(Dutch)	[Liwicki et al., 2011]/(4)	3.73%
SigComp11(Dutch)	[Liwicki et al., 2011]/(5)	3.65%
SigComp11(Dutch)	[Riesen and Schmidt, 2019]	3.24%
SigComp'11 (Dutch)	**min-max**	**1.58%**
SigComp'11 (Dutch)	**z-normalization**	**0.83%**
SigComp'11 (Dutch)	**centered min-max**	**0.72%**
SigComp11(Chinese)	[Liwicki et al., 2011]/(6)	17.6%
SigComp11(Chinese)	[Liwicki et al., 2011]/(1)	15.19%
SigComp11(Chinese)	[Liwicki et al., 2011]/(7)	14.68%
SigComp11(Chinese)	[Liwicki et al., 2011]/(4)	6.83%
SigComp'11 (Chinese)	**min-max**	**1.11%**
SigComp'11 (Chinese)	**z-normalization**	**0.86%**
SigComp'11 (Chinese)	**centered min-max**	**0.64%**
SigWiComp2015	**z-normalization**	**5.04%**
SigWiComp2015	**min-max**	**5.02%**
SigWiComp2015	**centered min-max**	**5.01%**

Sub-thesis I.7.: Systematic approach for online signature verification

In this sub-thesis, I took advantage of the previous observations of the effect of the main algorithms used for online signature verification to present a new competitive system.

According to previous observations, I can define several configurations that yield near-optimal results for most databases. Such a verifier should always consider all three features (X, Y, and P), use the Manhattan distance to calculate the DTW_D score of signature pairs, and remove the pen-up durations without filling these with artificial data. Resampling and rotation normalization should also be omitted. The best location and scale normalization techniques were z-normalization and the proposed centered min-max normalization. As non-centered min-max normalization is also widespread in the field, I show the verification results for all three approaches compared with recently published results in Table 6 for the SVC2004 dataset and Table 7 for the MCYT100 dataset. In Table 8, I compare the results with the recently published results that were derived from datasets of the SigComp'11 and SigWiComp'15 competitions.

The results of this sub-thesis showed that:

- The proposed online signature verification system showed competitive results compared to state of the art as shown in the previous tables.
- I suggest using the introduced centered min-max normalization, which may be an alternative to z-normalization and yielded competitive results.

Thesis II.: Sampling frequency based online signature verification system

Publications related to this thesis: **[3] [10] [12] [13]** .

Online signature verification considers signatures as time sequences of different measurements of the signing instrument. These signals are captured on digital devices and therefore consist of a discrete number of samples. To enrich or simplify this information, several verifiers employ re-sampling and interpolation as a preprocessing step to improve their results; however, their design decisions may be difficult to generalize.

In thesis II, I investigated the direct effect of the sampling rate of the input signals on the accuracy of online signature verification systems without using interpolation techniques and proposed a signer-dependent online signature verification system.

Sub-thesis II.1: The effect of sampling rate on the verification accuracy

In this sub-thesis, I conducted thousands of measurements on five different public signature databases, and the results of the experiments showed a different behavior from the mentioned expectations, see Figure 3. I did not use interpolation to avoid its effect on the results. The relation between the sampling frequency and the error rate was not monotonous in the majority of the cases; however, the error rate had a local minimum. Moreover, this local minimum was achieved in a similar range for several databases, see Figure 4.

Supported by 5600 experiments, the results of the sub-thesis suggest that:

- One can obtain better results by decreasing the sampling rate and the average number of sample points of the databases.
- The optimal sampling frequency should be between 15 Hz and 50 Hz.
- The optimal sample count for online signatures is between 60 and 240 points.
- Using frequencies lower than these ranges greatly decreased the accuracy, whereas using higher frequencies decreased or did not affect the accuracy in 92.5% of the configurations.
- Using sampling frequencies higher than 100 Hz will not improve the accuracy of the systems but will instead increase the computational cost of the verification.

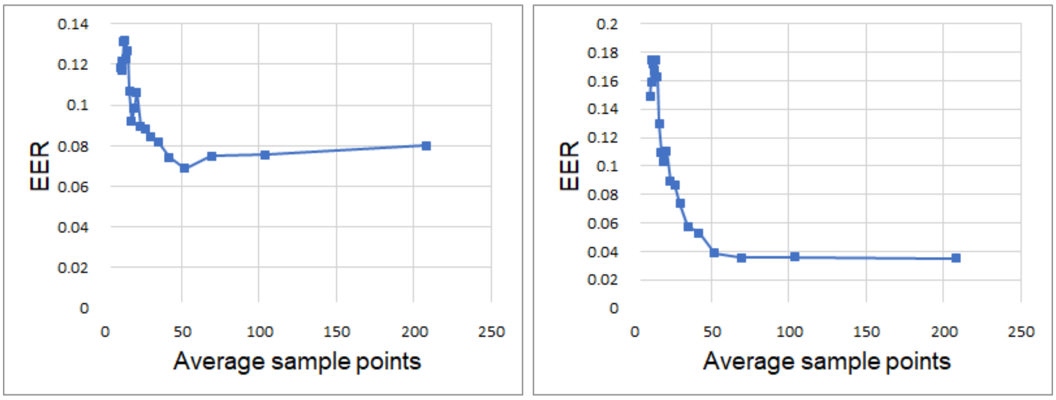

Figure 3: Expected (left) vs. typical (right) behavior of the EER as a function of sample points on SVC2004.

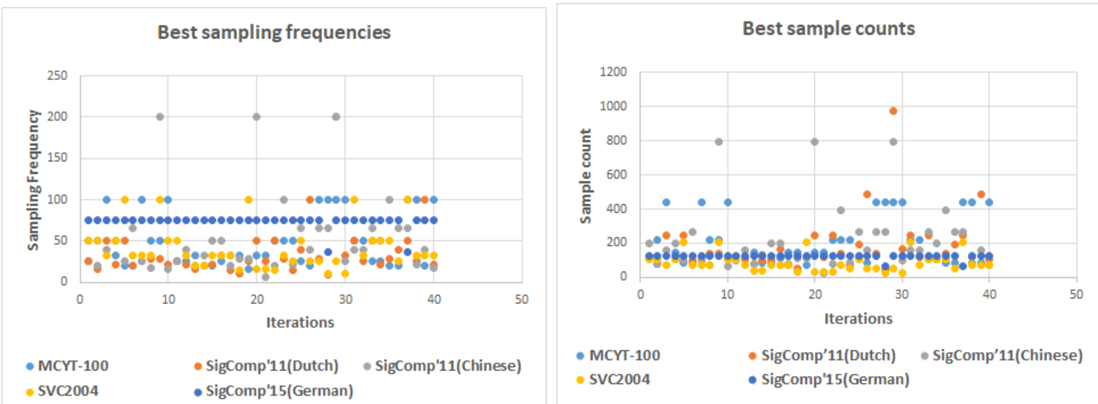

Figure 4: Sampling frequencies and sample points of the best results for all the tests.

Sub-thesis II.2: Accuracy improvement using signer-dependent sampling frequency

In this sub-thesis, I proposed a signer-dependent online signature verification system. The proposed technique is based on choosing the best sampling frequency for each signer before starting the verification process. It uses only the references for choosing the signer optimal sampling frequency, simulating a real-life scenario where no more genuine or forged signatures are available.

The results of the sub-thesis suggest that:

- Using signer-dependent sampling frequency increases the system's accuracy up to 8.4% comparing to the database common sampling frequency.
- The proposed system was tested using 500 different tests, the accuracy improved in around 80% of them.
- Using the optimal preprocessing method and sample number, the accuracy improved 92% of the tests as shown in Figure 5.

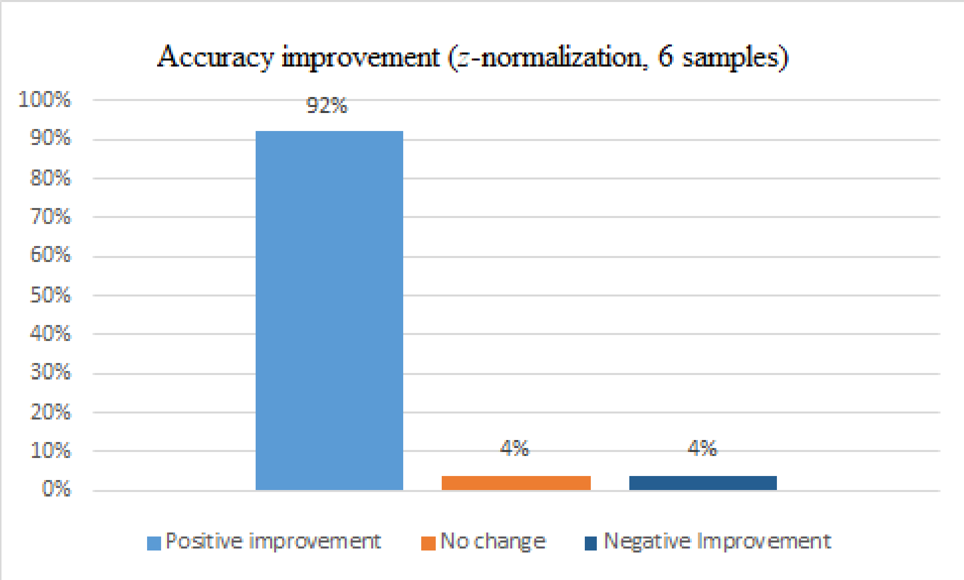

Figure 5: Accuracy improvement.

Table 9: A comparison with recent results in the field

Database	EER%	Reference
MCYT-100	3.69	[Xia et al., 2018]
MCYT-100	3.05	[Sharma and Sundaram, 2016c]
MCYT-100	2.76	[Sharma and Sundaram, 2018]
MCYT-100	1.81	[Lai and Jin, 2018]
MCYT-100	**1.8**	**Proposed**
MCYT-100	1.34	[Okawa, 2019]
MCYT-100	1.28	[Okawa, 2020]
SVC2004	6.65	[Wang et al., 2011]
SVC2004	3.41	[Alpar, 2018]
SVC2004	2.73	[Sharma and Sundaram, 2016b]
SVC2004	1.8	[Foroozandeh et al., 2020]
SVC2004	**1.6**	**Proposed**
SVC2004	0.83	[Fayyaz et al., 2015]
SigComp'11 (Dutch)	3.73	[Liwicki et al., 2011]
SigComp'11 (Dutch)	5.5	[Parodi and Gómez, 2014]
SigComp'11 (Dutch)	3.65	[Liwicki et al., 2011]
SigComp'11 (Dutch)	**0.6**	**Proposed**
SigComp'11 (Chinese)	14.68	[Liwicki et al., 2011]
SigComp'11 (Chinese)	8.93	[Parodi and Gómez, 2014]
SigComp'11 (Chinese)	6.83	[Liwicki et al., 2011]
SigComp'11 (Chinese)	**0.1**	**Proposed**

Sub-thesis II.3: Online signature verification system using signature down-sampling and signer-dependent sampling frequency approaches.

In this sub-thesis, I proposed an online signature verification system based on the previously concluded results in sub-theses 1 and 2. Table 9 shows that the proposed verification system achieved competitive accuracy results compared to the state of the art as shown in Table 10.

Table 10: The error rate for each database.

Database	**EER**
MCYT-100	1.8%
SVC2004	1.6%
SigComp'11 (Dutch)	0.6%
SigComp'11 (Chinese)	0.1%

Thesis III.: Optimized jk-nearest neighbor based online signature verification and evaluation of the main parameters

Publications related to this thesis: **[4] [5] [14] [15] [16] [17]**.

In thesis III, I investigated the k-nearest neighbor algorithm for online signature verification, improved it by using an optimized jk-NN algorithm, and created an online signature verification using the combination of k-NN and DTW algorithms. The summary of the thesis results are the following:

Sub-thesis III.1: Evaluation on the effect of the algorithm parameters

I have studied and evaluated the main parameters of the jk-NN algorithm using hundreds of experiments to eliminate the effect of irrelevant data. Furthermore, I built a jk-NN online signature verification system.

The experiment started by using various reference numbers and thresholds, evaluating and analyzing the verification accuracy's performance. After that, I examined several j and k values and the effect of adjusting the number of nearest neighbors upon the results and compared the k-NN and jk-NN algorithms. For the jk-NN online signature verifier, the j values, k values, number of references used, and threshold selection (value of θ) are the significant factors determining the performance. In the following subsections, each of these factors is discussed and evaluated.

The results of the sub-thesis suggest that:

- Better verification accuracy can be achieved using more than ten original signatures as references, particularly between 13 and 15.
- A threshold θ=1.5 provides the best results for the jk-NN algorithm for online signature verification.
- jk-NN enhanced accuracy compared to the k-NN algorithm when using the same verification system. The accuracy increased by 2.02% for the SVC2004 database, by 0.59% for the MCYT-100 database, by 0.73% for the SigComp'11 database, and by 10% for the SigWiComp'15 database.
- For j, the best results are centered around the value of 5. For k, the optimal value requires careful consideration to achieve optimal results.

Sub-thesis III.2: Optimized jk-NN algorithm for real-life scenarios signature verification

In this sub-thesis, I introduced a new optimized formula of the jk-NN algorithms, which is based on the minimum value of FRR reached within the training set to choose the best value of k, calculated in real-life circumstances where a certain number of signature references are available and used for this purpose. Using the previous evaluation of the algorithm parameters, the values j=5 and θ =1.5 are used with 15 reference signatures in the proposed algorithm.

The idea is to divide the references (R) into two groups, the first group (R_t) is used to calculate the threshold, while the other group is used for testing. FRR is calculated in each iteration using different values of k from the (K_s) group of values. The best value is assigned for k (K_{opt}) to the k that provided the minimum $FRR(K)$ (FRR using the k-NN) among all K_s. The SVS can use these values in the verification phase. The minimum value of (FRR) will not always provide the optimal value of (k) since the references R, and R_t are not the same and will provide different results. Still, it will indicate one of the best values of k that can produce a very accurate result. The new formula of the algorithm is presented in the following equations:

$$\frac{j \in J^* : \left[d\left(S_j, S_{jnn}\right) < \theta \frac{1}{K_{opt}} \sum_{k=1}^{K_{opt}-1} d\left(S_{jnn}, S_{jknn}\right) \right]}{5} > 0.5 \tag{3.12}$$

Where

$$J^* \in \{1, 2, 3, 4, 5\} \tag{3.13}$$

and K_{opt} is the value of k that provides:

$$\underset{\forall K \in K_s}{min} \; FRR(K) \tag{3.14}$$

The results of the sub-thesis suggest that:

The proposed method is ideal for real-life scenarios in which only the references of the signers are available. The achieved accuracies of the method are encouraging considering the applied scenario and can be summarised summarised in Table 11

Table 11: The AER for each database.

Database	**EER**
MCYT-100	3.26%
SVC2004	8%
SigComp’11	2.22%
SigWiComp’15	13%

Sub-thesis III.3: k-NN and DTW for mobile scenario online signature verification

In this sub-thesis, I presented an online signature verification system based on a novel combination of the DTW and the k-NN algorithms. The system considers local thresholds to classify signatures and uses DTW and k-nn algorithms in the classification phase. DTW is used as a distance measurement between the signatures. The k-NN algorithm is used to select the reference signatures and calculate the upper and lower threshold, which plays a significant part in calculating the prediction of the tested signature.

The distance is used to calculate the prediction (P_q) for the questioned signature (S_q) using a calculated forgery threshold F_{th}, genuine threshold G_{th}, and a scaling parameter s as following:

$$P_q = \frac{s * F_{th} - d}{s * F_{th} - G_{th}} \tag{3.15}$$

The proposed system was presented in the SVC2021 signature verification competition for both DeepSignDB and SVC2021_EvalDB databases and achieved the following:

- In the development stage, the proposed system ranked 2nd overall and achieved 7.74% EER in Task1.
- 5.81% in Task 2 (1st place)
- 7.71% in Task 3 (3rd place).
- In the evaluation stage, the proposed system achieved 13.29% in Task2 (bronze medal) and 4th overall.

Tables 12-14 shows the main results achieved by our team in the competition.

Table 12: SVC2021-DeepSignDB - Task2 results.

Position	**Team name**	**EER (%)**	**Points**
1	**SigStat**	**5.81**	**3**
2	**DLCV Lab**	**6.58**	**2**
3	**SIG-Team**	**9.43**	**1**
4	Baseline DTW	10.16	0
5	TUSUR KIBEVS	12.68	0
6	JAIRG	12.86	0

Table 13: SVC2021- SVC2021_EvalDB - Task2 results.

Position	**Team name**	**EER (%)**	**Points**
1	**DLCV Lab**	**7.4055**	**3**
2	**SIG-Team**	**10.1366**	**2**
3	**SigStat**	**13.2878**	**1**
4	TUSUR KIBEVS	13.3929	0
5	MaD Lab	17.2269	0
6	JAIRG	18.4349	0

Table 14: SigStat team results.

Task	DeepSignDB	SVC_EvalDB
Task1	7.74%	11.75%
Task2	5.81%	13.29%
Task3	7.71%	14.48%

4 Application of the new results

The results of this dissertation are based on numerous experiments which were built during the last few years. The majority of these experiments were developed and tested using the SigStat project. SigStat is an open-source .NET class library for offline and online signature verification. The objective of SigStat is to provide a .NET class library for anyone interested in the field of signature verification. One of the project's main goals was to create a modular system, which breaks down the verification system into separate steps. This approach allows students and enables researchers to focus on improving subtasks instead of coping with the whole verification process.

Most existing systems in the data processing area work with generic data types like vectors and multidimensional arrays [TFF, 2021][ML, 2021]. Within the SigStat system, we took advantage of the fact that we know that we are working with signatures. So we were not just trying to fit into some generic frameworks like machine learning models but provide a strongly typed framework for realizing, benchmarking, and debugging signature verifiers.

The development of the SigStat project started almost parallel with my Ph.D. studies. I have added, modified, and tested several algorithms in the project related to online signature verification during the last four years. It includes some database loaders, preprocessing algorithms, math helpers, and classification algorithms. Figure 6 shows an example of the structure of part of the project.

Loaders are used to acquire signatures from known databases, assign each signature to a signee and read their corresponding feature information. For this purpose I have created the following loaders:

- Loader for the SigWiComp 2015 German database (`SigComp15GermanLoader`).
- Loader for the signature competition 2011 Chinese set (`SigComp11ChineseLoader`).
- Loader for the MCYT database (`MCYTLoader`).

Some benchmarks and algorithms were also implemented to test and evaluate some signature verification systems. Up till now, I have implemented the following benchmarks and algorithms:

- `OnlineRotationBenchmark` that tests all the rotation algorithms effects on the signature verification algorithms.
- Rotation normalization algorithm (`NormalizeRotation`).
- Orthogonal rotation algorithm (`OrthogonalRotation`).

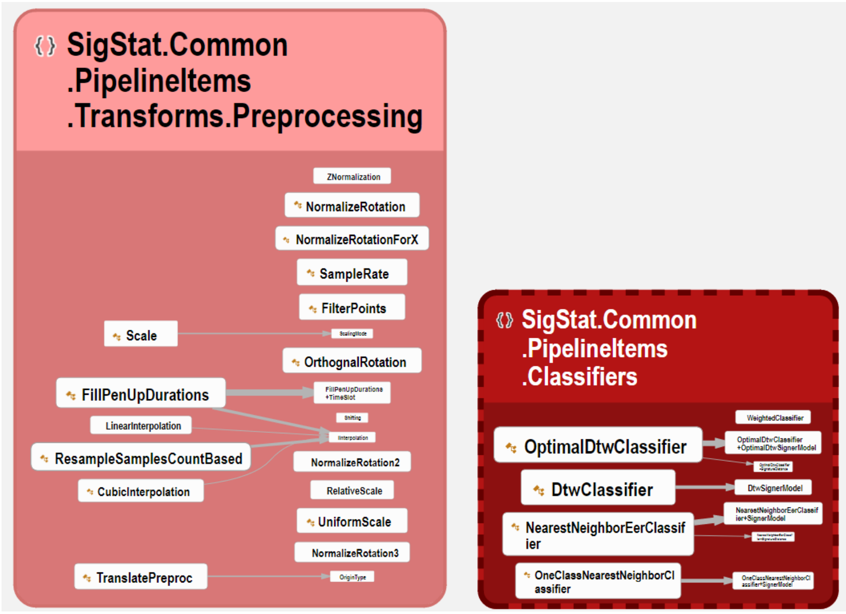

Figure 6: SigStat project structure.

- Shifting algorithm (`Shifting`)
- Algorithm for signature resampling using lower sampling frequency (`Resampling`).
- I have implemented a z-normalization preprocessing method.

Some classifiers were also created or extended by me to support my experiments and to presents more classifiers for the research community, such as:

- Optimal DTW classifier that finds and uses the optimal values of the parameters used in the DTW classifier (`OptimalDtwClassifer`).
- Classifier that uses different DTW windows (`DtwPyWindow`).
- I added some extra attributes to existing classes that can be used for specific benchmarks and algorithms.
- I created some mathematical algorithms that can be used for statistical analysis and classification algorithms.
- I have created a jk-NN classifier and testing code for it.
- I have implemented some functions to test the effect of the sampling frequency on the verification system using different combinations.
- I have implemented a signer-dependent sampling frequency-based classifier.
- I have implemented a classifier based on a combination of both k-NN and DTW algorithms.
- I have implemented a code for testing the optimal sampler.
- I have implemented a signer statistics helper.

5 Scientific Publications

International Journals

1 **M. Saleem** and B. Kovari, "Preprocessing approaches in DTW based online signature verification", Pollack Periodica, Volume 15, number 1, Akadémiai Kiadó, pages 148–157., 2020. (Scopus, Q3).

2 **M. Saleem**, C. Szucs and B. Kovari, "Systematic evaluation of preprocessing approaches in online signature verification", International Journal of Biometrics, 2021. (Scopus, WoS, Q3), **IF: 1.12**. (Under review)

3 **M. Saleem** and B. Kovari, "Online signature verification using signature down-sampling and signer dependent-sampling frequency", Neural Computing and Applications, Springer, Switzerland, 2021. (Scopus, WoS, **Q1**), **IF: 5.606**.

4 **M. Saleem** and B. Kovari, "Optimized jk-nearest neighbor based online signature verification and evaluation of the main parameters", Computer Science, Poland, 2021. (Scopus, Q4).

5 R. Tolosana et al., and **M. Saleem**, "SVC-onGoing: Signature Verification Competition", Pattern Recognition, 2021. (Scopus,Wos, **Q1**), **IF: 7.74**. (minor revisions)

Conference Proceedings

6 **M. Saleem** and B. Kovari, "Preprocessing algorithms in DTW based online signature verification", Automation and Applied Computer Science Workshop 2018, Budapest, Hungary, pages 253–266., 2018.

7 **M. Saleem** and B. Kovari, "Preprocessing algorithms in DTW based online signature verification", University of Pecs, Pecs, Hungary, pages 253–266., 2018.

8 **M. Saleem** and B. Kovari, "Survey of Signature Verification Databases", Multi-Science - XXXIII. microCAD International Multidisciplinary Scientific Conference, Miskolc, Hungary, pages 253–266., 2019.

9 **M. Saleem** and B. Kovari, "An evaluation of preprocessing approaches in online signature verification ", Automation and Applied Computer Science Workshop 2019, Budapest, Hungary, pages 253–266., 2019.

10 **M. Saleem** and B. Kovari, "The effect of signature down-sampling on online signature verification", Automation and Applied Computer Science Workshop 2020, Budapest, Hungary, pages 253–266., 2020.

11 **M. Saleem** and B. Kovari, "Survey of Preprocessing Techniques and Classification Approaches in Online Signature Verification", International Conference on Image Analysis and Recognition, Portugal, pages 253–266., 2020.

12 **M. Saleem** and B. Kovari, "Online signature verification based on signer dependent sampling frequency and dynamic time warping", ISCMI2020, Sweden, pages 253–266., 2020.

13 **M. Saleem** and B. Kovari, “Classification approaches in online signature verification”, University of Pecs, Pecs, Hungary, pages 253–266., 2020.

14 **M. Saleem** and B. Kovari, “ Performance evaluation of the JK-nearest neighbor online signature verification parameters ”, 4th International Conference on Data Storage and Data Engineering, Barcelona, Spain, pages 1–5, 2021.

15 **M. Saleem** and B. Kovari, “Online signature verification using a combination of k-nearest neighbor and the dynamic time warping algorithms on DeepSignDB database”, Automation and Applied Computer Science Workshop 2021, Budapest, Hungary, 2021.

16 R. Tolosana et al., and **M. Saleem**, “ICDAR 2021 Competition on On-Line Signature Verification”, 16th International Conference on Document Analysis and Recognition ICDAR,Lausanne, Switzerland, 2021.

17 **M. Saleem** and B. Kovari, “K-nearest Neighbour and Dynamic Time Warping for Online Signature Verification”, SPPR, Switzerland, 2021.

18 C. Szucs, **M. Saleem**, B. Kovari, and Z. Docs, “Elofeldolgozási algoritmusok szisztematikus vizsgálata az online aláírás-hitelesítésben”, Képfeldolgozók és Alakfelsimerők társaságának 13. konferenciája, Budapest, Hungary, 2021.

Other publications

19 **M. Saleem**, “Speed Enhancement and Parallelization of Saw-Tooth Fringe Projection Technique”, Al-Yarmouk University, Irbid, Jordan, 2017.

Citations

1 H. Takci and E. Ekinci, "Two-step Authentication with the Help of a Novel Biometric", 1 st International Conference on Cyber Security and Digital Forensics (ICONSEC), Yalova, Turkey, 2021.

2 L. De Luisa et al. "In-Air 3D Dynamic Signature Recognition using Haptic Devices.", IEEE International Workshop on Biometrics and Forensics (IWBF), Rome, Italy, pages 1-6, 2021.

3 S.F. Al-Hamdan, and M.A. Bawaneh, "Simulation of intensity based triangular fringe projection technique for surface shape measurements", IEEE International Conference on Electrical and Computing Technologies and Applications (ICECTA), RAK, Emirates, pages 1-5, 2017.

4 M. Caruana Montes. Nuevos esquemas de verificación de firma manuscrita dinámica: análisis de la complejidad y fusión de sistemas , MSc thesis, (2021).

5 M. Diaz et al. "One vs. One Offline Signature Verification: A Forensic Handwriting Examiners Perspective.", Conference: 54th IEEE International Carnahan Conference on Security Technology, 2021.

References

[Alpar, 2018] Alpar, O. (2018). Online signature verification by continuous wavelet transformation of speed signals. *Expert Systems with Applications*, 104:33–42.

[Barkoula et al., 2013] Barkoula, K., Economou, G., and Fotopoulos, S. (2013). Online signature verification based on signatures turning angle representation using longest common subsequence matching. *International Journal on Document Analysis and Recognition (IJDAR)*, 16(3):261–272.

[Bibi et al., 2020] Bibi, K., Naz, S., and Rehman, A. (2020). Biometric signature authentication using machine learning techniques: Current trends, challenges and opportunities. *Multimedia Tools and Applications*, 79(1):289–340.

[Chandra et al., 2021] Chandra, S., Singh, K. K., Kumar, S., Ganesh, K., Sravya, L., and Kumar, B. P. (2021). A novel approach to validate online signature using machine learning based on dynamic features. *Neural Computing and Applications*, pages 1–20.

[Fairhurst, 1997] Fairhurst, M. (1997). Signature verification revisited: promoting practical exploitation of biometric technology. *Electronics & communication engineering journal*, 9(6):273–280.

[Fauziyah et al., 2009] Fauziyah, S., Azlina, O., Mardiana, B., Zahariah, A., and Haroon, H. (2009). Signature verification system using support vector machine. In *2009 6th International Symposium on Mechatronics and its Applications*, pages 1–4. IEEE.

[Fayyaz et al., 2015] Fayyaz, M., Hajizadeh_Saffar, M., Sabokrou, M., and Fathy, M. (2015). Feature representation for online signature verification. *arXiv preprint arXiv:1505.08153*.

[Foroozandeh et al., 2020] Foroozandeh, A., Hemmat, A. A., and Rabbani, H. (2020). Online handwritten signature verification and recognition based on dual-tree complex wavelet packet transform. *Journal of Medical Signals and Sensors*, 10(3):145–157.

[Gruber et al., 2009] Gruber, C., Gruber, T., Krinninger, S., and Sick, B. (2009). Online signature verification with support vector machines based on lcss kernel functions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(4):1088–1100.

[Hu et al., 2019] Hu, H., Chen, D., and Zheng, J. (2019). Online handwriting signature verification based on template clustering. In *Proceedings of the 2019 3rd International Workshop on Education, Big Data and Information Technology*, pages 129–135.

[Ibrahim et al., 2010] Ibrahim, M. T., Khan, M. A., Alimgeer, K. S., Khan, M. K., Taj, I. A., and Guan, L. (2010). Velocity and pressure-based partitions of horizontal and vertical trajectories for online signature verification. *Pattern Recognition*, 43(8):2817–2832.

[Jain et al., 2020] Jain, A., Singh, S. K., and Singh, K. P. (2020). Handwritten signature verification using shallow convolutional neural network. *Multimedia Tools and Applications*, pages 1–26.

[Jia et al., 2019] Jia, Y., Huang, L., and Chen, H. (2019). A two-stage method for online signature verification using shape contexts and function features. *Sensors*, 19(8):1808.

[Kholmatov and Yanikoglu, 2009] Kholmatov, A. and Yanikoglu, B. (2009). Susig: an on-line signature database, associated protocols and benchmark results. *Pattern Analysis and Applications*, 12(3):227–236.

[Lai and Jin, 2018] Lai, S. and Jin, L. (2018). Recurrent adaptation networks for on-line signature verification. *IEEE Transactions on Information Forensics and Security*, 14(6):1624–1637.

[Lai et al., 2017] Lai, S., Jin, L., and Yang, W. (2017). 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)*, volume 1, pages 400–405. IEEE.

[Liwicki et al., 2011] Liwicki, M., Malik, M. I., Van Den Heuvel, C. E., Chen, X., Berger, C., Stoel, R., Blumenstein, M., and Found, B. (2011). Signature verification competition for online and offline skilled forgeries (sigcomp2011). In *2011 International Conference on Document Analysis and Recognition*, pages 1480–1484. IEEE.

[Malik et al., 2015] Malik, M. I., Ahmed, S., Marcelli, A., Pal, U., Blumenstein, M., Alewijns, L., and Liwicki, M. (2015). Icdar2015 competition on signature verification and writer identification for on-and off-line skilled forgeries (sigwicomp2015). In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1186–1190. IEEE.

[ML, 2021] ML (2021). Ml.net an open source and cross-platform machine learning framework. https://dotnet.microsoft.com/apps/machinelearning-ai/ml-dotnet. Accessed: 2021-06-26.

[Nagel and Rosenfeld, 1973] Nagel, R. N. and Rosenfeld, A. (1973). Steps toward handwritten signature verification. In *Proc. First Int. Joint Conf. on Pattern Recognition, Washington, DC*, pages 59–66.

[Nanni and Lumini, 2008] Nanni, L. and Lumini, A. (2008). A novel local online signature verification system. *Pattern Recognition Letters*, 29(5):559–568.

[Okawa, 2019] Okawa, M. (2019). Template matching using time-series averaging and dtw with dependent warping for online signature verification. *IEEE Access*, 7:81010–81019.

[Okawa, 2020] Okawa, M. (2020). Online signature verification using single-template matching with time-series averaging and gradient boosting. *Pattern Recognition*, 102:107227.

[Parodi and Gómez, 2014] Parodi, M. and Gómez, J. C. (2014). Legendre polynomials based feature extraction for online signature verification. consistency analysis of feature combinations. *Pattern Recognition*, 47(1):128–140.

[Rashidi et al., 2012] Rashidi, S., Fallah, A., and Towhidkhah, F. (2012). Feature extraction based dct on dynamic signature verification. *Scientia Iranica*, 19(6):1810–1819.

[Riesen and Schmidt, 2019] Riesen, K. and Schmidt, R. (2019). Online signature verification based on string edit distance. *International Journal on Document Analysis and Recognition (IJDAR)*, 22(1):41–54.

[Rua and Castro, 2012] Rua, E. A. and Castro, J. L. A. (2012). Online signature verification based on generative models. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(4):1231–1242.

[Sae-Bae and Memon, 2014] Sae-Bae, N. and Memon, N. (2014). Online signature verification on mobile devices. *IEEE transactions on information forensics and security*, 9(6):933–947.

[Sharma and Sundaram, 2016a] Sharma, A. and Sundaram, S. (2016a). An enhanced contextual dtw based system for online signature verification using vector quantization. *Pattern Recognition Letters*, 84:22–28.

[Sharma and Sundaram, 2016b] Sharma, A. and Sundaram, S. (2016b). An enhanced contextual dtw based system for online signature verification using vector quantization. *Pattern Recognition Letters*, 84:22–28.

[Sharma and Sundaram, 2016c] Sharma, A. and Sundaram, S. (2016c). A novel online signature verification system based on gmm features in a dtw framework. *IEEE Transactions on Information Forensics and Security*, 12(3):705–718.

[Sharma and Sundaram, 2018] Sharma, A. and Sundaram, S. (2018). On the exploration of information from the dtw cost matrix for online signature verification. *IEEE Transactions on Cybernetics*, 48(2):611–624.

[Tang et al., 2017] Tang, L., Kang, W., and Fang, Y. (2017). Information divergence-based matching strategy for online signature verification. *IEEE Transactions on Information Forensics and Security*, 13(4):861–873.

[TFF, 2021] TFF (2021). Tensorflow federated: Machine learning on decentralized data. https://www.tensorflow.org/federated. Accessed: 2021-06-26.

[Wang et al., 2011] Wang, K., Wang, Y., and Zhang, Z. (2011). online signature verification using wavelet packet. In *2011 International Joint Conference on Biometrics (IJCB)*, pages 1–6. IEEE.

[Xia et al., 2017] Xia, X., Chen, Z., Luan, F., and Song, X. (2017). Signature alignment based on gmm for online signature verification. *Pattern Recognition*, 65:188–196.

[Xia et al., 2018] Xia, Z., Shi, T., Xiong, N. N., Sun, X., and Jeon, B. (2018). A privacy-preserving handwritten signature verification method using combinational features and secure knn. *IEEE Access*, 6:46695–46705.

[Yeung et al., 2004] Yeung, D.-Y., Chang, H., Xiong, Y., George, S., Kashi, R., Matsumoto, T., and Rigoll, G. (2004). Svc2004: First international signature verification competition. In *International conference on biometric authentication*, pages 16–22. Springer.