



SVC-onGoing: Signature verification competition

Ruben Tolosana^{a,*}, Ruben Vera-Rodriguez^a, Carlos Gonzalez-Garcia^a, Julian Fierrez^a, Aythami Morales^a, Javier Ortega-Garcia^a, Juan Carlos Ruiz-Garcia^a, Sergio Romero-Tapiador^a, Santiago Rengifo^a, Miguel Caruana^a, Jiajia Jiang^b, Songxuan Lai^b, Lianwen Jin^b, Yecheng Zhu^b, Javier Galbally^c, Moises Diaz^d, Miguel Angel Ferrer^d, Marta Gomez-Barrero^e, Ilya Hodashinsky^f, Konstantin Sarin^f, Artem Slezkin^f, Marina Bardamova^f, Mikhail Svetlakov^f, Mohammad Saleem^g, Cintia Lia Szcs^g, Bence Kovari^g, Falk Pulsmeier^h, Mohamad Wehbi^h, Dario Zanca^h, Sumaiya Ahmadⁱ, Sarthak Mishraⁱ, Suraiya Jabinⁱ

^a Biometrics and Data Pattern Analytics Lab, UAM, Spain

^b South China University of Technology, China

^c European Commission - Joint Research Centre, Italy

^d Universidad de las Palmas de Gran Canaria, Spain

^e Hochschule Ansbach, Germany

^f Tomsk State University of Control Systems and Radioelectronics, Russia

^g Budapest University of Technology and Economics, Hungary

^h Machine Learning and Data Analytics Lab, FAU, Germany

ⁱ Jamia Millia Islamia, India

ARTICLE INFO

Article history:

Received 14 September 2021

Revised 9 February 2022

Accepted 22 February 2022

Available online 24 February 2022

Keywords:

SVC-onGoing

SVC 2021

Biometrics

Handwriting

Signature verification

DeepSignDB

SVC2021_EvalDB

ABSTRACT

This article presents SVC-onGoing¹, an on-going competition for on-line signature verification where researchers can easily benchmark their systems against the state of the art in an open common platform using large-scale public databases, such as DeepSignDB² and SVC2021_EvalDB³, and standard experimental protocols. SVC-onGoing is based on the ICDAR 2021 Competition on On-Line Signature Verification (SVC 2021), which has been extended to allow participants anytime. The goal of SVC-onGoing is to evaluate the limits of on-line signature verification systems on popular scenarios (office/mobile) and writing inputs (stylus/finger) through large-scale public databases. Three different tasks are considered in the competition, simulating realistic scenarios as both random and skilled forgeries are simultaneously considered on each task. The results obtained in SVC-onGoing prove the high potential of deep learning methods in comparison with traditional methods. In particular, the best signature verification system has obtained Equal Error Rate (EER) values of 3.33% (Task 1), 7.41% (Task 2), and 6.04% (Task 3). Future studies in the field should be oriented to improve the performance of signature verification systems on the challenging mobile scenarios of SVC-onGoing in which several mobile devices and the finger are used during the signature acquisition.

© 2022 The Author(s). Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

On-line handwritten signature verification has always been a very active area of research due to its high popularity for authentication scenarios [1] and the variety of open challenges that are still under research nowadays [2], e.g., one/few-shot learning [3–5], device interoperability [6,7], aging [8,9], types of impostors [10], signature complexity [11,12], template storage [13], etc. Despite all these challenges, the performance of on-line signature verification

* Corresponding Author.

E-mail address: ruben.tolosana@uam.es (R. Tolosana).

¹ <https://competitions.codalab.org/competitions/27295>.

² <https://github.com/BiDALab/DeepSignDB>.

³ https://github.com/BiDALab/SVC2021_EvalDB.

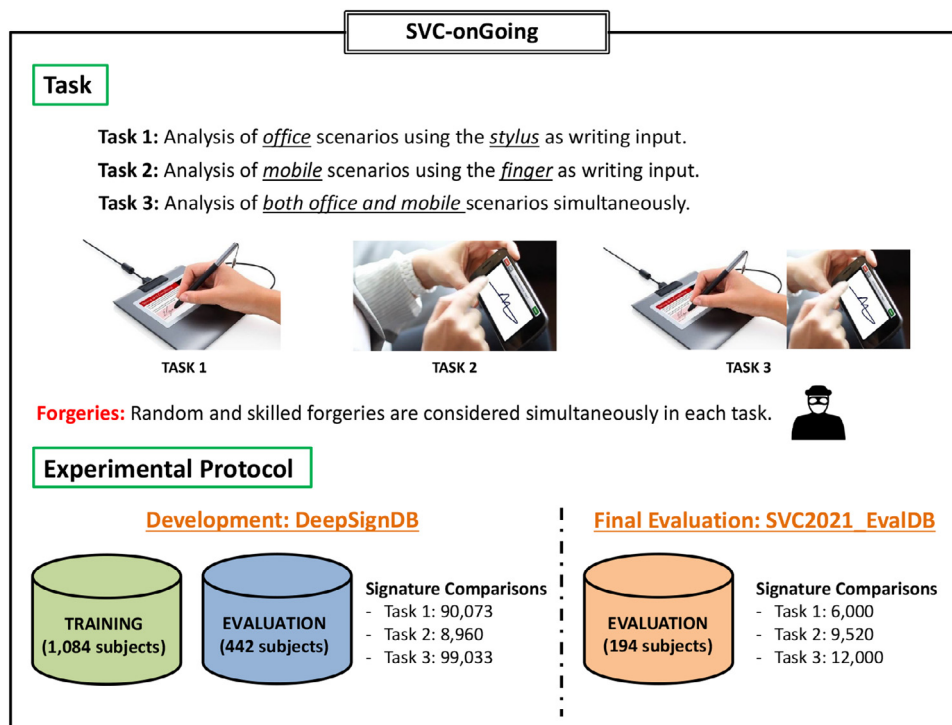


Fig. 1. Description of the tasks and experimental protocol details considered in SVC-onGoing. Two large-scale public databases are considered in the competition: DeepSignDB [17] and the novel SVC2021_EvalDB acquired for the competition.

systems has been improved in the last years due to several factors, especially: *i*) the evolution in the acquisition technology going from devices specifically designed to acquire handwriting and signature in office-like scenarios through a pen stylus (e.g. Wacom devices) to the current touch screens of mobile scenarios in which signatures can be captured anywhere using our own personal smartphone through the finger [6,14], and *ii*) the extended usage of deep learning technology in many different areas, overcoming traditional handcrafted approaches and even human performance [15–17]. So, with all these aspects in mind, the question is: what are the current performance limits of the on-line signature verification technology under realistic scenarios?

This article describes the experimental framework and results of SVC-onGoing⁴, which is based on ICDAR 2021 Competition on On-Line Signature Verification (SVC 2021) [18]. The goal of SVC-onGoing is to provide the research community with an open computing platform where researchers can easily benchmark their systems against the state of the art using public databases and common experimental protocols. Fig. 1 graphically summarises the main details of the competition. Three different tasks are considered in the competition:

- **Task 1:** Office scenarios using the stylus as input.
- **Task 2:** Mobile scenarios using the finger as input.
- **Task 3:** Both office and mobile scenarios simultaneously.

The motivations for the design of these three tasks in the competition are: *i*) covering popular office-like scenarios using the stylus as input [1], as they are deployed in several real-world applications [2], and evaluate the impact of recent deep learning techniques in this scenario compared with traditional approaches such as Dynamic Time Warping (DTW) [15]; *ii*) assess the performance of state-of-the-art signature verification systems on recent challenging mobile scenarios [6,11], in which the intra-user variability

is higher due to several factors such as the quality of the acquisition sensors (e.g., sampling frequency), the usability of the finger as writing input, etc.; and *iii*) assess novel application scenarios of the signature verification technology, for example, through the acquisition of signatures using both stylus and finger tools indifferently. In addition, we simulate in SVC-onGoing the following realistic operational conditions, to the best of our knowledge, not considered in previous on-line signature verification competitions [1]:

- Over 1700 subjects and 100 different acquisition devices are considered in the competition, using both Wacom devices (office scenarios) and general-purpose devices such as tablets and smartphones (mobile scenarios).
- Random and skilled forgeries (a.k.a. bona fide and presentation attacks in ISO/IEC 30107-1:2016 [19]) are simultaneously considered in each task. In addition, different types of skilled forgeries are considered in the competition such as static (i.e., only the image of the signature to forge is available for the attacker) and dynamic forgeries (i.e., both image and dynamics are available for the attacker), in both trained and blueprint cases [10].
- Realistic intra-subject variability across time, as different acquisition set-ups are considered in the competition ranging from 1 to 5 sessions and with a time gap between sessions from days to months, therefore enabling realistic template aging and template update studies.

This realistic scenario has been achieved thanks to the public DeepSignDB database [17] and the novel SVC2021_EvalDB database (this later one specifically acquired for this competition). Besides, we have designed realistic and challenging experimental protocols making public the corresponding signature comparisons files and the benchmarking platform.

A preliminary version of this article was published in [18]. This article significantly improves [18] in the following aspects: *i*) we provide a review of key previous on-line signature verification competitions in the literature, highlighting the motivation of SVC-onGoing; *ii*) we adapt and benchmark our recent Time-Aligned Re-

⁴ <https://competitions.codalab.org/competitions/27295>.

current Neural Network (TA-RNN) presented in [20] using the experimental framework of SVC-onGoing to serve as a representation of the state of the art for our comparative experiments; *iii*) we provide a more extensive description of the on-line signature verification systems evaluated in SVC 2021, including key figures and tables regarding the system architectures and features extracted; and *iv*) we provide an in-depth analysis of the results achieved on DeepSignDB and SVC2021_EvalDB databases of the competition. Also, we perform an independent analysis of the evaluated systems on skilled and random forgery impostors to see the robustness against different attack scenarios [10].

The remainder of the article is organised as follows. Section 2 summarises previous on-line signature verification competitions. Section 3 and 4 describe the details of the databases and the set-up considered in the competition, respectively. Section 5 provides a description of the submitted on-line signature verification systems. Section 6 describes the results of the competition. Finally, Section 7 draws the final conclusions and points out some lines for future work.

2. Previous on-line signature verification competitions

Many international on-line signature verification competitions have been organised in the last decades [1], focused on the analysis of different challenges. One of the most popular competitions is The First International Signature Verification Competition (SVC 2004) [21]. The goal of SVC 2004 was to establish common benchmark databases and benchmarking rules for the first time in the signature verification community. Two different tasks were considered in the competition. Task 1 considered only information related to the X and Y spatial coordinates of the signatures whereas Task 2 contained additional information such as pen orientation and pressure. Results in terms of Equal Error Rate (EER) were reported for each task, achieving values of 2.84% and 2.89% EERs for Task 1 and 2, respectively. Despite the valuable know-how obtained from the competition, the SVC 2004 database lacks of realistic operational conditions nowadays as: *i*) contributors were advised not to use their real signatures, but to design new ones, and *ii*) an old-fashion device was considered in the acquisition.

Another popular competition is the BioSecure Signature Evaluation Campaign (BSEC'2009) [22]. The goal of BSEC'2009 was to evaluate different on-line signature verification systems depending on the quality of the signatures. As a result, two different tasks were considered. The first one aimed at studying the influence of acquisition conditions (i.e., DS2: Wacom Intuos 3 A6 tablet, and DS3: PDA HP iPAQ hx2790) while the second one aimed at studying the impact of the information content in signatures (entropy-based quality measure). Experimental results revealed the high degradation performance on mobile devices for skilled forgeries, i.e., 2.2% EER for DS2 vs. 4.97% EER for DS3.

Also in 2009, Blankers *et al.* organised the ICDAR 2009 Signature Verification Competition [23]. The goal of the competition was to combine realistic forensic casework with automated methods by testing systems on a forensic-like new dataset. As a result, both on- and off-line signature verification scenarios were considered in the competition. The best results achieved in the competition were 2.85% and 9.15% EERs for the on- and off-line methods, respectively. In addition, the off-line results achieved by the automatic systems were compared to Forensic Handwriting Experts (FHEs) of the Netherlands Forensic Institute (NFI) who achieved incorrect conclusion in 3.13% of the verifications using only the image of the signatures. However, as the authors commented in the paper, the comparison between automatic systems and FHEs is not straightforward as different metrics and conditions are considered. In order to solve this problem, in SigComp2011 [24] the authors introduced the usage of the likelihood ratio as a performance metric.

In addition, both Western and Chinese signatures were considered in the SigComp2011 competition.

Finally, the latest on-line signature verification competitions (e.g., SASIGCOM 2020 [25]) have been focused on the analysis of the system performance with emphasis on handwritten signatures acquired in different countries such as the Dutch, Japanese, Bengali, Italian, German, and Thai databases.

Despite the massive efforts carried out by the research community in the organization of previous international competitions, and the valuable know-how obtained from them, there is still one critical aspect that must be tackled in order to move forward in the signature verification field: the proposal of a signature verification benchmark based on the usage of large-scale publicly available databases, common experimental protocols, and an open internet platform that all researchers can easily use to compare their results with the state of the art. This is the key contribution of SVC-onGoing⁵.

3. SVC-ongoing: Databases

Two databases are considered in SVC-onGoing: DeepSignDB and SVC2021_EvalDB. These databases are publicly available for the research community and can be downloaded following the instructions included in^{6,7}. We provide next a description of them.

3.1. DeepSignDB

The DeepSignDB database [17] comprises on-line signatures from a total of 1526 subjects from four different well-known databases: MCYT (330 subjects) [26], BiosecrID (400 subjects) [27], Biosecure DS2 (650 subjects) [22], eBioSign (65 subjects) [14], and a novel signature database composed of 81 subjects. DeepSignDB comprises more than 70K signatures acquired using both stylus and finger writing inputs in both office and mobile scenarios. A total of 8 different devices are considered in the acquisition (i.e., 5 Wacom devices and 3 Samsung general purpose devices). In addition, different types of impostors and number of acquisition sessions are considered along the database. For more details about DeepSignDB, we refer the reader to the published article [17].

3.2. SVC2021_EvalDB

The SVC2021_EvalDB database is a novel database specifically acquired in SVC 2021. Two acquisition scenarios are considered:

- **Office scenario:** on-line signatures from 75 total subjects were acquired using a Wacom STU-530 device with the stylus as writing input. This acquisition was carried out using members of Universidad Autonoma de Madrid. Regarding the acquisition protocol, the device was placed on a desktop and subjects were able to rotate it in order to feel comfortable with the writing position. It is important to highlight that the subjects considered in the acquisition of SVC2021_EvalDB are different compared to the ones considered in the DeepSignDB database. Signatures were collected in two separated sessions with a time gap between them of at least 1 week. For each subject, there are 8 total genuine signatures (4 signatures/session) and 16 skilled forgeries (8 signatures/type) performed by 4 different subjects in 2 different sessions. Regarding the skilled forgeries, both static and dynamic forgeries were considered in the first

⁵ <https://competitions.codalab.org/competitions/27295>.

⁶ <https://github.com/BiDALab/DeepSignDB>.

⁷ https://github.com/BiDALab/SVC2021_EvalDB.

and second acquisition sessions, respectively. Information related to X and Y spatial coordinates, pressure, and timestamp is recorded for the Wacom device. In addition, pen-up trajectories are also available.

- **Mobile scenario:** on-line signatures from 119 total subjects were acquired using the same acquisition framework considered in MobileTouchDB [20]. It is important to highlight that the acquisition of this mobile dataset was carried out before the acquisition of the office dataset, considering a different set of users compared with the office scenario as subjects did not need any special device for the acquisition, just their own smartphone. Regarding the acquisition protocol, we implemented an Android App and uploaded it to the Play Store in order to study an unsupervised mobile scenario at international level. This way all subjects could download the App and use it on their own devices without any kind of supervision, simulating a practical scenario in which subjects can generate touch-screen on-line signatures in any possible scenario, e.g., standing, sitting, walking, indoors, outdoors, etc. As a result, 94 different smartphone models from 16 different brands were used during the acquisition.

Regarding the acquisition protocol, between 4 and 6 separated sessions in different days were considered with a total time gap between the first and last session of at least 3 weeks. For each subject, there are at least 8 total genuine signatures (2 signatures/session) and 16 skilled forgeries (8 signatures/type) performed by 4 different subjects. Regarding the skilled forgeries, both static and dynamic forgeries were considered, similar to the office scenario. Information related to X and Y spatial coordinates, and timestamp is recorded for all devices. Pen-up information is not available in this case.

4. SVC-onGoing: Competition Set-Up

4.1. Tasks

The goal of SVC-onGoing is to evaluate the limits of on-line signature verification systems on popular scenarios (office/mobile) and writing inputs (stylus/finger) through large-scale public databases. As a result, the following three tasks are considered in the competition:

- **Task 1:** Office scenarios using the stylus as input.
- **Task 2:** Mobile scenarios using the finger as input.
- **Task 3:** Both office and mobile scenarios simultaneously.

In addition, SVC-onGoing simulates realistic operational conditions considering random and skilled forgeries simultaneously in each task.

4.2. Public web platform

A public web platform is set up for the SVC-onGoing competition⁸. Participants can test their on-line signature verification systems on each task using the signature comparisons files provided (see Section 4.4) to obtain the scores and update them to the web platform, which automatically computes their EER performance on the different tasks. Results are updated in the platform in real time, and they are visible to everyone in a ranking dashboard. Additional information of the submitted systems such as the individual EER of skilled and random forgery scenarios are shown in the official web of the competition⁹.

In addition, we want to highlight that participants/teams can test published/unpublished approaches in the web platform without any restrictions. In fact, one of the main motivations of this competition is to facilitate the research community the comparison of novel approaches with public and popular benchmarks before publishing them. This way the revision process would be much fairer compared to the actual practice in which each paper uses or proposes a different experimental protocol, which makes it difficult to compare to previous works.

Finally, to guarantee the quality of the competition, the organizers of SVC-onGoing might need to verify the truthfulness of the scores submitted if necessary, asking the participants/teams for more details.

4.3. Evaluation criteria

The SVC-onGoing competition follows a ranking based on points. Each task is evaluated separately, having three winners with their corresponding points (gold medal: 3, silver medal: 2, and bronze medal: 1). The participant/team that gets more points in total (Task 1, 2, and 3) in the final evaluation stage of the competition is the actual winner of SVC-onGoing. It is important to highlight the following aspects of the competition: *i*) participants/teams can test their signature verification approaches on specific tasks of the competition, it is not required to test them on all three tasks, *ii*) there are no restrictions on the number of signature verification approaches, *iii*) there are no restrictions on the number of scores submissions to the web platform, *iv*) the point evaluation is carried out at team level, not signature verification approach. These aspects have been defined to motivate the use of the platform by the research community.

The evaluation metric considered is the popular Equal Error Rate (%) similar to most on-line signature verification studies in the literature.

4.4. Experimental protocol

The two following stages are considered in SVC-onGoing:

- **Development:** the goal of this stage is to provide the participants with the data needed to train the on-line signature verification systems. **Only the DeepSignDB database** is provided to the participants in this stage of the competition. In addition, participants can freely use other databases to train their systems.

In order to allow the participants to test their trained systems under similar conditions considered in the final evaluation stage of the competition, we divide the DeepSignDB database into training and evaluation datasets. The training dataset is based on 1084 subjects whereas the evaluation dataset comprises the remaining 442 subjects of the database. For the training of the systems (1,084 subjects), no instructions are given to the participants. They can use the data as they like. Nevertheless, for the evaluation of the systems (442 subjects), we provide the participants with the signature comparisons to run (without ground-truth labels). This way participants can obtain a quantitative measure of the performance of the developed systems for the final evaluation stage of the competition.

- **Final Evaluation:** the final evaluation of SVC-onGoing is carried out using **only the novel SVC2021_EvalDB database**. This database comprises a different set of subjects not considered in the development database of the competition. This way, the proposed experimental protocol can also evaluate the robustness of the signature verification systems against unseen users. The database together with the corresponding signature comparisons files (one file per task) are sent to the participants after signing the corresponding license agreement. It is important

⁸ <https://competitions.codalab.org/competitions/27295>.

⁹ <https://sites.google.com/view/SVC2021/home>.

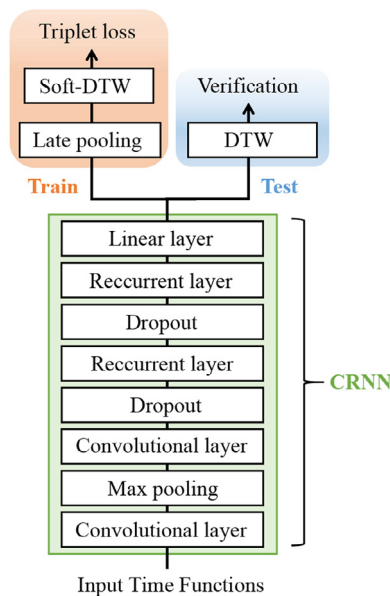


Fig. 2. DLVC-Lab team: Framework of the deep soft-DTW (DSDTW) model.

to highlight that all signatures are included in a single folder, and both the nomenclature of the signatures and the signature comparisons files are randomized to avoid cheating. Ground-truth labels are not provided to the participants. In addition, and in order to consider a very challenging impostor scenario, the skilled forgery comparisons included in the corresponding files are optimised using machine learning methods, selecting only the best high-quality forgeries.

Finally, the Task 3 of the competition (analysis of office and mobile scenarios together) has been designed combining a balanced set of signature comparisons from Tasks 1 (office scenario) and 2 (mobile scenario).

5. SVC-onGoing: Description of systems

A total of 56 participants/teams have registered in SVC-onGoing. However, only 6 teams have finally submitted their scores so far with a total of 12 different on-line signature verification systems. Next, we describe briefly the systems provided by each of the teams of the competition.

5.1. DLVC-Lab team

The DLVC-Lab team is composed of members of the South China University of Technology and the Guangdong Artificial Intelligence and Digital Economy Laboratory.

The DLVC-Lab team proposed an end-to-end trainable deep soft-DTW (DSDTW) model, which greatly enhances the classical DTW method with the capability of deep representation learning. In particular, they use neural networks to learn deep time functions as inputs for DTW. As DTW is not fully differentiable with regards to its inputs, they introduce its smoothed formulation, soft-DTW [28], and incorporate the soft-DTW distances of signature pairs into a triplet loss function for optimization. As soft-DTW is differentiable, the entire system is end-to-end trainable and achieves a perfect integration of neural networks and DTW. Fig. 2 provides a graphical representation of the framework proposed by the DLVC-Lab team.

Three different approaches have been submitted to SVC-onGoing. System 1 is based on Convolutional Recurrent Neural Networks (CRNN) whereas System 2 and 3 are based on fully Con-

Table 1
DLVC-Lab team: CRNN architecture.

Layer	Parameters
Convolution	$c = 64, k = 7, s = 1, p = 3$
Max Pooling	$k = 2, s = 2$
Convolution	$c = 128, k = 3, s = 1, p = 1$
Dropout	prob = 0.1
Recurrent	128 GARUs [16]
Dropout	prob = 0.1
Recurrent	128 GARUs
Linear	$c = 64$

Table 2
DLVC-Lab team: Set of 12 time functions used in the on-line signature verification approach [5].

#	Feature
1-2	First-order derivatives of X- and Y-coordinates: \dot{x}_n, \dot{y}_n
3	Velocity magnitude: $v_n = \sqrt{\dot{x}_n^2 + \dot{y}_n^2}$
4	Path-tangent angle: $\theta_n = \arctan(\dot{y}_n/\dot{x}_n)$
5-7	$\cos(\theta_n), \sin(\theta_n),$ and pressure z_n
8-9	First-order derivatives of v_n and θ_n : $\dot{v}_n, \dot{\theta}_n$
10	Log curvature radius: $\rho_n = \log(v_n/\dot{\theta}_n)$
11	Centripetal acceleration: $c_n = v_n \cdot \dot{\theta}_n$
12	Total acceleration: $a_n = \sqrt{\dot{v}_n^2 + c_n^2}$

volutional Neural Networks (CNN). The detailed CRNN architecture parameters of System 1 is described in Table 1, where $c, k, s, p,$ and $prob$ denote channels, kernel size, stride, padding size, and probability, respectively. The convolutional layers are followed by ReLU activation. The fully CNN architecture (Systems 2 and 3) is obtained by replacing the two recurrent layers with two convolutional layers of the same size. Systems 2 and 3 only differ in the training data. Concretely, Systems 1 and 2 use the development set of the DeepSignDB database for training (1,084 subjects), including both stylus-written and finger-written signatures. System 3 uses only finger-written signatures for training and, therefore, it is only tested on Task 2 (finger input scenario), unlike Systems 1 and 2 that are tested on all three tasks of the competition.

Regarding the preprocessing and time function extraction, the timestamp information of each signature file is considered to estimate the sampling rate and to resample the signature at about 100 Hz. After that, the 12 time functions described in Table 2 are extracted for each signature.

These time functions are fed to DSDTW. Except pressure z for finger-written signatures, each time function is normalised to have zero mean and unit variance. Pressure z for finger-written signatures is set as constant 1.0.

Finally, regarding the training process, each system is trained for 20 epochs using stochastic gradient descent with Nesterov momentum. The momentum is a constant 0.9. The learning rate is initially 0.01 and exponentially decays by 0.9 after each epoch. The evaluation set of the DeepSignDB database (442 subjects) is used to pick out the best model.

5.2. SIG Team

The Spanish-Italian-German (SIG) team is composed of members of the European Commission (Italy), Universidad de las Palmas de Gran Canaria (Spain), and Hochschule Ansbach (Germany).

The signature verification system presented is based on the main principle laid out in [29]: the generation of synthetic off-line signatures from the real on-line samples and the fusion of both types of data can lead to the overall improvement of the on-line verification performance.

Following that rationale, the system submitted is based on the combination of on- and off-line signature information. A general

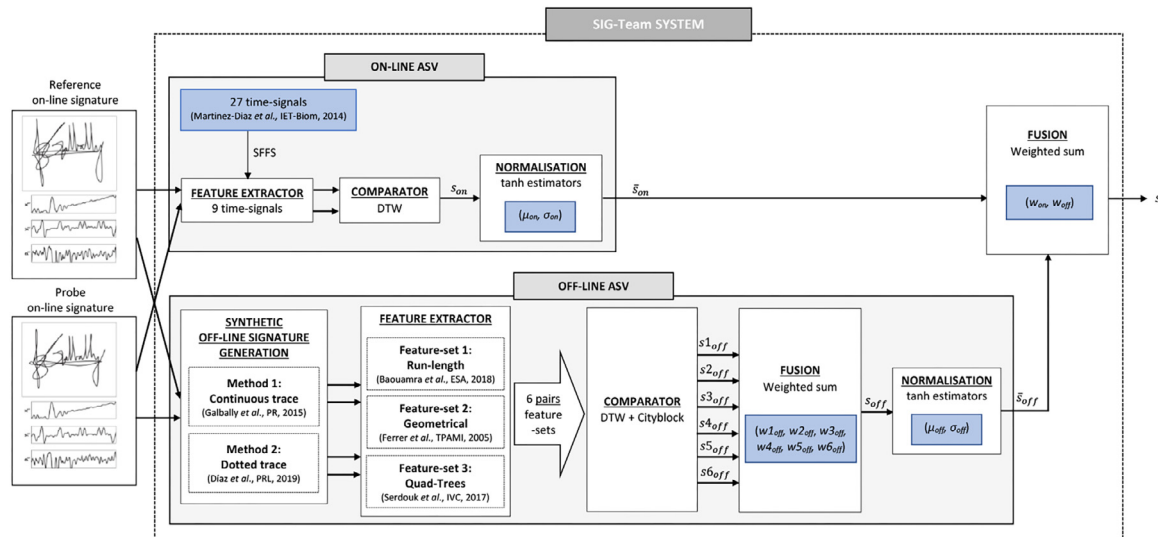


Fig. 3. SIG team: Diagram of the system submitted based on the combination of on- and off-line signature information. Blue-shaded boxes show the parameters optimised during the development stage of the competition using DeepSignDB [17].

Table 3

SIG team: Subset of 9 time functions used in the on-line signature verification approach. The features numbering and description follow that of Table 2 in [30].

#	Feature
1-2	X- and Y-coordinates: x_n, y_n
5	Path velocity magnitude: $v_n = \sqrt{y_n^2 + x_n^2}$
10, 12	First-order derivatives of pen pressure z_n and v_n : z'_n, v'_n
13	First-order derivative of path-tangent angle: $\Theta_n = \arctan(y'_n/x'_n)$
21	Ratio of the minimum and the maximum speed over a window of 5 samples: $v'_n = \min\{v_{n-4}, \dots, v_n\} / \max\{v_{n-4}, \dots, v_n\}$
23	First order derivative of the angle of consecutive samples: $\alpha_n = \arctan((y_n - y_{n-1}) / (x_n - x_{n-1}))$
25	Cosine of the angle of consecutive samples: $c_n = \cos(\alpha_n)$

diagram of the system is shown in Fig. 3, where the parameters optimised during the development stage of the competition are highlighted in blue.

The **on-line signature approach** is based on local features and the well-known DTW algorithm. In particular, the system is based on a subset of the initial 27 time functions introduced in [30] and selected using the Sequential Floating Forward Selection (SFFS) algorithm. Table 3 provides a description of the 9 time functions selected for the task. The specific implementation of the DTW algorithm uses the Euclidean Distance to compute the optimal path in between signatures and outputs as score s_{on} the last value of the optimal path, normalised by the path length. For the cases where pressure p is not available (i.e., mobile scenario in Task 2 of the competition), that time signal is simply discarded, together with any other time function derived from it.

Regarding the **off-line signature approach**, the first step performed is the generation of the synthetic off-line data starting from the real on-line signatures. Two different methods are used for this purpose: *i*) continuous trace [29], and *ii*) dotted trace [31]. Once the two synthetic off-line signatures are created (for each dynamic signature given as input), three different handcrafted features are extracted: *i*) run-length distribution [32], *ii*) geometrical features [33], and *iii*) quad-tree implementation of histogram of templates [34].

The score for each of the three feature sets is obtained by comparing the reference and probe vectors using the DTW algorithm followed by the cityblock distance. This process leads to six off-line intermediate scores ($s1_{off}, s2_{off}, \dots, s6_{off}$) for each on-line comparison in the competition (recall that each individual on-line signature is converted to two off-line synthetic signatures, defined by three feature sets).

The six intermediate scores obtained by the off-line approach are finally fused into one unique off-line score s_{off} using a weighted sum. The weights for the fusion are empirically calculated on the training databases of the competition optimising the EER for each of the tasks.

Finally, the on- and off-line scores (s_{on} and s_{off}) are normalised to the [0,1] range using the tanh-estimators and fused into the final score s given as output by the system based on the weighted sum.

Only the DeepSignDB database provided in the development stage of SVC 2021 was considered for training and evaluating the system.

5.3. TUSUR KIBEVS Team

The TUSUR KIBEVS team is composed of members of the Tomsk State University of Control Systems and Radioelectronics.

The on-line signature verification system presented is based on the use of global features and a gradient boosting classifier. Fig. 4 graphically summarises the approach considered. First, signatures are normalised in terms of position, rotation, and size according to the procedures described in [35]. A set of 100 global features is extracted for each enrolled and test signatures ($F_{enrolled}$ and F_{test}) based on previous approaches in the literature [30]. Then, a new feature vector F is obtained based on the subtraction of the previous enrolled and test feature vectors: $F = |F_{enrolled} - F_{test}|$. The resulting feature vector F is introduced to CatBoost, a fast, scalable, and high performance Gradient Boosting on Decision Trees (GBDT) that is available as an open source library¹⁰.

¹⁰ <https://catboost.ai/>.

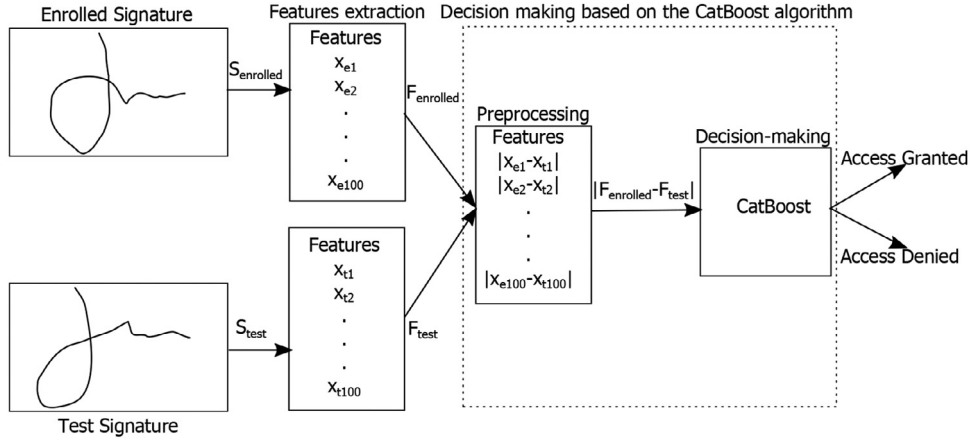


Fig. 4. TUSUR KIBEVS team: Description of the signature verification system based on global features and CatBoost.

Regarding the training procedure, only the DeepSignDB database provided in the development stage of SVC-onGoing is considered. A total of 10K signature comparisons are randomly selected (5K genuine and 5K forgeries), considering both office (stylus) and mobile (finger) scenarios simultaneously. Forgery comparisons included 2.5K skilled forgeries and 2.5K random forgeries.

5.4. SigStat team

The SigStat team is composed of members of the Budapest University of Technology and Economics.

Three different on-line signature verification systems were presented. All of them are implemented using the SigStat framework¹¹. First, all signatures go through a preprocessing stage. Time samples with zero pressure are removed from the stylus-based signatures to reduce noise and remove some artifacts. Finally, X, Y, and pressure information are scaled to the [0,1] range and shifted by the average of their values. After this preprocessing stage, the biometric information is used to calculate different distance scores between signature pairs, considering three different approaches.

The first system considers local thresholds to detect whether the query signature is genuine or forgery. In particular, it uses DTW to calculate signature distances and the k-Nearest Neighbours (k-NN) approach to set a lower and an upper threshold for each reference signature. During the development stage, the system is tested on the evaluation subset of the DeepSignDB (442 users). The distances and comparisons between the signatures are used to calculate and tune several parameters, selecting the optimal values of the genuine G_{th} and forgery F_{th} thresholds and a scaling parameter s for the classification purpose.

For testing, the distance d between the questioned signature (S_q) and the reference signature (S_r) is obtained using DTW. The final score P_q is calculated as follows:

$$P_q = \frac{s \cdot F_{th} - d}{s \cdot F_{th} - G_{th}} \quad (1)$$

The second system considers global thresholds and is based on 4 classifiers and a linear fusion of them. The first three classifiers take advantage of global features such as the standard deviation of X and Y spatial coordinates, and the signing time duration. The last classifier is based on the DTW distance of signature pairs.

In the development stage, the evaluation subset of DeepSignDB is used to make genuine-genuine and genuine-forgery comparisons. For each comparison, the calculated DTW distance, the de-

vice input, and the expected prediction are stored. Next, the comparisons and their results are sorted into four different groups based on expected prediction and input device (genuine finger, genuine stylus, forgery finger, and forgery stylus). For each group some statistical parameters such as the minimum and median values are calculated and used to set the global thresholds for the system.

For testing, the score of the questioned signature P_q is calculated based on the DTW distance of the reference-questioned pair d , the minimum distance of genuine comparisons $d_{g_{min}}$ and the median distance of forgery comparisons $d_{f_{med}}$:

$$P_q = 1 - \frac{d_{f_{med}} - d}{d_{f_{med}} - d_{g_{min}}} \quad (2)$$

In case of $d < d_{g_{min}}$, the score P_q is automatically 0 and when $d > d_{f_{med}}$ is 1. A similar approach is considered for the remaining three classifiers based on global features.

Finally, the third system extends the set of global features considered in the second system, for example including the DTW distance as feature. Contrary to previous systems, a gradient boosting classifier (XGBoost) is considered for the final prediction.

5.5. MaD-Lab Team

The MaD-Lab team is composed of members of the Machine Learning and Data Analytics Lab (FAU).

The proposed system consists of a 1D CNN trained to classify pairs of signatures as matching or not matching. Features are extracted using a mathematical concept called *path signature* together with statistical features. These features are then used to train an adapted version of ResNet-18 [36].

Regarding the preprocessing stage, the X and Y spatial coordinates are normalised to a [-1, 1] range whereas the pressure information to [0,1]. In case that no pressure information is available (Task 2, mobile scenario), a vector with all one values is considered.

For the feature extraction, the global features included in Table 4 are extracted for each signature. Besides, additional features are extracted using the signature path method [37]. This is a mathematical tool that extracts features from paths. It is able to encode linear and non-linear features from the signature path. The path signature method is applied over the raw X and Y spatial coordinates, their first-order derivatives, the perpendicular vector to the segment, and the pressure.

Finally, for classification, a 1D adapted version of the ResNet-18 CNN is considered. To adapt the ResNet-18 image version, every 2D operation is exchanged with a 1D one. Also, a sigmoid activation

¹¹ <http://www.sigstat.org>.

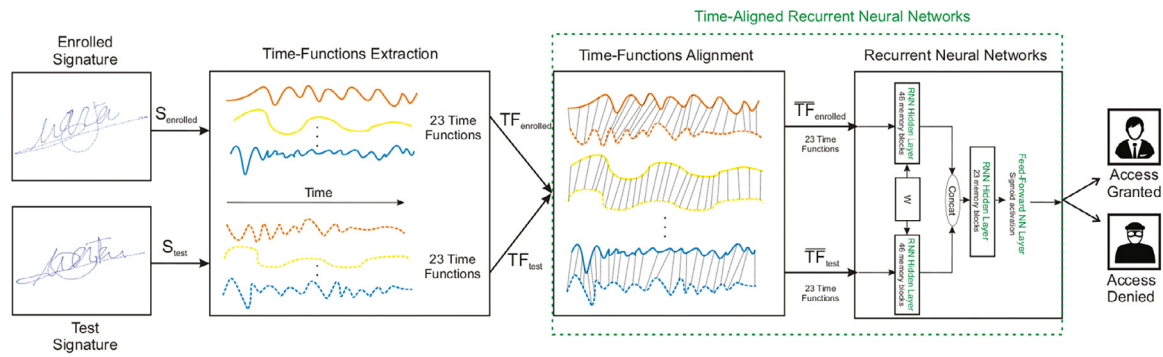


Fig. 5. BiDA-Lab team: Architecture of the on-line signature verification system based on Time-Aligned Recurrent Neural Networks. S denotes one signature sample, and TF and \overline{TF} the original and pre-aligned 23 time functions [30], respectively. The details of the Recurrent Neural Networks block are included in Section 5.7. Diagram taken from [17].

Table 4
MaD-Lab team: Set of global features used in the on-line signature verification approach.

#	Feature
1	Number of time steps
2	Percentage of coordinates with x -values larger than 0
3	Percentage of coordinates with x -values smaller than 0
4	Percentage of coordinates with y -values larger than 0
5	Percentage of coordinates with y -values smaller than 0
6	Mean of the x -values of the coordinates
7	Mean of the y -values of the coordinates
8	Median of the x -values of the coordinates
9	Median of the y -values of the coordinates
10	Standard deviation of the x -values of the coordinates
11	Standard deviation of the y -values of the coordinates
12	Skewness of the x -values of the coordinates
13	Skewness of the y -values of the coordinates

function is added in the last layer to output values between 0 and 1. Pairs of signatures are presented to the network as two different channels.

Regarding the training parameters of the network, binary cross-entropy is used as the loss function. The network is optimised using stochastic gradient descent (SGD) with a momentum of 0.9 and a learning rate of 0.001. The learning rate is decreased by a factor 0.1 if the accumulated loss in the last epoch is larger than the epoch before. In case the learning rate drops to 10^{-6} , the training process is stopped. Also, if the learning rate does not decrease below 10^{-6} , the training process is stopped after 50 epochs.

5.6. JAIRG Team

The JAIRG team is composed of members of the Jamia Millia Islamia.

Three different systems were presented, all of them focused on Task 2 (mobile scenarios). The on-line signature verification systems considered are based on an ensemble of different deep learning models training with different sets of features. The ensemble is formed using a weighted average of the scores provided by five individual systems. The specific weights to fuse the scores in the ensemble approach are obtained using a Genetic Algorithm (GA) [38].

For the feature extraction, three different approaches are considered: *i*) a set of 18 time functions related to X and Y spatial coordinates [7], *ii*) a subset of 40 global features [30], and *iii*) a set of global features extracted after applying 2D Discrete Wavelet Transform (2D-DWT) over the image of the signatures.

For classification, Bidirectional Gated Recurrent Unit (BGRU) models with a Siamese architecture are considered [15]. Different models are studied varying the number of hidden layers, input features, and training parameters. Finally, an ensemble of the best

BGRU models in the evaluation of DeepSignDB is considered, selecting the fusing weight parameters through a GA.

5.7. BiDA-Lab team

The BiDA-Lab team is composed of members of the Universidad Autonoma de Madrid. Although they did not participate in the ICDAR SVC 2021 competition (they were the organizers), it is very interesting now to benchmark their signature technology using the same experimental framework of SVC-onGoing. In particular, they consider the same signature verification system presented in [17] based on Time-Aligned Recurrent Neural Network (TA-RNN) [20]. Fig. 5 provides a graphical representation of that architecture.

For the input of the system, they feed the network with 23 time functions extracted from the signature [30]. Information related to the azimuth and altitude of the pen angular orientation is not considered. The TA-RNN architecture is based on two consecutive stages: *i*) time sequence alignment through DTW, and *ii*) feature extraction and matching using a RNN. The RNN system comprises three layers. The first layer is composed of two Bidirectional Gated Recurrent Unit (BGRU) hidden layers with 46 memory blocks each, sharing the weights between them. The outputs of the first two parallel BGRU hidden layers are concatenated and serve as input to the second layer, which corresponds to a BGRU hidden layer with 23 memory blocks. Finally, a feed-forward neural network layer with a sigmoid activation is considered, providing an output score for each pair of signatures.

This learning model was presented in [17] and has been re-trained here for SVC-onGoing¹² adapted to the stylus scenario by using only the stylus-written signatures of the development set of DeepSignDB (1,084 users). The best model has been then selected using a partition of the development set of DeepSignDB, leaving out of the training the DeepSignDB evaluation set (442 users).

6. SVC-Ongoing: Experimental results

This section analyses the results achieved by the participants in the development and evaluation stages of SVC-onGoing.

6.1. Development results: DeepSignDB

In this first stage of the competition, participants can test the performance of their systems using the evaluation dataset (442 subjects) of DeepSignDB [17] through the public SVC-onGoing web platform¹³. Table 5 shows the results achieved in each of the three

¹² <https://competitions.codalab.org/competitions/27295>.

¹³ <https://competitions.codalab.org/competitions/27295>.

Table 5
Development results of SVC-onGoing using the evaluation dataset (442 users) of DeepSignDB [17].

Task 1: Office Scenario			Task 2: Mobile Scenario			Task 3: Office/Mobile Scenario		
Position	Team	EER(%)	Position	Team	EER(%)	Position	Team	EER(%)
1	DLVC-Lab	3.32%	1	SigStat	5.81%	1	DLVC-Lab	4.18%
2	BiDA-Lab	4.31%	2	DLVC-Lab	6.58%	2	BiDA-Lab	5.01%
3	SIG	5.35%	3	SIG	9.43%	3	SigStat	7.71%
4	TUSUR KIBEVS	7.19%	4	Baseline DTW	10.16%	4	TUSUR KIBEVS	7.77%
5	Baseline DTW	7.71%	5	BiDA-Lab	11.25%	5	Baseline DTW	7.91%
6	SigStat	7.74%	6	TUSUR KIBEVS	12.68%	6	MaD	13.63%
7	MaD	14.57%	7	JAIRG	12.86%			
			8	MaD	13.36%			

Table 6
Final evaluation results of SVC-onGoing using the novel SVC2021_EvalDB database.

Task 1: Office Scenario			Task 2: Mobile Scenario			Task 3: Office/Mobile Scenario		
Points	Team	EER(%)	Points	Team	EER(%)	Points	Team	EER(%)
3	DLVC-Lab	3.33%	3	DLVC-Lab	7.41%	3	DLVC-Lab	6.04%
2	BiDA-Lab	4.08%	2	BiDA-Lab	8.67%	2	BiDA-Lab	7.63%
1	TUSUR KIBEVS	6.44%	1	SIG	10.14%	1	SIG	9.96%
0	SIG	7.50%	0	SigStat	13.29%	0	TUSUR KIBEVS	11.42%
0	MaD	9.83%	0	TUSUR KIBEVS	13.39%	0	MaD	14.21%
0	SigStat	11.75%	0	Baseline DTW	14.92%	0	SigStat	14.48%
0	Baseline DTW	13.08%	0	MaD	17.23%	0	Baseline DTW	14.67%
			0	JAIRG	18.43			

Table 7
Global ranking of SVC-onGoing.

Position	Team	Technology Description	Total Points
1	DLVC-Lab	Deep Soft-DTW: Deep learning systems enhanced with classical DTW	9
2	BiDA-Lab	TA-RNN: Deep learning system enhanced with classical DTW	6
3	SIG	DTW system: Fusion of real on-line and synthetic off-line signature information	2
4	TUSUR KIBEVS	System based on global features and gradient boosting classifier (CatBoost)	1
5	SigStat	Systems based on DTW, K-NN, and gradient boosting classifier (XGBoost)	0
6	MaD	Deep learning system trained with mathematical and statistical features	0
7	JAIRG	Deep learning system trained with multiple feature approaches	0

tasks. A Baseline DTW system (similar to the one described in [7]) based on X , Y spatial time signals, and their first- and second-order derivatives is included in the table for a better comparison of the results.

First, in all tasks we can see that the three best systems submitted to SVC-onGoing have outperformed the traditional Baseline DTW. In general, four different teams dominate the development stage of the competition: DLVC-Lab, BiDA-Lab, SigStat, and SIG. For Task 1, focused on the analysis of office scenarios using the stylus as writing input, the deep learning approach presented by the DLVC-Lab team has outperformed the other approaches with a 3.32% EER. The second and third positions are achieved by the BiDA-Lab and SIG teams with 4.31% and 5.35% EERs, respectively.

Regarding Task 2, focused on mobile scenarios using the finger as writing input, a considerable system performance degradation is generally observed compared to the best results of Task 1 (e.g., 3.32% vs. 5.81% EER for the DLVC-Lab). It is interesting to highlight the results achieved by the SigStat team, as contrary to the other teams, they have been able to obtain much better results for Task 2 than for Task 1 (5.81% vs. 7.74% EER), achieving the first position in the ranking of Task 2. In addition, the deep learning system proposed by BiDA-Lab has achieved a 11.25% EER, much worse compared with the results of Task 1 (4.31% EER). This result proves the bad generalisation of the stylus model to the finger scenario (Task 2) as the model considered was trained using only signatures acquired through the stylus, not the finger.

Finally, in general, most signature verification approaches seem to achieve good results in Task 3 when both office (Task 1) and mobile (Task 2) scenarios are evaluated together. For example, if

we analyse the approach presented by DLVC-Lab, the results are 3.32%, 5.81%, and 4.18% for the Tasks 1, 2, and 3, respectively. These results prove the good generalisation ability of the approaches against different acquisition scenarios, office and mobile, and writing inputs, stylus and finger.

6.2. Evaluation results: SVC2021_EvalDB

This section describes the final evaluation results of the competition using the novel SVC2021_EvalDB database acquired for the competition. It is important to highlight that the winner of SVC-onGoing is based only on the results achieved in this stage of the competition as described in Section 4.3. Tables 6 and 7 show the results achieved by the participants in each of the three tasks, and the current ranking of SVC-onGoing based on the total points, respectively. Similar to the previous section, we include in Table 6 a Baseline DTW system (similar to the one described in [7]) based on X , Y spatial coordinates, and their first- and second-order derivatives for a better comparison of the results.

As can be seen in Tables 6 and 7, the DLVC-Lab team is the current winner of SVC-onGoing (9 points), followed by the BiDA-Lab (6 points), SIG (2 points), and TUSUR KIBEVS (1 point) teams. The on-line signature verification systems proposed by the DLVC-Lab team achieve the best results in all three tasks. Nevertheless, these results are very close to the TA-RNN approach presented by BiDA-Lab, despite of the fact that just a single model trained for the stylus scenario (Task 1) is considered. In particular, there is an EER absolute difference of 0.75%, 1.26%, and 1.59% in each of the tasks of the competition. Also, it is interesting to compare the best

results achieved in each task with the results obtained using traditional approaches in the field (Baseline DTW). Concretely, for each of the tasks, the DLVC-Lab team achieves relative improvements of 74.54%, 50.34%, and 58.3% EER compared to the Baseline DTW. These results prove the high potential of deep learning approaches such as DSDTW and TA-RNN for the on-line signature verification field, as commented in previous studies [4,5,17].

Other approaches not based on deep learning, like the ones presented by the SIG team that uses on- and off-line signature information, have provided very good results, achieving points in most tasks. The same happens with the system proposed by the TUSUR KIBEVS team based on global features and a gradient boosting classifier (CatBoost). In particular, the approach presented by TUSUR KIBEVS has outperformed the approach proposed by the SIG team for the office scenario (6.44% vs. 7.50% EER). Nevertheless, much better results are obtained by the SIG team for the mobile and office/mobile scenarios (10.14% and 9.96% EERs) compared to the TUSUR KIBEVS results (13.39% and 11.42% EERs).

Another key aspect to analyse in SVC-onGoing is the generalisation ability of the proposed systems against new users and acquisition conditions (e.g., new devices). This analysis is possible as different databases are considered in the development and final evaluation of the competition. Tables 5 and 6 show the results achieved using the DeepSignDB and SVC2021_EvalDB databases, respectively. For Task 1, we can observe the good generalisation ability of the DLVC-Lab and BiDA-Lab systems based on deep learning, achieving results of 3.32% and 4.31% EERs for the development, and 3.33% and 4.08% EERs for the evaluation, respectively. In addition, good generalisation results are provided by the TUSUR KIBEVS and MaD teams. Regarding Task 2, it is interesting to note the poor generalisation ability showed by the SigStat system. In particular, they achieve the first position in the development stage with a 5.81% EER, but this result has increased to a final 13.29% EER in the final evaluation, achieving the fourth position. As a result, the DLVC-Lab, BiDA-Lab, and SIG teams have achieved the best generalisation results for Task 2. Similar trends are observed in Task 3.

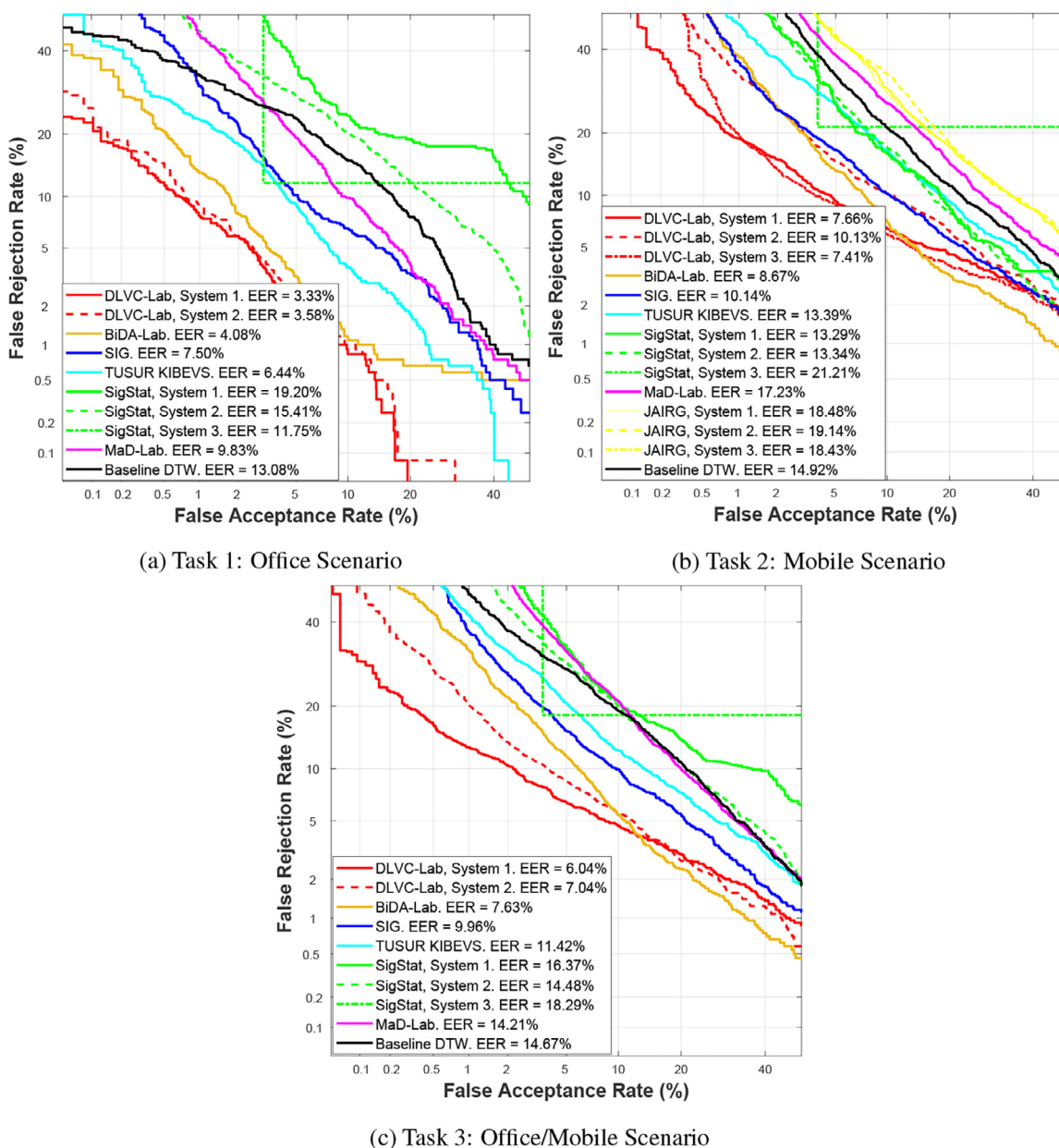


Fig. 6. Systems results in terms of DET curves over the final evaluation dataset of SVC-onGoing.

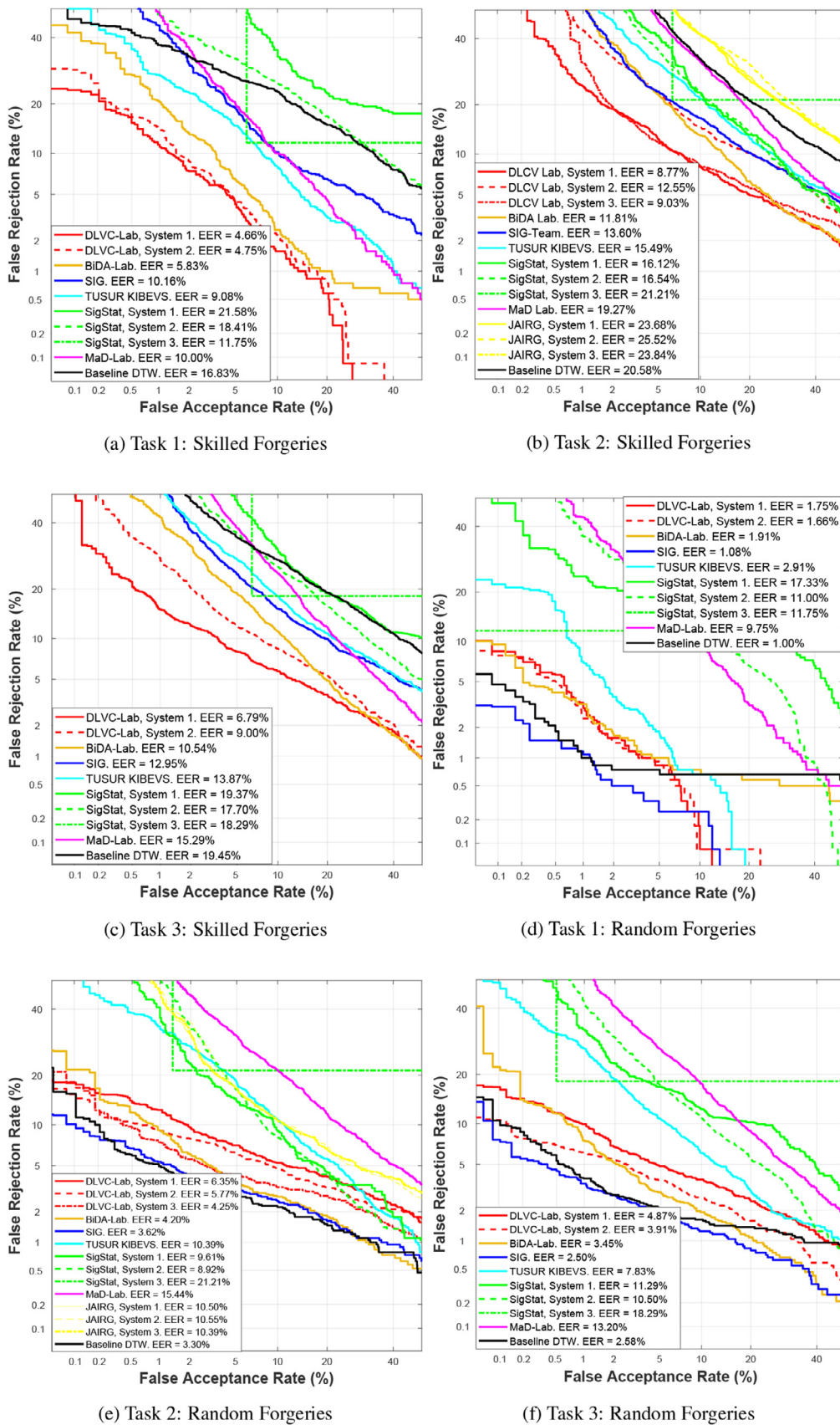


Fig. 7. Forgery Analysis: DET curves over the final evaluation dataset of the SVC-onGoing.

Finally, for completeness, we also analyse the False Acceptance Rate (FAR) and False Rejection Rate (FRR) results of all submitted systems. Fig. 6 shows the Detection Error Tradeoff (DET) curves of each task. As indicated in Section 4.3, it is important to highlight that participants/teams are not required to test their signature verification approaches on all three tasks. This is the reason why, for example, the System 3 of DLVC-Lab is tested only on Task 2 (mobile scenario) as it was specifically designed for that scenario, using only finger-written signatures (see Section 5.1).

In general, for low values of FAR (i.e., high security), the DLVC-Lab systems achieve the best results in all tasks. It is interesting to remark that depending on the specific task, the FRR values for low values of FAR are very different. For example, analysing the best system of DLVC-Lab team, for a FAR value of 0.1%, the FRR value is around 20% for Task 1. However, the FRR value increases over 40% for Task 2, showing the challenging conditions considered in real mobile scenarios using the finger as writing input. A similar trend is observed for low values of FRR (i.e., high convenience).

6.3. Forgery analysis

This section analyses the impact of the type of forgery in the proposed on-line signature verification systems [10]. In the evaluation of SVC-onGoing, both random and skilled forgeries are considered simultaneously in order to simulate real scenarios. Therefore, the winner of the competition is the system that achieves the highest robustness against both types of impostors at the same time. We now analyse the level of security of the submitted systems for each type of forgery, i.e., random and skilled. Fig. 7 shows the DET curves of each task and type of forgery, including also the EER results, over the final SVC2021_EvalDB dataset.

Analysing the skilled forgery scenario (Fig. 7a, 7b, and 7c), in all cases the deep learning systems proposed by the DLVC-Lab and BiDA-Lab teams achieve the best results in terms of EER. These results are much better compared to the non deep learning systems proposed by the SIG and TUSUR KIBEVS teams, especially for the Office scenario (Task 1). Similar results are obtained in all three tasks by SIG and TUSUR KIBEVS teams, outperforming the traditional Baseline DTW system.

Regarding the random forgery scenario, interesting results are observed in Fig. 7(d), 7(e), and 7(f). In general, the system proposed by the SIG team outperforms the deep learning systems in all three tasks, achieving better EER results. In fact, a simple approach like the Baseline DTW system (included for comparison) is able to outperform all submitted systems achieving EER results of 1.00%, 3.30%, and 2.58% for each of the corresponding tasks of the competition, proving the potential of DTW for the detection of random forgeries. A similar trend was already discovered in previous studies in the literature [15], highlighting also the difficulties of deep learning models to detect both skilled and random forgeries simultaneously. This aspect has been partly improved by the DLVC-Lab and BiDA-Lab through the incorporation of DTW to the deep learning architectures (DSDTW and TA-RNN, respectively).

Finally, seeing the results included in Fig. 7, we also want to highlight the very challenging conditions considered in SVC-onGoing compared with previous international competitions. This is produced mainly due to the real scenarios studied in the competition, e.g., several acquisition devices and types of impostors, large number of subjects, etc.

7. Conclusions

This article has described the experimental framework and results of SVC-onGoing¹⁴, an on-going competition for on-line sig-

nature verification where researchers can easily benchmark their systems against the state of the art in an open common platform using large-scale public databases such as DeepSignDB¹⁵ and SVC2021_EvalDB¹⁶, and standard experimental protocols. The goal of SVC-onGoing is to evaluate the limits of on-line signature verification systems on popular scenarios (office/mobile) and writing inputs (stylus/finger) through large-scale public databases. The following tasks are considered in the competition: *i*) Task 1, analysis of office scenarios using the stylus as input; *ii*) Task 2, analysis of mobile scenarios using the finger as input; and *iii*) Task 3, analysis of both office and mobile scenarios simultaneously. In addition, both random and skilled forgeries are simultaneously considered in each task in order to simulate realistic scenarios.

The results achieved in the final evaluation stage of SVC-onGoing have proved the high potential of deep learning methods compared to traditional approaches such as Dynamic Time Warping (DTW). In particular, the current winner of SVC-onGoing is the DLVC-Lab team that has proposed an end-to-end trainable deep soft-DTW (DSDTW). The results achieved in terms of Equal Error Rates (EER) are 3.33% (Task 1), 7.41% (Task 2), and 6.04% (Task 3). Similar results are also obtained by the Time-Aligned Recurrent Neural Network (TA-RNN) proposed by BiDA-Lab team. These results prove the challenging conditions considered in SVC-onGoing compared to previous international competitions, specially for the mobile scenario (Task 2).

A posterior analysis of the on-line signature verification systems over random and skilled forgeries independently have also produced interesting findings: *i*) deep learning methods seem crucial to improve the performance of the systems against skilled forgeries, and *ii*) traditional and simple approaches based on DTW [39] are able to outperform deep learning methods when considering random forgeries. This aspect highlights the difficulties of deep learning models to detect both skilled and random forgeries simultaneously. This has been partly improved by the DLVC-Lab and BiDA-Lab through the incorporation of DTW to the deep learning architectures (DSDTW and TA-RNN, respectively).

Finally, we would like to highlight that SVC-onGoing follows the line already pointed out in [1]: “Finally, as for many fields where successful commercial applications have been developed, these breakthroughs come from the development of robust algorithms tested and validated on huge representative databases, from which benchmarks can be designed and comparative analysis can be conducted.”

On this regard, it would be advisable that future studies proposing new systems and/or algorithmic contributions in the field of signature, are tested on SVC-onGoing and present results on this common benchmark, so that the evolution of the state of the art of this technology can be objectively assessed.

Future studies may be oriented to improve the mobile scenario (Task 2) as: *i*) this is a very important scenario for commercial applications nowadays, and *ii*) poor results are currently obtained, 8.77% and 3.30% EERs for skilled and random forgeries, respectively. Also, more efforts are needed for the proposal of novel deep learning models and loss functions more robust against both skilled and random forgeries simultaneously [40].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

¹⁵ <https://github.com/BiDALab/DeepSignDB>.

¹⁶ https://github.com/BiDALab/SVC2021_EvalDB.

¹⁴ <https://competitions.codalab.org/competitions/27295>.

Acknowledgments

This work has been supported by projects: PRIMA (H2020-MSCA-ITN-2019-860315), TRESPASS-ETN (H2020-MSCA-ITN-2019-860813), INTER-ACTION (PID2021-126521OB-I00 MICINN/FEDER), Orange Labs, and by UAM-Cecabank.

References

- [1] M. Diaz, M.A. Ferrer, D. Impedovo, M.I. Malik, G. Pirlo, R. Plamondon, A perspective analysis of handwritten signature technology, *ACM Comput Surv* 51 (2019) 1–39.
- [2] M. Faundez-Zanuy, J. Fierrez, M.A. Ferrer, M. Diaz, R. Tolosana, R. Plamondon, Handwriting biometrics: applications and future trends in e-Security and e-Health, *Cognit Comput* 12 (5) (2020) 940–953.
- [3] M. Okawa, Online signature verification using single-template matching with time-series averaging and gradient boosting, *Pattern Recognit* 102 (2020) 107227.
- [4] R. Tolosana, P. Delgado-Santos, A. Perez-Urbe, R. Vera-Rodriguez, J. Fierrez, A. Morales, DeepWriteSYN: On-line handwriting synthesis via deep short-term representations, in: *Proc. AAAI Conference on Artificial Intelligence*, 2021.
- [5] S. Lai, L. Jin, Y. Zhu, Z. Li, L. Lin, SynSig2Vec: forgery-free learning of dynamic signature representations by sigma lognormal-based synthesis, *IEEE Trans Pattern Anal Mach Intell* (2021).
- [6] N. Sae-Bae, N. Memon, Online signature verification on mobile devices, *IEEE Trans. Inf. Forensics Secur.* 9 (6) (2014) 933–947.
- [7] R. Tolosana, R. Vera-Rodriguez, J. Ortega-Garcia, J. Fierrez, Preprocessing and feature selection for improved sensor interoperability in online biometric signature verification, *IEEE Access* 3 (2015) 478–489.
- [8] J. Galbally, M. Martinez-Diaz, J. Fierrez, Aging in biometrics: an experimental analysis on on-line signature, *PLoS ONE* 8 (7) (2013) e69897.
- [9] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, Reducing the template ageing effect in on-line signature biometrics, *IET Biom.* 8 (6) (2019) 422–430.
- [10] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, Presentation attacks in signature biometrics: types and introduction to attack detection, S. Marcel, M.S. Nixon, J. Fierrez and N. Evans (Eds.), *Handbook of Biometric Anti-Spoofing* (2nd Edition), Springer (2019).
- [11] R. Tolosana, R. Vera-Rodriguez, R. Guest, J. Fierrez, J. Ortega-Garcia, Exploiting complexity in pen- and touch-based signature biometrics, *International Journal on Document Analysis and Recognition* 23 (2020) 129–141.
- [12] R. Vera-Rodriguez, R. Tolosana, M. Caruana, G. Manzano, C. Gonzalez-Garcia, J. Fierrez, J. Ortega-Garcia, DeepSignCX: signature complexity detection using recurrent neural networks, in: *Proc. International Conference on Document Analysis and Recognition*, 2019.
- [13] O. Delgado-Mohatar, J. Fierrez, R. Tolosana, R. Vera-Rodriguez, Biometric template storage with blockchain: a first look into cost and performance trade-offs, in: *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
- [14] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, J. Ortega-Garcia, Benchmarking desktop and mobile handwriting across COTS devices: the e-BioSign biometric database, *PLoS ONE* 12 (5) (2017) 1–17.
- [15] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, Exploring recurrent neural networks for on-line handwritten signature biometrics, *IEEE Access* 6 (2018) 5128–5138.
- [16] S. Lai, L. Jin, Recurrent adaptation networks for online signature verification, *IEEE Trans. Inf. Forensics Secur.* 14 (6) (2018) 1624–1637.
- [17] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, DeepSign: deep on-line signature verification, *IEEE Transactions on Biometrics, Behavior, and Identity Science* 3 (2) (2021) 229–239.
- [18] R. Tolosana, R. Vera-Rodriguez, C. Gonzalez-Garcia, J. Fierrez, S. Rengifo, A. Morales, J. Ortega-Garcia, J.C. Ruiz-Garcia, S. Romero-Tapiador, J. Jiang, S. Lai, L. Jin, Y. Zhu, J. Galbally, M. Diaz, M. Ferrer, M. Gomez-Barrero, I. Hodashinsky, K. Sarin, A. Slezkin, M. Bardamova, M. Svetlakov, M. Saleem, C. Szucs, B. Kovari, F. Pulsmeier, M. Wehbi, D. Zanca, S. Ahmad, S. Mishra, S. Jabin, ICDAR 2021 competition on on-line signature verification, in: *Proc. International Conference on Document Analysis and Recognition*, 2021.
- [19] ISO/IEC JTC1 SC37 Biometrics, ISO/IEC 30107-1. Information Technology - Biometric presentation attack detection, International Organization for Standardization, 2016.
- [20] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, BioTouchPass2: touchscreen password biometrics using time-Aligned recurrent neural networks, *IEEE Trans. Inf. Forensics Secur.* 15 (2020) 2616–2628.
- [21] D.-Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, G. Rigoll, SVC2004: first international signature verification competition, in: *Proc. International Conference on Biometric Authentication*, 2004.
- [22] N. Houmani, A. Mayoue, S. Garcia-Salicetti, B. Dorizzi, M. Khalil, M. Moustafa, H. Abbas, D. Muramatsu, B. Yanikoglu, A. Kholmatov, M. Martinez-Diaz, J. Fierrez, J. Ortega-Garcia, J.R. Alcob e, J. Fabregas, M. Faundez-Zanuy, J. Pascual-Gaspar, V. Cardesoso-Payo, C. Vivaracho-Pascual, Biosecure signature evaluation campaign (BSEC'2009): evaluating on-Line signature algorithms depending on the quality of signatures, *Pattern Recognit* 45 (3) (2012) 993–1003.
- [23] V.L. Blankers, C.E. van den Heuvel, K. Franke, L. Vuurpijl, ICDAR 2009 signature verification competition, in: *Proc. International Conference on Document Analysis and Recognition*, 2009.
- [24] M. Liwicki, M.I. Malik, C.E. Van Den Heuvel, X. Chen, C. Berger, R. Stoel, M. Blumenstein, B. Found, Signature verification competition for online and offline skilled forgeries (SigComp2011), in: *Proc. International Conference on Document Analysis and Recognition*, 2011.
- [25] A. Das, H. Suwanwivat, U. Pal, M. Blumenstein, ICFHR 2020 competition on short answer assessment and Thai student signature and name components recognition and verification (SASIGCOM 2020), in: *Proc. International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2020.
- [26] J. Ortega-Garcia, J. Fierrez-Aguilar, et al., MCYT baseline corpus: A bimodal biometric database, *Proc. IEEE Vision, Image and Signal Processing* 150 (6) (2003) 395–401.
- [27] J. Fierrez, J. Galbally, J. Ortega-Garcia, M.R. Freire, F. Alonso-Fernandez, D. Ramos, D.T. Toledano, J. Gonzalez-Rodriguez, J.A. Siguenza, J. Garrido-Salas, et al., BiosecurID: a multimodal biometric database, *Pattern Analysis and Applications* 13 (2) (2010) 235–246.
- [28] M. Cuturi, M. Blondel, Soft-DTW: a differentiable loss function for time-series, in: *Proc. International Conference on Machine Learning*, 2017.
- [29] J. Galbally, M. Diaz-Cabrera, M.A. Ferrer, M. Gomez-Barrero, A. Morales, J. Fierrez, On-line signature recognition through the combination of real dynamic data and synthetically generated static data, *Pattern Recognit* 48 (9) (2015) 2921–2934.
- [30] M. Martinez-Diaz, J. Fierrez, R.P. Krish, J. Galbally, Mobile signature verification: feature robustness and performance comparison, *IET Biom.* 3 (2014) 267–277.
- [31] M. Diaz, M.A. Ferrer, D. Impedovo, G. Pirlo, G. Vessio, Dynamically enhanced static handwriting representation for parkinsons disease detection, *Pattern Recognit Lett* 128 (2019) 204–210.
- [32] W. Bouamra, C. Djeddi, B. Nini, M. Diaz, I. Siddiqi, Towards the design of an offline signature verifier based on a small number of genuine samples for training, *Expert Syst Appl* 107 (2018) 182–195.
- [33] M.A. Ferrer, J.B. Alonso, C.M. Travieso, Offline geometric parameters for automatic signature verification using fixed-Point arithmetic, *IEEE Trans Pattern Anal Mach Intell* 27 (6) (2005) 993–997.
- [34] Y. Serdouk, H. Nemmour, Y. Chibani, Handwritten signature verification using the quad-tree histogram of templates and a support vector-based artificial immune classification, *Image Vis Comput* 66 (2017) 26–35.
- [35] E. Hancer, I. Hodashinsky, K. Sarin, A. Slezkin, A wrapper metaheuristic framework for handwritten signature verification, *Soft comput* 25 (13) (2021) 86658681.
- [36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016.
- [37] I. Chevrete, A. Kormilitzin, A primer on the signature method in machine learning, *arXiv preprint arXiv:1603.03788* (2016).
- [38] J. Galbally, J. Fierrez, M.R. Freire, J. Ortega-Garcia, Feature selection based on genetic algorithms for on-line signature verification, in: *Proc. IEEE Workshop on Automatic Identification Advanced Technologies*, 2007.
- [39] M. Okawa, Time-series averaging and local stability-weighted dynamic time warping for online signature verification, *Pattern Recognit* 112 (2021) 107699.
- [40] A. Morales, J. Fierrez, A. Acien, R. Tolosana, I. Serna, SetMargin loss applied to deep keystroke biometrics with circle packing interpretation, *Pattern Recognit* 122 (2022) 108283.

Ruben Tolosana received the M.Sc. and Ph.D. degrees in 2014 and 2019. Since 2021, he is Assistant Professor at Universidad Autonoma de Madrid. His research interests are mainly focused on pattern recognition and machine learning, particularly in the areas of handwritten signature, human-computer interaction, DeepFakes, biometrics, and privacy.

Ruben Vera-Rodriguez received the Ph.D. degree in electrical and electronic engineering from Swansea University, U.K., in 2010. Currently he is an Associate Professor at Universidad Autonoma de Madrid, Spain. His research interests include signal and image processing, machine learning, and biometrics. He has authorized more than 130 scientific articles.

Carlos Gonzalez-Garcia received his B.Sc. degree in Computer Science Engineering and his M.Sc. degree from the Universidad Autonoma de Madrid, where he currently is a Ph.D. student at the BiDA Lab research group. His work is focused on machine learning, handwritten biometrics and Human-Computer Interaction.

Julian Fierrez received the M.Sc. and Ph.D. degrees in 2001 and 2006, respectively. He is now Associate Professor at UAM. His research interests include signal and image processing, pattern recognition, security, and biometrics. He received the IAPR Young Biometrics Investigator Award 2017 and is Associate Editor of Elseviers INFORMATION FUSION.

Aythami Morales Moreno M.Sc. in Electrical Engineering, and Ph.D from the Universidad de LPGC in 2006 and 2011. Since 2017, he is Associate Professor with the Universidad Autonoma de Madrid. In his work, he combines his interests in machine learning, biometric processing, security, and privacy.

Javier Ortega-Garcia received the M.Sc. and Ph.D. degrees in electrical engineering in 1989 and 1996, respectively. He is founder and co-director of the Biomet-

ric Recognition Group ATVS. His research interests are focused on biometrics signal processing: on-line signature verification, speaker recognition, web-based biometrics, data fusion and multibiometrics.

Juan Carlos Ruiz-Garcia received his B.Sc. degree in Computer Science Engineering from the Universidad de Granada and got the M.Sc. degree from the Universidad Autonoma de Madrid, where he currently a PhD student at the research group BiDA Lab focusing his research on e-Learning, e-Health and Computer-Human Interaction.

Sergio Romero-Tapiador received the BSc in Computer Science and Engineering in 2020 from Universidad Autonoma de Madrid. Since September 2019, he is a member of the Biometrics and Data Pattern Analytics - BiDA Lab. Among the research activities, he is mainly working on Pattern Recognition and Machine Learning.

Santiago Rengifo received a 2-year degree in Web and Application development in 2020. He is currently working at Biometrics and Data Pattern Analytics (BiDA-Lab) at Universidad Autonoma de Madrid as a laboratory assistant.

Miguel Caruana received his M.Sc. degree in Telecommunication Engineering in 2021 from Universidad Autónoma de Madrid. He has been collaborating with the Biometrics and Data Pattern Analytics (BiDA-Lab) at the Universidad Autonoma de Madrid through a research grant during his studies.

Jijia Jiang received the B.S. degree in South China University of Technology in 2020, and she is currently pursuing the master degree in signal and information processing. Her research interests include machine learning and handwritten biometric recognition.

Songxuan Lai received the Ph.D degree in electronics and information engineering from South China University of Technology in 2021. He is currently working as an artificial intelligence engineer at Huawei Cloud. His research interests include handwriting analysis and recognition, OCR systems, and machine learning.

Lianwen Jin received the Ph.D. degree from South China University of Technology in 1996. He is professor in the College of Electronic and Information Engineering at the South China University of Technology. He is the author of more than 200 scientific papers and serves as AC/SPC/PC in many international conferences.

Yecheng Zhu received the master degree in South China University of Technology in 2021. His research interests include machine learning, handwriting analysis and computer vision.

Javier Galbally received the Ph.D. degree in Electrical Engineering in 2009. In 2013, he joined the European Commission in DG JRC, where he is currently a Scientific/Technical Officer. He is the Chair of the EAB Research Projects Conference and has authored over 100 publications mainly focused on biometrics.

Moises Diaz received the M.Tech., M.Sc., and Ph.D. degrees in engineering from Universidad de Las Palmas de Gran Canaria, Spain, in 2010, 2011, and 2016, respectively. He joined that University as an Associate Professor in 2021. His current research interests include pattern recognition, document analysis, handwriting recognition, and biometrics.

Miguel Angel Ferrer received an M.Sc. and a Ph.D. from the Universidad Politcnica de Madrid, Spain, in 1988 and 1994. He joined the University of Las Palmas de Gran Canaria, Spain, in 1989, where he is currently a Full Professor. His research interests include pattern recognition, biometrics, and computer vision.

Marta Gomez-Barrero received the Ph.D. degree in Electrical Engineering in 2016. Since 2020, she is a Professor for IT-Security and technical data privacy at the Hochschule Ansbach, in Germany. She has received several distinctions, including: EAB European Biometric Industry Award 2015 and Best Paper Award at ICB 2015.

Ilya Hodashinsky received the Ph.D. degree in 1984, the Dr. Sc. degree in 2004 from the Tomsk State University of Control Systems and Radioelectronics (TUSUR), Russia, and the Professor title at the 2011. His research interests include computational intelligence, fuzzy modeling, and pattern recognition.

Konstantin Sarin graduated from the Faculty of Control Systems, TUSUR. Received the Ph.D. degree in 2016. He is assistant professor of the Department of Complex

Information Security of Computer Systems in TUSUR. His research interests include computational intelligence, fuzzy modeling, and machine learning.

Artem Slezkin received the engineer's degree at specialty Information Security of Automated Banking TUSUR, in 2019. He is currently pursuing the PhD degree with the Laboratory of Biological Signal Extraction, Analysis and Management at TUSUR. His research interests include pattern recognition, machine learning, and fuzzy systems.

Marina Bardamova graduated from the Faculty of Security, TUSUR in 2017. She is a lecturer of the Department of Complex Information Security of Computer Systems in TUSUR and a junior researcher of the Laboratory. Her main research interests include computational intelligence, data mining, fuzzy modeling, and machine learning.

Mikhail Svetlakov received the Specialist of Information Security degree from TUSUR. He is currently a Junior Research Assistant at the Laboratory for collection, analysis, and control of biological signals, TUSUR. Research interests - fuzzy systems, computer vision.

Mohammad Saleem is a Ph.D. candidate in computer engineering at Budapest University of Technology and Economics, Hungary. His research interests include online signature verification. Previously, he worked as a researcher and teacher assistant at Yarmouk University, Irbid, Jordan. He is a member of the Jordanian engineers' association.

Cintia Lia Szucs is a Ph.D. candidate in software engineering at Budapest University of Technology and Economics, Hungary. Her research interests include online signature verification. She is a member of the Hungarian Association for Image Processing and Pattern Recognition, a member of the John von Neumann Computer Society.

Bence Kovari received his Ph.D. degree in software engineering from Budapest University of Technology and Economics, Hungary, in 2013, studying the automated verification of handwritten signatures. He is a member of the Hungarian Association for Image Processing and Pattern Recognition and a member of the John von Neumann Computer Society.

Falk Pulsmeier received his B.Sc. and M.Sc. Degrees in Mathematics from the RWTH Aachen University. He recently started at the MaD Lab, FAU Erlangen-Nrnberg, as PhD student with a focus on Machine Learning and Medical Engineering.

Mohamad Wehbi received B.Eng. in Computer Engineering and Informatics from the Beirut Arab University, Lebanon. He received his M.Sc. Degree in Robotics and Automation Engineering from the University of Siena, Italy. He is currently a researcher pursuing his PhD at the AIBE Department, FAU Erlangen-Nrnberg, Erlangen, Germany.

Dario Zanca received B.Sc. and M.Sc. Degrees in Mathematics from the University of Palermo, Italy, and Ph.D. in Smart Computing from the University of Florence, Italy. He worked as postdoc researcher at the DSMCN, University of Siena, Italy. He is currently postdoc researcher at AIBE Department, FAU Erlangen-Nrnberg, Erlangen, Germany.

Sumaiya Ahmad received B.Sc. in Computer application from Aligarh Muslim University, India in 2015. She received master in computer application (MCA) from Department of Computer Science, JMI, Delhi, India in 2018 and currently pursuing Ph.D. in computer science from here. Her research interests include AI, biometrics, and Information Security.

Sarthak Mishra received BCA from Indraprastha University, New Delhi, India in 2015. He received MCA from Department of Computer Science, JMI, New Delhi, India in 2018 and currently pursuing Ph.D. in computer science from here. His research interests include AI, biometrics, video surveillance, and crowd monitoring.

Suraiya Jabin received MCA in 2002 and Ph.D. in 2009 from Jamia Hamdard university, Delhi, India. In Dec 2006, she joined as Assistant Professor in JMI, India where she continues to work, as a Professor since Nov 2019. Her research interests include AI, behavioral biometrics, and computational biology.