DeepSignCX: Signature Complexity Detection using Recurrent Neural Networks

Ruben Vera-Rodriguez, Ruben Tolosana, Miguel Caruana, Gustavo Manzano,
Carlos Gonzalez-Garcia, Julian Fierrez, and Javier Ortega-Garcia
Biometrics and Data Pattern Analytics - BiDA Lab
Universidad Autonoma de Madrid, Madrid, Spain
{ruben.vera, ruben.tolosana, julian.fierrez, javier.ortega}@uam.es

***Abstract*—This paper proposes a novel approach for on-line signature complexity detection based on Recurrent Neural Networks (RNNs). Complexity of handwritten signatures can vary from very simple ones (just a simple flourish) to very complex signatures (including the handwritten full name and complex flourish). Three different complexity levels are proposed: low, medium, and high. Time functions are extracted from the on-line signatures and a system based on RNNs (BLSTM in particular) is trained to classify the three levels of complexity over a ground truth manually labelled database (BiosecurID with 400 subjects). This initial model is used to automatically label a very large database (DeepSignDB[1]) containing over 1500 subjects, which is then used to train the proposed RNN for signature complexity detection. Promising results ca. 85% of accuracy are achieved. This complexity detector could be used as a first stage in a signature verification system in order to train a specific biometric system per signature complexity level and improve the overall system performance.**

***Keywords*-biometrics; on-line signature; signature complexity; recurrent neural networks; deep learning**

I. Introduction

Handwritten signature verification systems have been shown to be highly sensitive to signature complexity [1]–[3]. In [4], Alonso-Fernandez *et al.* evaluated the effect of the complexity and legibility of the signatures for off-line signature verification (i.e., signatures with no available dynamic information) pointing out the differences in performance for several matchers. Signature complexity has also been associated to the concept of entropy, defining entropy as the inherent information content of biometric samples [5]. In [6] a "personal entropy" measure based on Hidden Markov Models (HMM) was proposed in order to analyse the complexity and variability of on-line signatures regarding three different levels of entropy. In addition, the same authors proposed in [7] a new metric known as "relative entropy" for classifying users into animal groups (see the biometric menagerie [8]) where skilled forgeries were also considered. Recently, we proposed in [2], [3] a novel system for on-line signature complexity detection based on computing the number of lognormals obtained from the Sigma LogNormal writing generation model [9]. Then, after the complexity of the signature was detected, we trained specific signature verification systems adapted to each complexity level group obtaining significant improvements of the system performance. Although, the good results achieved confirmed the success of our approach, the proposed signature complexity detector may not be very robust in some cases as it is only based on a single measure (i.e., the number of lognormals).

Deep learning isapproaches are the state-of-the-art technology used in other biometric recognition traits such as the face [10] or the voice [11]. However, this technology has not been widely used in applications such as on-line signature verification due mainly to the lack of large amounts of data that can improve the performance over traditional approaches. In [12] a system based on Recurrent Neural Networks (RNNs) with a Siamese architecture was proposed for on-line signature verification with very promising results.

In this work, we propose the usage of RNNs in order to develop an on-line signature complexity detection system through a semi-supervised process. First, an initial model is trained over the BiosecurID database [13], for which the manual labels of the signatures' complexity are available as low, medium and high complexity. Then, this model is used to automatically obtain the complexity labels of a much larger database with more than 1500 subjects (DeepSignDB [14]). Finally, based on these automatic labels, a new RNN signature complexity detection system is developed.

The remainder of the paper is organized as follows. Sec. II describes the proposed on-line signature complexity detector. Sec. III describes the on-line signature databases considered in the experimental work. Sec. IV describes the experimental protocol and the results achieved. Finally, Sec. V draws the final conclusions and points out some lines for future work.

II. On-Line Signature Complexity Detector

The on-line signature complexity detector proposed in this work is based on time functions (a.k.a. local system) [15], [16]. For each signature acquired, signals related to X and Y pen coordinates and pressure are used to extract a set of 23 time functions as in [12].

For the classification of each signature into a specific complexity level, our proposed detector is based on a RNN deep learning technology. Two different approaches are considered:

[1]https://github.com/BiDAlab/DeepSignDB

2379-2140/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDAR.2019.00214

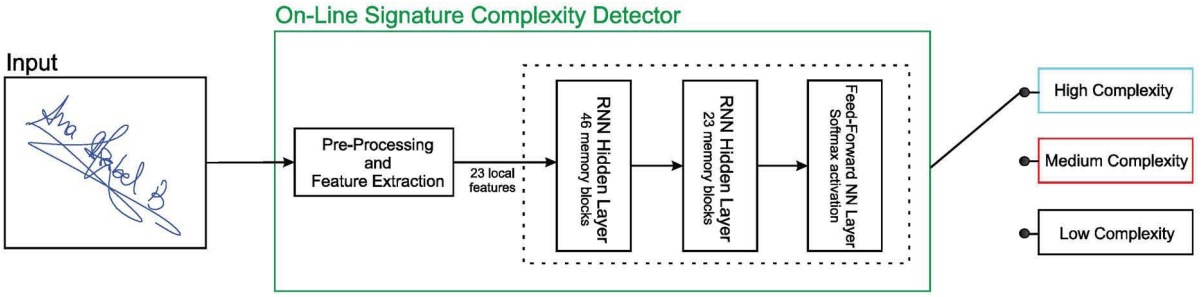


Fig. 1: Architecture of our proposed RNN on-line signature complexity detector.

i) Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU). In particular, bidirectional schemes (i.e., BLSTM and BGRU), which also allow access to future context, are considered in this study as they have recently achieved much better results compared to the original schemes for the task of on-line signature verification [12]. Fig. 1 graphically summarizes the architecture of our proposed RNN on-line signature complexity detector. For the input of the detector, we feed the network with the 23 time functions extracted from the signature. These time functions are preprocessed following a zero mean and unit standard deviation normalization. For the RNN different configurations were tested with different number of hidden layers, neurons per layer, learning rate, etc. These details are described in Sect. IV. Finally, a feed-forward neural network layer with a softmax activation is considered, providing an output score for each of the three complexity levels considered.

III. On-line Signature Databases

The following two databases are considered in the experimental work:

A. BiosecurID Database

This database is comprised of 16 original signatures and 12 skilled forgeries per user, captured in 4 separate acquisition sessions [13]. Each session was captured leaving a two month interval between them. There are a total of 400 subjects and signatures were acquired considering a controlled and supervised office-like scenario. Users were asked to sign on a piece of paper, inside a grid that marked the valid signing space, using an inking pen. The paper was placed on a Wacom Intuos 3 pen tablet that captured the following time signals of each signature: X and Y pen coordinates (5080 dpi resolution), pressure (1024 levels) and timestamp (100 Hz). In addition, pen-ups trajectories are available.

For this database, signature complexity labels generated manually were taken from [2]. Three different complexity levels (high, medium and low) were considered based on previous works [7]. Users with signatures longer in writing time and with an appearance more similar to handwriting were labelled as high-complexity users whereas those users with signatures shorter in time and with generally simple flourish with no legible information were labelled as low-complexity users. This served as a ground truth to both train specific models per complexity level and evaluate the signature complexity detector.

B. DeepSignDB Database

The DeepSignDB database[2] [14] comprises data from a total of 1526 subjects from four different well-known on-line signature databases: MCYT (330 subjects) [17], BiosecurID (400 subjects) [13], Biosecure DS2 (650 subjects) [18], e-BioSign (65 subjects) [19] and a novel signature database not presented yet (with 81 subjects). This database comprises more than 70K signatures acquired using both stylus and finger inputs. Two acquisition scenarios are considered, office and mobile, with a total of 8 different devices. Additionally, different types of impostors and number of acquisition sessions are considered along the database. For the results presented in this work, only signatures performed with pen stylus are considered. For this database there are no complexity labels available (except for BiosecurID), so this database is used to train the complexity detector in a semi-supervised mode.

IV. Experimental Work

A. Experimental Protocol

The experimental protocol has been designed to evaluate the signature complexity detector proposed in this work in order to classify users into three different complexity levels. In this sense, only genuine signatures have been considered, as done in previous related works [2], [6], [7].

Two experiments have been carried out: *i*) we first train the RNN complexity detector using the manual labels from BiosecurID database; and *ii*) this initial model has been used to classify the complexity of the signatures from DeepSignDB. These new labels together with the manual ones are finally used to train a more robust complexity detector.

In the first experiment (Section IV-B1), BiosecurID database is divided into development and evaluation datasets, considering all available genuine signatures from 50 and 13 subjects per complexity level respectively. This makes a total of 150 subjects for development and 39 users for evaluation. The remaining subjects were discarded as for training RNN models

[2]https://github.com/BiDAlab/DeepSignDB

it is important to consider balanced classes. Considering all signatures from the development dataset, a RNN model was trained to classify an input signature into one of the three possible complexity levels. The evaluation results are given as the accuracy of the automatic labels against the manual ground truth over the evaluation dataset.

In the second experiment (Section IV-B2), the previous RNN model is used to classify the complexity level of all signatures of DeepSignDB (except for BiosecurID database whose original manual labels are kept). In this case signatures from a total of 231 new subjects compared to the previous experiment for each complexity class were available. The database was divided into development (with data from 256 subjects per complexity class, i.e., 231 new subjects from DeepSignDB plus 25 subjects from the development set of BiosecurID), validation (the remaining 25 subjects from the development set of BiosecurID per complexity class) and evaluation (the 13 subjects from the evaluation set of BiosecurID per complexity class). The evaluation results are obtained over the same evaluation dataset of BiosecurID in order to carry out a comparative analysis of the results achieved using this semi-supervised approach and the previous fully supervised one.

B. Experimental Results

*1) **RNN Model over BiosecurID Database**:* For the first experiment different RNN architectures were tested in order to achieve optimal complexity detection results over BiosecurID database. Two types of RNN layers were used: BLSTMs and BGRUs. Different experiments were performed varying the number of layers (one to three BLSTM and BGRU layers), the number of neurons per layer, learning rate (0.001, 0.0001 and 0.00001), and also varying the recurrent dropout rate. Finally, the best results were achieved for the case of having two RNN hidden layers (for both BLSTM and BGRU approaches). The first layer is composed of 46 memory blocks. The output of the first hidden layer serves as input to the second RNN hidden layer, which is composed of 23 memory blocks. Finally, a feed-forward neural network layer with a softmax activation is set to provide an output score for each of the three complexity levels considered. Also, best results were achieved when having a learning rate of 0.00001 without recurrent dropout.

Table I shows the accuracy (in %) of the complexity classification over the evaluation set of BiosecurID. Classification accuracies of 84.3% and 85.4% are obtained for both BGRU and BLSTM RNN models respectively. As the BLSTM model achieves better results, this model architecture is the one used in the next experiments. Figure 2 shows the training process with accuracy over the several epochs. The training process was carried out for 80 epochs, but only the results for the first 40 epochs are shown in Figure 2 for better visualization. As shown in the figure, results over the development dataset are able to achieve 100% of accuracy. However, the final BLSTM model chosen is the one obtained at epoch 26 as it achieves the best accuracy over the evaluation dataset (85.4%).

An analysis of the learnt features of the optimal BLSTM model was also undertaken to improve our understanding of

TABLE I: Accuracy results of two RNN architectures, BGRU and BLSTM over the evaluation set of BiosecurID

	Accuracy (in %)
BGRU	84.3
BLSTM	85.4

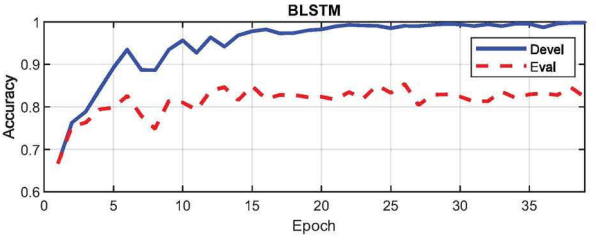


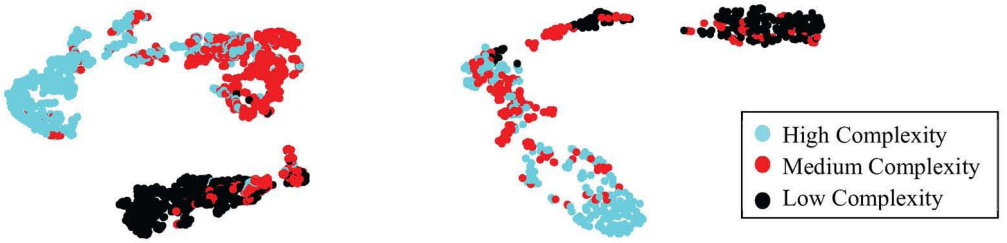
Fig. 2: Training process of the best BLSTM RNN model of Sect. IV-B1 for signature complexity detection with accuracy for each training epoch for both Development and Evaluation datasets of BiosecurID.

the feature learning models. Using the t-SNE technique for feature clustering visualisation, we can visually evaluate in Figure 3 the spatial representation of the signatures regarding the ground truth complexity labels for both the development and evaluation datasets. Light blue dots represent signatures with high complexity manual label, red dots represent signatures with medium complexity manual label, and finally black dots represent signatures with low complexity manual label. This representation is carried out using the manual labels, as the representation using the automatic labels gives perfect clustering results. As shown in the figure, the majority of the signatures are well clustered into the three complexity groups. However, there are some cases of signatures placed in a wrong cluster regarding their manual complexity label. Later on, we perform a more thorough analysis of the data and give some possible reasons for this.

*2) **RNN Model over DeepSignDB Database**:* The second experiment was designed to check whether the use of a large database (DeepSignDB), but without having any ground truth regarding the level of complexity of the signatures, could improve the complexity RNN model following a semi-supervised setting.

In this case, the best complexity RNN model (BLSTM) trained in the previous experiment (Sect. IV-B1) was used to classify the complexity level of all DeepSignDB signatures (except for BiosecurID database whose original manual labels are kept). Then, these new complexity labels together with the manual labels from BiosecurID were used to train a new BLSTM model using a much larger amount of signatures (see Section. IV-A for the details on the division of the database).

It is worth noting that the majority of the signatures of DeepSignDB were classified as medium complexity, which is probably the most common class in real life. Therefore, many signatures of the medium complexity group were not considered further in order to have balanced classes. Also, as the



(a) Development Set (b) Evaluation Set

Fig. 3: t-SNE representation of features learned for the BLSTM RNN model from Sect. IV-B1. The color of the dots indicate the original manual label: light blue dots represent signatures with high complexity manual label, red dots represent medium complexity, and black dots represent low complexity.

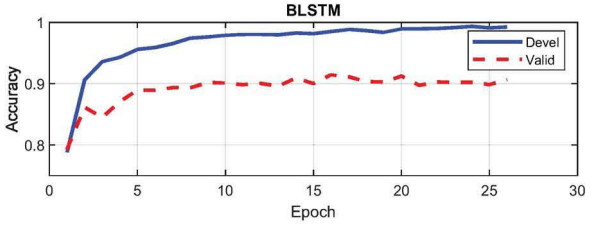


Fig. 4: Training process of the best BLSTM RNN model of Sect. IV-B2 for signature complexity detection with accuracy for each training epoch for both Development and Validation datasets of DeepSignDB.

complexity detector is applied to each signature individually, in order to form balanced sets for training this new BLSTM model, a final decision of the complexity level of the signatures of a given subject was taken by assigning a specific label if more than 2/3 of all signatures had the same complexity label.

With this new semi-supervised dataset of signatures from 256 subjects for each complexity level, a new BLSTM RNN model was trained. In this case, the same configuration of the RNN model from the previous section was followed. Figure 4 shows the training process of the BLSTM model with accuracy over the several epochs for both the development set and the validation set. The model obtained for epoch 16 with best accuracy of 91.42% over the validation set was chosen. Finally applying this model to the evaluation set, an accuracy of 85.7% was achieved, which is a bit higher than the accuracy of 85.4% obtained in Sect. IV-B1 only using signatures from the development set of BiosecurID. It is important to note that this new model has been trained with signatures from 5 different databases and seven different acquisition devices. Therefore, it is likely that this new model is able to generalize much better over new signatures compared to the previous one, just trained with data from BiosecurID and having only one acquisition device.

Finally, Fig. 5, 6 and 7 show examples of the complexity classification of some signatures from the evaluation set of the BiosecurID database produced by this last BLSTM RNN model trained over the development set of DeepSignDB. Figure 5 shows on top three signatures correctly classified as high complexity signatures. Also, on the bottom part of the figure it is possible to see examples of signatures wrongly classified as having a high complexity, because the manual label in the three examples was of medium complexity. As can be seen, visually these signatures look very similar compared to the correctly classified ones. So even if the manual label is medium complexity, it is clear they could also have been manually labelled as of high complexity, showing the ambiguity of the manual labelling for some signatures. Figures 6 and 7 show similar examples for the other two cases of complexity, also providing similar conclusions. It is possible to visually check that the BLSTM complexity classifier is doing a good job detecting some cases of wrong manual complexity assignments.

This approach could be compared to the complexity detection approach based on applying the Sigma Lognormal writing generation model proposed in [2], which also considered the BiosecurID database. In that case the accuracy obtained was of 64% for complexity detection. Here the accuracy achieved is over 85%, making a very significant improvement of performance.

V. Conclusions and Future Work

This paper has proposed a novel approach for on-line signature complexity detection based on RNNs. The architecture providing the best performance has been based on two BLSTM hidden layers with a final feed-forward neural network layer with a softmax activation, providing an output score for each of the three complexity levels considered. First, an initial model is trained over BiosecurID database, for which we have the manual labels for the signatures as low, medium and high complexity. Then, this model is used to automatically obtain the complexity labels of a much larger database with more than 1500 subjects (DeepSignDB). Based on these labels a new signature complexity detection system if finally developed. Best results of ca. 85% of accuracy are achieved. A visual analysis of the model prediction shows very good results. It

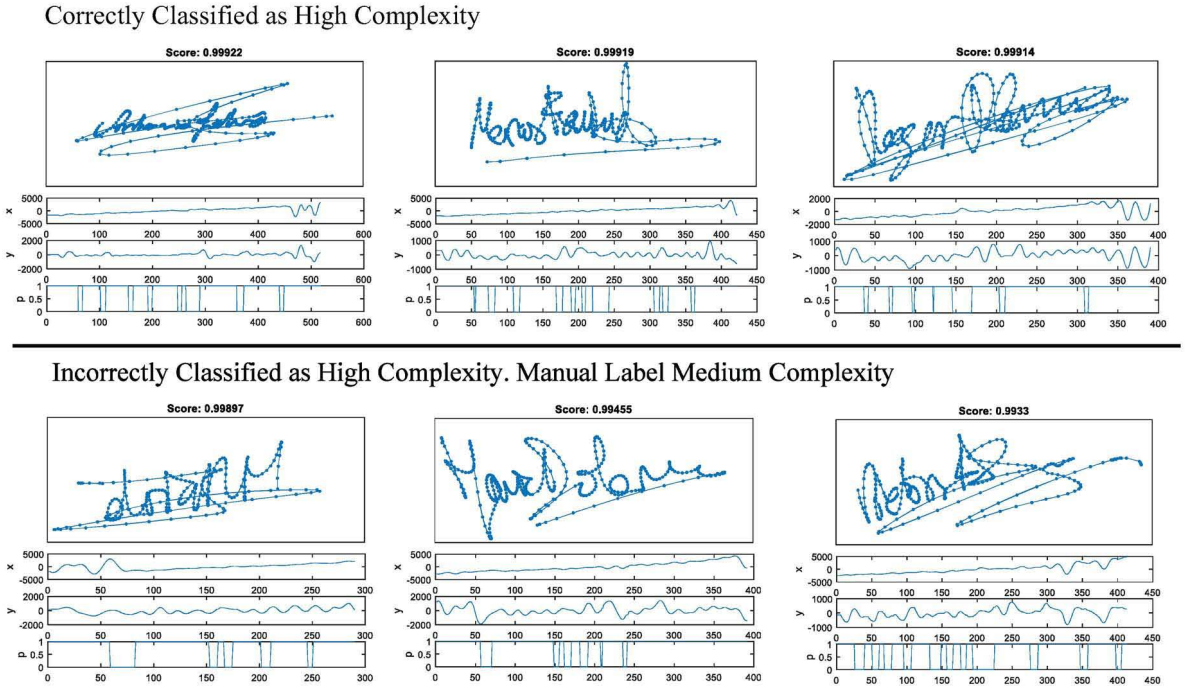


Fig. 5: Examples of signatures classified as high complexity correctly (top), and incorrectly (bottom).

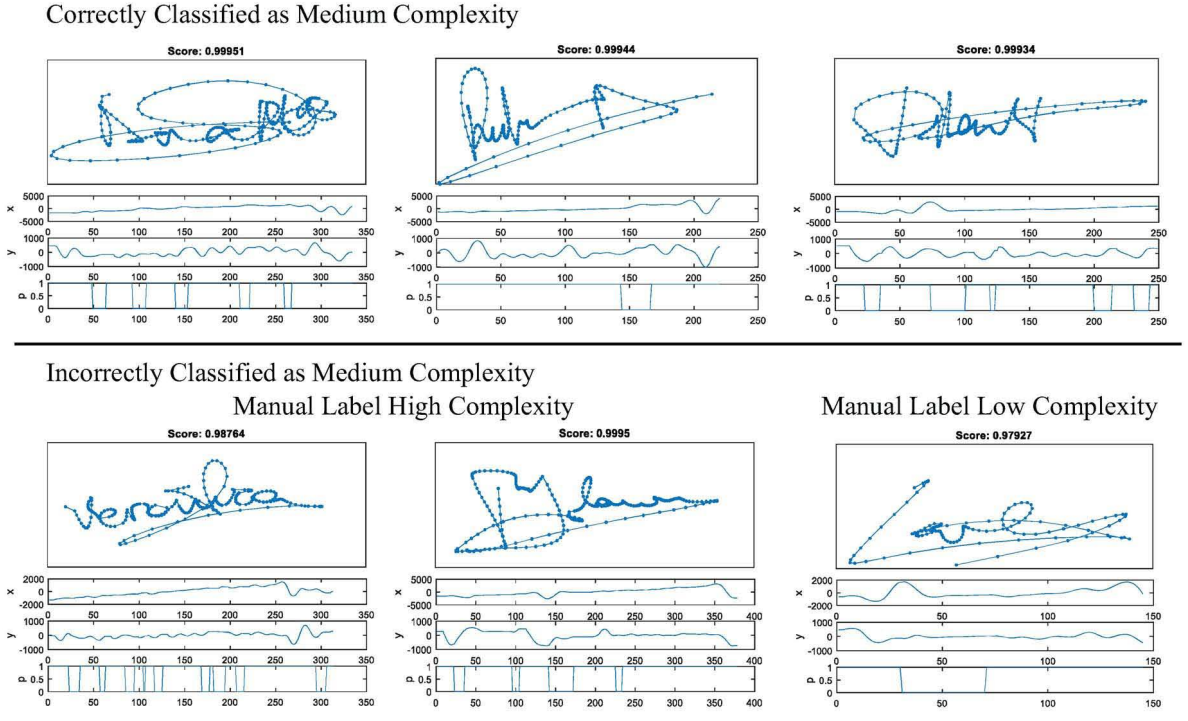


Fig. 6: Examples of signatures classified as medium complexity correctly (top), and incorrectly (bottom).

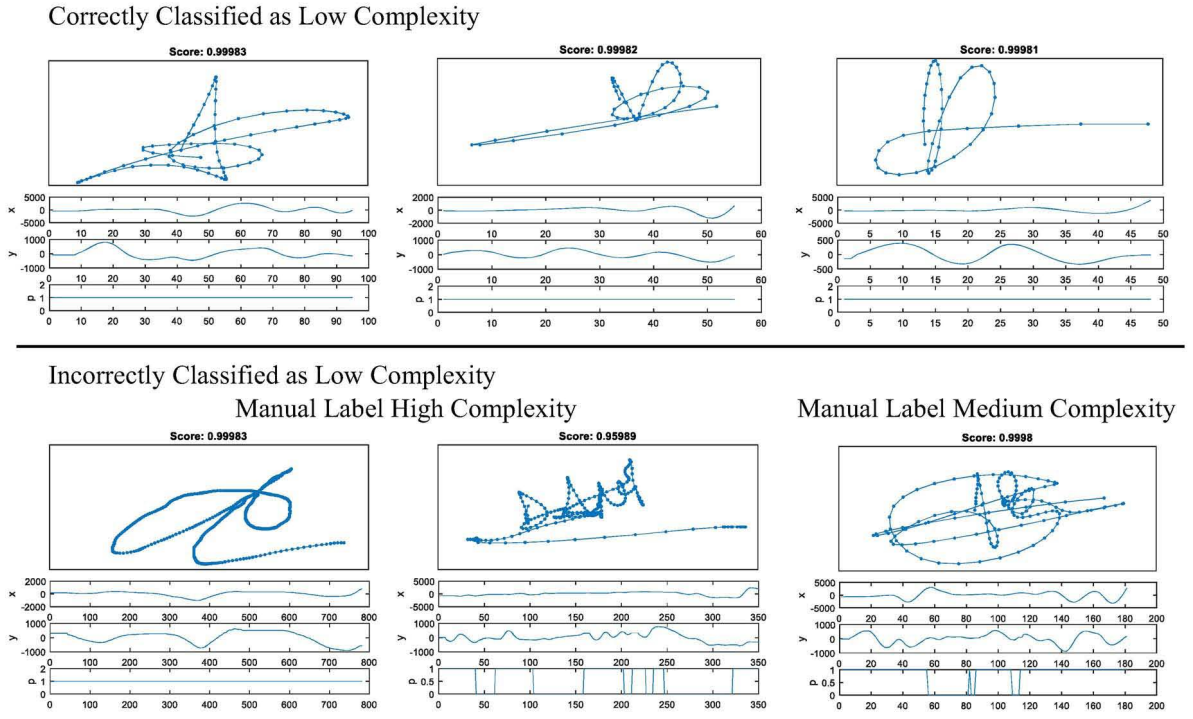


Fig. 7: Examples of signatures classified as low complexity correctly (top), and incorrectly (bottom).

is worth noting that some manual labels could be wrong, as there are cases of signatures that even for humans are not easy to classify to a specific complexity level.

For future work, this complexity detection system will be used to train a specific on-line signature verification system for each complexity level in order to improve the overall signature verification system performance.

ACKNOWLEDGMENTS

Work supported by projects: BIBECA (RTI2018-101248-B-I00), Bio-Guard (Ayudas Fundación BBVA a Equipos de Investigación Científica 2017) and by CecaBank.

REFERENCES

[1] J. Fierrez, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Target Dependent Score Normalization Techniques and Their Application to Signature Verification," *IEEE Transactions on Systems, Man, and Cybernetics. Part C*, vol. 35, no. 3, pp. 418–425, 2005.

[2] R. Tolosana, R. Vera-Rodriguez, R. Guest, J. Fierrez, and J. Ortega-Garcia, "Complexity-based biometric signature verification," in *Proc. IAPR ICDAR*, 2017.

[3] R. Tolosana, R. Vera-Rodriguez, R. Guest, J. Fierrez, and J. Ortega-Garcia, "Exploiting Complexity in Pen- and Touch-based Signature Biometrics," *arXiv:1905.03676*, May 2019.

[4] F. Alonso-Fernandez, M. Fairhurst, J. Fierrez, and J. Ortega-Garcia, "Impact of Signature Legibility and Signature Type in Off-Line Signature Verification," *In Proc. IEEE Biometrics Symposium*, 2007.

[5] Z. Zhou, *"Biometric Entropy"*. S.Z. Li and A. Jain (Eds.), *Encyclopedia of Biometrics*, Springer, 2009.

[6] N. Houmani, S. Garcia-Salicetti, and B. Dorizzi, "A Novel Personal Entropy Measure Confronted to Online Signature Verification Systems' Performance," *in Proc. BTAS*, 2008.

[7] N. Houmani and S. Garcia-Salicetti, "On Hunting Animals of the Biometric Menagerie for Online Signature," *PLOS ONE*, 2016.

[8] N. Yager and T. Dunstone, "The Biometric Menagerie," *IEEE Transactions on PAMI*, vol. 32, no. 2, pp. 220–230, 2010.

[9] C. O'Reilly and R. Plamondon, "Development of a Sigma-Lognormal Representation for On-Line Signatures," *Pattern Recognition*, vol. 42, no. 12, pp. 3324–3337, 2009.

[10] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.

[11] G. Heigold, I. Moreno, S. Bengio, and N. Shazeer, "End-to-end text-dependent speaker verification," in *Proc. ICASSP*, 2016, pp. 5115–5119.

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

[13] J. Fierrez, J. Galbally, and J. O.-G. *et al.*, "BiosecurID: A Multimodal Biometric Database," *Pattern Analysis and Applications*, vol. 13, no. 2, pp. 235–246, 2010.

[14] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Do you need more data? the DeepSignDB on-line handwritten signature biometric database," in *Proc. 15th IAPR Int. Conference on Document Analysis and Recognition, ICDAR*, 2019.

[15] M. Diaz, M. Ferrer, D. Impedovo, M. Malik, G. Pirlo, and R. Plamondon, "A Perspective Analysis of Handwritten Signature Technology," *ACM Computing Surveys*, vol. 51, no. 6, pp. 1–39, 2019.

[16] J. Fierrez and J. Ortega-Garcia, *"On-Line Signature Verification", Handbook of Biometrics*. Springer, 2008, pp. 189–209.

[17] J. Ortega-Garcia, *et al.*, "MCYT Baseline Corpus: A Bimodal Biometric Database," *IEE Proceedings Vision, Image and Signal Processing*, vol. 150, no. 6, pp. 395–401, December 2003.

[18] N. Houmani, A. Mayoue, S. Garcia-Salicetti, *et al.*, "BioSecure Signature Evaluation Campaign (BSEC'2009): Evaluating On-Line Signature Algorithms Depending on the Quality of Signatures," *Pattern Recognition*, no. 3, pp. 993–1003, 2012.

[19] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Benchmarking Desktop and Mobile Handwriting across COTS Devices: the e-BioSign Biometric Database," *PLOS ONE*, vol. 12, 2017.