Pós-Graduação em Ciência da Computação

VICTOR LORENA DE FARIAS SOUZA

Writer-independent offline handwritten signature verification system based on the dichotomy transformation, prototype selection and feature selection

Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
http://cin.ufpe.br/~posgraduacao

Recife
2020

VICTOR LORENA DE FARIAS SOUZA

Writer-independent offline handwritten signature verification system based on the dichotomy transformation, prototype selection and feature selection

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação.

Área de Concentração: Inteligência Computacional

Orientador: Adriano Lorena Inácio de Oliveira

Coorientador: Robert Sabourin (École de Technologie Supérieure - Université du Québec, Montreal, Québec, Canada)

Coorientador: Rafael Menelau Oliveira e Cruz (École de Technologie Supérieure - Université du Québec, Montreal, Québec, Canada)

Recife

2020

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

S729w Souza, Victor Lorena de Farias
Writer-independent offline handwritten signature verification system based on the dichotomy transformation, prototype selection and feature selection / Victor Lorena de Farias Souza. – 2020.
113 f.: il., fig., tab.

Orientador: Adriano Lorena Inácio de Oliveira.
Tese (Doutorado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2020.
Inclui referências e apêndices.

1. Inteligência computacional. 2. Transformação dicotômica. I. Oliveira, Adriano Lorena Inácio de (orientador). II. Título.

006.31 CDD (23. ed.) UFPE - CCEN 2020 - 158

Victor Lorena de Farias Souza

"Writer-independent offline handwritten signature verification system based on the dichotomy transformation, prototype selection and feature selection"

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação.

Aprovado em: 12/08/2020.

_____

Orientador: Prof. Dr. Adriano Lorena Inacio de Oliveira

BANCA EXAMINADORA

_____

Prof. Dr. George Darmiton da Cunha Cavalcanti
Centro de Informática /UFPE

_____

Prof. Dr. Cleber Zanchettin
Centro de Informática / UFPE

_____

Prof. Dr. André Carlos Ponce de Leon Ferreira de Carvalho
Instituto de Ciências Matemáticas e de Computação / USP

_____

Prof. Dr. Luiz Eduardo Soares de Oliveira
Departamento de Informática / PUC-PR

_____

Prof. Dr. Jean Paul Barddal
Escola Politécnica / PUC-PR

I dedicate this thesis to all my family, friends and professors who gave me the necessary support to get here.

ACKNOWLEDGEMENTS

Acima de tudo agradeço a Deus pela motivação e pela força necessárias ao longo dessa jornada.

Agradeço, ainda, a todos aqueles que me apoiaram no desenvolvimento deste trabalho.

Um agradecimento especial aos meus pais Jorge Souza e Débora Farias, aos meus irmãos Jorge Lorena e Roberta Lorena e à minha namorada, Marcela Ximenes, por sempre acreditarem em mim, me incentivarem e me apoiarem em todos os momentos da realização deste trabalho.

Agradeço ainda aos Professores Adriano Oliveira e Robert Sabourin e ao Rafael Cruz, pelo apoio, confiança, oportunidade, críticas construtivas, orientações e excelentes conselhos fornecidos e nestes últimos anos.

Aos professores do Centro de Informática pelos ensinamentos passados.

Aos professores que aceitaram fazer parte da banca examinadora.

ABSTRACT

High number of writers, small number of training samples per writer with high intra-class variability and heavily imbalanced class distributions are among the challenges and difficulties of the offline Handwritten Signature Verification (HSV) problem. A good alternative to tackle these issues is to use a writer-independent (WI) framework. In WI systems, a single model is trained to perform signature verification for all writers from a dissimilarity space generated by the dichotomy transformation. Among the advantages of this framework is its scalability to deal with some of these challenges and its ease in managing new writers, and hence of being used in a transfer learning context. In this work, a deep analysis of this approach is presented, highlighting how it handles the challenges as well as the dynamic selection of reference signatures through fusion function, and its application for transfer learning. All the analyses are carried out using the instance hardness (IH) measure. By having these findings at the instance level, we develop an approach that uses prototype selection (Condensed Nearest Neighbors) and feature selection (based on Binary Particle Swarm Optimization) techniques well suited to our WI-HSV scenario. These techniques allowed us to handle the redundancy of information in both sample and the feature levels present in the dissimilarity space. Specifically in the feature selection scenario, we also propose a global validation strategy with an external archive to control overfitting during the search process. The experimental results reported herein show that the use of prototype selection and feature selection in the dissimilarity space allows a reduction in its redundant information and the complexity of the classifier without degrading its generalization performance. In addition, the results show that the WI classifier is scalable enough to be used in a transfer learning approach, with a resulting performance comparable to that of a classifier trained and tested in the same dataset. Finally, using the IH analysis, we were able to characterize "good" and "bad" quality skilled forgeries as well as the frontier region between positive and negative samples.

Keywords: Offline signature verification. Dichotomy transformation. Instance hardness. Prototype selection. Transfer learning. Feature selection.

RESUMO

Grande número de escritores, poucas amostras de treinamento por escritor, com alta variabilidade intra-classe e distribuições de classes fortemente desequilibradas, estão entre os desafios e as dificuldades da Verificação de Assinatura Manuscrita (HSV) offline. Uma boa alternativa para resolver esses problemas é usar um método independente de escritor (WI). Nos sistemas WI, um único modelo de classificação é treinado para executar a verificação de assinatura de todos os escritores a partir de um espaço de dissimilaridade gerado pela transformação dicotômica. Entre as vantagens dessa estrutura estão: a escalabilidade para lidar com alguns desses desafios listados e a facilidade no gerenciamento de novos escritores, e, portanto, a sua utilização em um contexto de transferência de aprendizado. Neste trabalho, apresentamos uma análise aprofundada dessa abordagem, destacando como ela lida com os desafios, a seleção dinâmica de assinaturas de referência por meio da função de fusão e sua aplicação na transferência de aprendizado. Todas as análises são realizadas usando a medida de dificuldade da instância (IH). Tendo por base os resultados dessas análises, desenvolvemos uma abordagem que usa técnicas de seleção de protótipos (vizinhos mais próximos condensados) e de seleção de características (com base na otimização de enxame de partículas binárias) adequadas ao nosso cenário WI-HSV. Essas técnicas nos permitiram lidar com a redundância de informações nos níveis das amostras e das características presentes no espaço de dissimilaridades. Especificamente no cenário de seleção de características, também propomos uma estratégia de validação global com um arquivo externo para controlar o overfitting durante o processo de busca. Os resultados experimentais relatados aqui mostram que o uso da seleção de protótipos e seleção de características no espaço de dissimilaridade permite uma redução em suas informações redundantes e na complexidade do classificador sem degradar seu desempenho de generalização. Além disso, os resultados mostram que o classificador WI é escalável o suficiente para ser usado em uma abordagem de aprendizado de transferência, com um desempenho resultante comparável ao de um classificador treinado e testado no mesmo conjunto de dados. Por fim, os resultados experimentais mostram que, utilizando a análise IH, conseguimos caracterizar falsificações especializadas de qualidade "boa" e "ruim", bem como a região fronteiriça entre amostras positivas e negativas.

Palavras-chaves: Verificação de assinaturas offline. Transformação dicotômica. Dificuldade da instância. Seleção de protótipos. Transferência de aprendizado. Seleção de características.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

BPSO	Binary Particle Swarm Optimization
CNN	Condensed Nearest Neighbors
DCNN	Deep Convolutional Neural Network
DS	Dissimilarity Space
DT	Dichotomy Transformation
EER	Equal Error Rate
FS	Feature Selection
HSV	Handwritten Signature Verification
IH	Instance Hardness
KNN	K-Nearest Neighbors
PS	Prototype Selection
PSO	Particle Swarm Optimization
SVM	Support Vector Machine
TL	Transfer Learning
WD	Writer-Dependent
WI	Writer-Independent

LIST OF SYMBOLS

μ	Adjustment factor
c_1	Cognitive parameter
$\varphi(t)$	Detection function
w_{final}	Final inertial weight
f	Fitness function
w	Inertial weight
$w_{initial}$	Initial inertial weight
K_{max}	Maximum number of iterations
$\mathbf{p}_i$	Particle position vector
$\mathbf{v}_i$	Particle velocity vector
c_2	Social parameter
ξ	Threshold
$\mathbf{x}_q$	Feature vector of the questioned signature
$\mathbf{x}_r$	Feature vector of the reference signature
$\mathbf{u}$	Dissimilarity vector

CONTENTS

1 INTRODUCTION

Handwritten Signature Verification (HSV) systems are used to automatically recognize whether the signature provided by a writer belongs to the claimed person (GURU et al., 2017). In offline HSV, the signature is acquired after the writing process is completed, and the system deals with the signature as an image. For instance, bank cheques and document authentication are among real-world applications using HSV systems (HAFEMANN; SABOURIN; OLIVEIRA, 2017b; ZOIS; ALEXANDRIDIS; ECONOMOU, 2019).

In the HSV problem, genuine signatures are the ones produced by the claimed person (original writer) and forgeries are those created by an impostor (forger). In general, forgeries can be categorized, based on the knowledge of the forger, into the following types (MASOUDNIA et al., 2019):

- Random forgeries: the forger has no information about the original writer.
- Simple forgeries: the forger knows the name of the original writer, but does not have access to the signature pattern.
- Skilled forgeries: the forger has information about both the name and the genuine signature pattern of the original writer, resulting in forgeries that are more similar to genuine signatures.

Figure 1 depicts examples of genuine signatures and skilled forgeries, obtained from (HAFEMANN; SABOURIN; OLIVEIRA, 2017a). Each column shows two genuine signatures from the same writer and a skilled forgery, from the GPDS dataset.

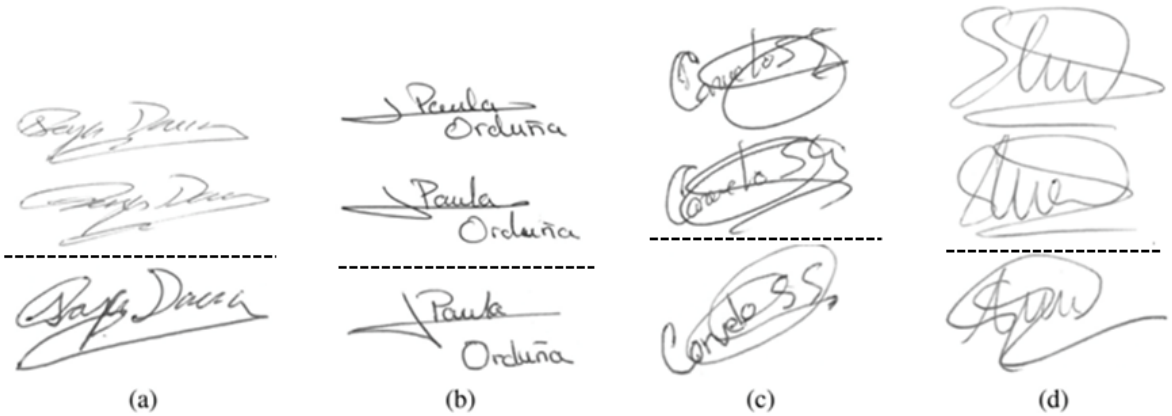

Fig. 1 – Signature examples from the GPDS dataset. Each column shows two genuine signatures from the same writer (above the line) and a skilled forgery (under the line).

A first aspect that should be considered when working with HSV is the decision about which classification strategy to use, that is, writer-dependent vs. writer-independent

(BOUAMRA et al., 2018). If a verification model is trained for each writer, the system is called Writer-Dependent (WD). This approach is the most commonly used and, in general, achieves better classification accuracies. However, requiring a classifier for each writer increases the complexity, and the computational cost of the system operations as more writers are added (ESKANDER; SABOURIN; GRANGER, 2013).

On the other hand, in Writer-Independent (WI) systems, a single model is trained for all writers. In this scenario, the systems usually operate on the dissimilarity space generated by the Dichotomy Transformation (DT) (RIVARD; GRANGER; SABOURIN, 2013). In this approach, a dissimilarity (distance) measure is used to compare signatures and a dichotomizer (two class classifier) is responsible for classifying them as belonging to the same writer or not. (ESKANDER; SABOURIN; GRANGER, 2013). When compared to the WD approach, WI systems are less complex, but in general obtain worse accuracy (HAFEMANN; SABOURIN; OLIVEIRA, 2017b).

Some of the challenges related to the offline HSV are: (C_1) the high number of writers (classes), (C_2) the high-dimensional feature space, (C_3) small number of training samples per writer with high intra-class variability (Figure 2 shows an example of this problem in the genuine signatures), and (C_4) the heavily imbalanced class distributions (HAFEMANN; SABOURIN; OLIVEIRA, 2017b; MASOUDNIA et al., 2019).

Fig. 2 – Overlaid genuine signature images of a single writer, illustrating the intra-class variability of the data.

Still, the main challenge is faced when dealing with skilled forgeries (C_5). Even though they are the most similar to genuine signatures, in general, they are not available for training purposes in real HSV applications. Thus, the systems are trained with partial knowledge as the classifiers are trained without sufficient information to distinguish between genuine signatures and skilled forgeries (HAFEMANN; SABOURIN; OLIVEIRA, 2017b).

The Dichotomy Transformation (DT), propopsed by Cha and Srihari (CHA; SRIHARI, 2000), can be applied to deal with some of these challenges and therefore facilitate the

signature verification task. The samples in the Dissimilarity Space (DS) generated by the dichotomy transformation are formed through the pairwise comparisons of signatures (a questioned and a reference signature) in the feature space (SOUZA; OLIVEIRA; SABOURIN, 2018).

Thus, the classification only depends on the input reference signature, by using the DT in a writer-independent approach, the model can verify signatures of writers for whom the classifier was not trained. So, it can easily manage new incoming writers (C_6). In this way, WI systems have the advantages of being scalable and adaptable (SHAO; ZHU; LI, 2015).

It is worth noting that, in the dissimilarity space generated by the WI dichotomy transformation, regardless of the number of writers, there are only two classes: (i) The positive class, composed of dissimilarity vectors computed from samples of the same writer. (ii) The negative class, composed of distances vectors computed from samples of different writers. Therefore, having a good feature representation of the signatures is very important for DT to work, i.e., to obtain little or no overlap between the positive and negative samples in the tranposed space (BERTOLINI et al., 2010).

Recently, Hafemann et al. (HAFEMANN; SABOURIN; OLIVEIRA, 2017a) proposed a new feature representation for the HSV problem. The approach carries out feature learning from the signature images in a WI format, using a Deep Convolutional Neural Network (DCNN) called *SigNet*. After being trained, the DCNN is used to extract representative features from the signatures. Their results showed a significant improvement in performance when compared to the previous state-of-the-art methods.

1.1 OBJECTIVES

The main objective of this thesis is to investigate how a writer-independent (WI) approach can handle the challenges of the HSV problem presented. Resulting in a general WI framework that uses targeted techniques to deal with each of the challenges presented. Due to its main characteristics, the *Signet* is used as the original feature space in this work (HAFEMANN; SABOURIN; OLIVEIRA, 2017a).

Considering the main characteristics of the WI dichotomy transformation, it is able to handle some challenges of the HSV problems that WD systems are not capable of. (i) The DT reduced the high number of classes to a 2-class problem. (ii) The problem is no longer imbalanced as both positive and negative classes may have the same number of samples. (iii) The small number of samples is no longer a problem. Since each dissimilarity vector generated by the DT is formed by the difference between the features of a questioned signature and a reference signature, this approach can increase the number of samples in the WI-HSV scenario. In this way, the dichotomy transformation naturally deals with some of the difficulties present in the HSV problem.

A disadvantage of the DT is that many of these samples may be redundant and have little influence for training the verification model. Thus, one goal of this study is to analyse the use of Prototype Selection (PS) techniques in the dissimilarity space to reduce the complexity and the training time of the classifier used without degrading its generalization (GARCIA et al., 2012).

In the WI case, a single model is trained for all users, and the classification depends solely on the input reference signature. Thus, another goal of this study is to analyse the use of a WI classifier in a transfer learning scenario, i.e., where the classifier is trained in one dataset, and is used to verify signatures in other datasets (SHAO; ZHU; LI, 2015).

In the considered feature representation, *SigNet* feature vectors are composed of 2048 dimensions (HAFEMANN; SABOURIN; OLIVEIRA, 2017a). However, when using this representation in a WI context, some of the features may be redundant and have little importance in the generated dissimilarity space (just like the redundant samples). In this context, another aspect the worth being analysed is whether Swarm optimization algorithms can be used for Feature Selection (FS) to obtain only the relevant dimensions on the transposed space (CRUZ; SABOURIN; CAVALCANTI, 2017)

Additional discussions on dichotomy transformation will be made throughout this study. With a further maturation of the related concept, it will be explained how this approach deals with the other challenges related to the faced problem.

The analyses of this study are also carried out based on the Instance Hardness (IH) measure (SMITH; MARTINEZ; GIRAUD-CARRIER, 2014) to maintain the findings at the instance level (LORENA et al., 2019). According to (SMITH; MARTINEZ; GIRAUD-CARRIER, 2014), understanding why instances are misclassified can lead to the development of learning algorithms that tackles directly the causes of the misclassification. In our scenario, the instance hardness is also used to characterize and analyse "good" and "bad" skilled forgeries.

1.2 RESEARCH QUESTIONS

Based on this context, the main research questions investigated in this thesis are:

1. How the writer-independent dichotomy transformation can handle the HSV data difficulties?

2. Does the number of reference signatures used influence the ability to verify signatures in the writer-independent model?

3. What is the best fusion function to be used to combine partial decisions in a scenario with multiple reference signatures?

4. Does the dissimilarity space generated by the dichotomy transformation have samples with redundant information, i.e., with little importance for training purposes? Can we use prototype selection methods for eliminating redundant training data?

5. Can the writer-independent approach be used in the context of transfer learning and still obtain good verification performance?

6. Can skilled forgeries be characterized as having "good" or "bad" quality based on the measure of instance hardness?

7. Does the generated dissimilarity space have redundant features?

8. Can overfitting control improve the performance of the optimization in the feature selection scenario?

1.3 CONTRIBUTIONS

The studies carried out in the present work stand out for presenting an innovative approach for offline writer-independent handwritten signatures verification in a image stream context.

The following papers were published during this research:

- SOUZA, V. L. F.; OLIVEIRA, A. L. I.; CRUZ, R. M. O.; SABOURIN, R. Improvingbpso-based feature selection applied to offline wi handwritten signature verification-through overfitting control. In:2020 Genetic and Evolutionary Computation Conference Companion. 2020. (GECCO '20), p. 69–70.

  – Contribution of this paper: Analysis on the redundancy of the features in the dissimilarity space and how the overfitting control can improve the performance of the optimization in the feature selection

- SOUZA, V. L. F.; OLIVEIRA, A. L. I.; CRUZ, R. M. O.; SABOURIN, R. A white-box analysis on the writer-independent dichotomy transformation applied to offline handwritten signature verification. Expert Systems with Applications, v. 154, p. 113397, 2020.

  – Contribution of this paper: Extension of studies related to the difficulties of the HSV problem data and the use of transfer learning. Also, the characterization of "good" and "bad" quality skilled forgeries.

- SOUZA, V. L. F.; OLIVEIRA, A. L. I.; CRUZ, R. M. O.; SABOURIN, R. Characterization of handwritten signature images in dissimilarity representation space. In: 2019 International Conference on Computational Science (ICCS). Springer International Publishing, 2019. p. 192–206

  - Contribution of this paper: Analysis on the difficulties of the HSV problem data and how writer-independent dichotomy transformation can handle it

- SOUZA, V. L. F.; OLIVEIRA, A. L. I.; CRUZ, R. M. O.; SABOURIN, R. On dissimilarity representation and transfer learning for offline handwritten signature verification. In: 2019 International Joint Conference on Neural Networks (IJCNN). 2019.

  - Contribution of this paper: Analysis on the use of prototype selection in the dissimilarity space and the use of the WI approach in a transfer learning context

- SOUZA, V. L. F.; OLIVEIRA, A. L. I.; SABOURIN, R. A writer-independent approach for offline signature verification using deep convolutional neural networks features. In: IEEE. 2018 7th Brazilian Conference on Intelligent Systems (BRACIS). 2018. p. 212–217.

  - Contribution of this paper: Analysis on the number of reference signatures and the fusion functions to be used during the verification.

1.4 ORGANIZATION

This thesis is organized as follows. In this chapter the introduction, the problem statement and the objectives for the accomplishment of this work were presented.

In chapter 2 the basic concepts and literature review related to this work are discussed. Firstly, the Handwritten Signature Verification Systems are presented, focusing on writer-independent approaches and on Dichotomy Transformation. Then, the feature representation used in this thesis is presented. After, prototype selection, transfer learning, instance hardness and feature selection which are themes involved in this study, are also addressed.

Chapter 3 presents the main contribution of this thesis: a deep analysis on the writer-independent (WI) dichotomy transformation applied to the offline handwritten signature verification problem. By having a good knowledge and mastery of the worked context we chose the fusion function, the prototype selection technique and the way to deal with transfer learning and the feature selection technique that best suit our problem.

Chapter 4 presents the experiments conducted in order to answer the research questions formulated in this thesis. In that chapter, the experimental protocol, experimental setup, the datasets used in the experiments and the main results obtained are presented and discussed.

In Chapter 5 the general conclusions and future works of this study are presented.

In Appendix A, we present a study on “good” and “bad” quality skilled forgeries at image level. Analyzes related to the location of these samples in the dissimilarity space and the respective instance hardness (IH) values are also presented. Next, in Appendix B and

C, we present the relationship of IH and the accuracy (%) of the writer-independent model, respectively, for the BRAZILIAN and the CEDAR datasets in the original dissimilarity space. In Appendix D, the relationship of IH and the accuracy (%) for the CEDAR dataset when considering the optimized feature space generated when using the proposed feature selection with global validation and external archive technique.

2 LITERATURE REVIEW

This chapter contains the literature review and description of the basic concepts related to the research areas of the proposed work, including 1) Handwritten Signature Verification Systems, 2) Feature representation, 3) Prototype Selection, 4) Feature Selection, 5) Transfer learning and 6) Instance Hardness.

In the Handwritten Signature Verification (HSV) Systems section, this problem is defined, and writer-dependent, writer-independent and hybrid approaches are discussed. More attention will be given to the writer-independent models, including the Dichotomy Transformation, as this approach is the one investigated in this thesis. In sequence, we will present how to use the Deep Convolutional Neural Networks DCNN proposed by (HAFEMANN; SABOURIN; OLIVEIRA, 2017a) to obtain features in the HSV context.

In the Prototype Selection section, we discussed a little about the Condensation, Edition and Hybrid approaches. Next, the main concepts related to Transfer learning and Instance Hardness are presented. Finally, Feature Selection based on Particle Swarm Optimization (PSO) is contextualized and the IDPSO (a variation of PSO - Particle Swarm Optimization) is presented.

2.1 HANDWRITTEN SIGNATURE VERIFICATION (HSV)

The handwritten signature is one of the oldest accepted biometric characteristics used to verify whether a person is whom he/she claims to be. The key task for an HSV system is deciding whether a given signature image is genuine or a forgery. Intuitively, genuine signatures are those that really belong to the indicated person and forgeries are those created by someone else (HAFEMANN; SABOURIN; OLIVEIRA, 2017b). The forgeries can be segmented into the following types (BHARATHI; SHEKAR, 2013): random forgeries, simple forgeries and skilled forgeries.

Authors recommend that random forgeries should be distinguished from other types of forgeries and should not affect the overall evaluation of the system (BOUAMRA et al., 2018).

Systems that deal with the offline HSV can be divided into Writer-Dependent (WD) or Writer-Independet (WI) systems. While in the first case, a classifier is trained for each writer, in WI systems a single model is trained for all writers from a dissimilarity space generated by the dichotomy transformation (DT) (SOUZA; OLIVEIRA; SABOURIN, 2018). In DT, a dissimilarity (distance) measure is used to compare two samples as belonging to the same writer or not (ESKANDER; SABOURIN; GRANGER, 2013). When compared to the WD approach, WI systems have the advantages of being less complex and more scalable, but in general, obtain worse accuracy (HAFEMANN; SABOURIN; OLIVEIRA, 2017b).

2.1.1 Writer-Dependent Approaches

(VARGAS et al., 2011) focused on the features extraction task in the WD handwritten signature verification context. The proposed features are based on statistical textural analysis of the gray level information from the handwritten signature images. The ink distribution in the signature strokes are also used to reduce the influence of different writing ink pens used by writers, since they provide information about rotation and luminance invariance. A WD Support Vector Machine (SVM) is trained on genuine and random forgeries samples and tested on random and skilled forgeries.

(BATISTA; GRANGER; SABOURIN, 2012) proposed a hybrid generative–discriminative ensemble of classifiers which dynamically selects the classifiers for building a writer-dependent HSV system. During the generative stage, the signatures are divided in a grid format and multiple discrete left-to-right Hidden Markov Models (HMMs) are trained with different number of states and codebook sizes, so the model can work at different levels of perception. Then, the HMM likelihoods for each enrolled signature are computed and grouped into a feature vector that is used through a specialized Random Subspace Method to build a pool of two-class classifiers (discriminative stage). For the verification task, the authors propose a new dynamic selection strategy based on the K-nearest-oracles (KNORA) algorithm and on Output Profiles (OP) to select the most accurate ensemble to classify the given signature.

(BHARATHI; SHEKAR, 2013) proposed a WD approach based on the combination of chain code histogram features improved by the Laplacian of Gaussian filter with a SVM classifier. The proposed chain code histogram approach can be divided as follows: (i) in the first step, the images are binarized and the noise eliminated, then signature contour is extracted. (ii) In sequence, a 4-directional chain code histogram is created on the grid of the extracted contour. (iii) At the end, the Laplacian of Gaussian filter is applied. The feature matrix obtained is then used to train a two-class SVM, which is responsible of performing the verification. While the training is based on genuine signatures and random forgeries, the verification is carried out on genuine signatures and skilled forgeries.

(SOLEIMANI; ARAABI; FOULADI, 2016) proposed a Deep Multitask Metric Learning (DMML) system for HSV. This approach combines Histogram of Oriented Gradients (HOG) and Discrete Radon Transform (DRT) with DMML. The system learns to compare two signatures, by learning a distance metric between them. The signatures are processed using a feedforward neural network, where the same weights are used in the bottom layers for all users and the last layer is specific and specializes for each individual.

(HAFEMANN; SABOURIN; OLIVEIRA, 2016) proposed a Deep Convolutional Neural Networks (DCNN) feature learning approach for the offline HSV problem, in order to obtain better feature representations than by using hand-crafted features. This approach can be divided in two steps: a Writer-Independent feature learning phase followed by Writer-Dependent classification, performed by a WD SVM. The feature learning phase uses a

surrogate classification task for learning the feature representations, where a DCNN is trained to discriminate between signatures from users not enrolled in the system. Then, this approach uses this DCNN as a feature extractor and trains a writer-dependent classifier for the verification of each user.

(HAFEMANN; SABOURIN; OLIVEIRA, 2017a) proposed an approach to deal with the offline HSV problem that uses concepts from both WI and WD systems. This model represents an evolution from the one cited above, with a more complex architecture and more complete experimental protocol. The approach carries out feature learning from the signature images in a WI format, using a Deep Convolutional Neural Network called *SigNet*. After being trained, the DCNN is used to extract representative features from the signatures, which are used to train a writer-dependent SVM classifier for each writer. The experimental results showed a significant improvement in performance when compared to the previous state-of-the-art methods.

(BOUAMRA et al., 2018) proposed a writer-dependent approach for offline HSV that employs One-Class Support Vector Machine (OC-SVM) with features based on run-length distributions of signatures binary images to perform the verification. The option of using OC-SVM is to simulate the real world scenarios, in which only genuine signatures are available for training the classifier model. In the training phase, the system searches for the optimal OC-SVM parameters - i.e., proportion of outliers (ν) and kernel parameter (γ) - and also selects the best decision threshold values (which is responsible for accepting or rejecting the signature as genuine) using a small subset of the development dataset. The results show that the proposed system is able to detect the skilled forgeries, especially when there is only one reference signature in the training set.

(OKAWA, 2018) proposed an offline signature verification system based on bag-of-visual words (BoVW) and a vector of locally aggregated descriptors (VLAD) to detect salient regions of the signature structure. Then, KAZE features are used to obtain information about the contours of strokes and the relationships between strokes. The KAZE features approach is a multiscale 2-D feature detection and description algorithm that is applied to nonlinear scale spaces. Finally, principal component analysis (PCA) is used to reduce the data dimensionality and the SVM is responsible for verifying signatures.

The HSV system proposed by (ZOIS et al., 2019) utilizes a Sparse Representation (SR) in order to learn local features and construct a global signature descriptor. In this study, the authors investigate the selection of the appropriate SR approach, which can be segmented in greedy and convex relaxation. The effects of the associated parameters, the sparsity level and the regularization function are also evaluated. Results showed that, greedy techniques can deploy the full potential of SR in a signature verification system.

2.1.2 Writer-Independent Approaches

In the WI scenario, (BERTOLINI et al., 2010) proposed a writer-independent approach for handwritten signature verification. This approach applies the ideas of the dissimilarity representation and SVMs as classifiers. The two main contributions of the authors are the following: (i) introduce a new graphometric feature set based on the curvature of the most important segments, simulated by using Bezier curves. (ii) The use of an ensemble of classifiers structure to improve the reliability and to reduce the false acceptance of the model. This ensemble is built using a standard genetic algorithm and a pool of base classifiers trained with four different graphometric feature sets.

(KUMAR; SHARMA; CHANDA, 2012) proposed a new feature set based on the surroundedness property for writer-independent HSV. The verification is performed by two-class classifiers, for example, RBF-SVM (Support Vector Machine with RBF kernel) or MLP (Multilayer perceptron). The proposed approach was able to find distinctive and representative features that represent both shape and texture properties of the signature. The shape of the signature is computed by considering the distribution of surrounded signature black pixels. The texture is computed through the correlation between a questioned signature pixel and the reference signature pixels. Experimental results indicate that the proposed feature set was sufficiently general to handle data.

(RIVARD; GRANGER; SABOURIN, 2013) have proposed a writer-independent approach that combines multiple feature extraction, Dichotomy Transformation (DT) and boosting feature selection. The authors report that the accuracy and reliability of the system can be improved by integrating features from different sources of information. Initially the authors employ some techniques to extract features at different scales. Then, the Dichotomy Transformation, which reduces the pattern recognition problem to a 2-class problem, is used. A good point that deserves to be highlighted is that with this transformation the system alleviates the challenges of dealing with limited number of reference signatures from a large number of users. Finally, an ensemble is built using boosting feature selection that uses low-cost classifiers capable of automatically selecting relevant features during training.

(HAMADENE; CHIBANI, 2016) proposed a WI framework for HSV using both the Contourlet Transformation (CT) and the Feature Dissimilarity Measure (FDM) thresholding for classification. The CT describes the writer's handwriting style through the following characteristics of each user: (i) which directions are contained in signatures, (ii) the amount of each direction, (iii) the spatial distribution of the directions toward users. The classification is performed by using straightforward FDM thresholding and the writer-independent threshold is defined using a signature stability criterion. This criterion is based on users genuine-genuine dissimilarities stability and the most appropriate frontier of the stable signatures is selected. This stability principle is based on considering a signature pair as more stable when their feature dissimilarity is lower. Experimental results

showed that the proposed system was able to perform the verification even with a unique threshold for accepting or rejecting a questioned signature, a reduced number of writers and a limited number of reference signatures.

(RANTZSCH; YANG; MEINEL, 2016), proposed a writer-independent approach based on deep similarity metrics. To this end, the signatures are embedded into a high-dimensional space and their Euclidean Distance, in that space, can be used to compute their similarity. Therefore, since genuine signatures generated from the same writer are more similar, they are embedded close to each other and the forgeries are embedded far from them. This can be obtained by training a Deep Neural Network (DNN) using a triplet-based loss function. The experimental results showed that the proposed approach was able to outperform the state-of-the-art from the ICDAR SigWiComp 2013 challenge on offline signature verification.

(GUERBAI; CHIBANI; HADJADJI, 2015) proposed a writer-dependent HSV system based on One-Class SVM (OC-SVM) that tries to reduce the difficulties of having a large numbers of users. As a one-class classification problem, the proposed approach models only one class (genuine signatures), which is a good characteristic, as, in general, the system only has the genuine signatures for each writer to train the classifier. Nevertheless, the low number of genuine signatures is still an important challenge.

(DUTTA; PAL; LLADOS, 2016) proposed an approach that uses hybrid features that consider the spatial information between local features and their global statistics in the signature image to perform the handwritten signature verification in a writer-independent way. To avoid the excess of computational burden when learning the condensed set of higher order neighbouring features based on visual words, the authors also create a code of local pairwise features which are represented as joint descriptors. Finally, local features are paired based on the edges of a graph representation built upon the Delaunay triangulation to to perform the verification task.

(ZOIS; ALEXANDRIDIS; ECONOMOU, 2019) proposed a writer independent framework that uses the dissimilarity approach on a new feature set obtained by detecting and counting asymmetric first order $5x5$ pixel mask transitions, instead of transitions between pixel assortments. By doing this, the proposed model obeys the inclusion property. A decision stump committee with Gentle AdaBoost (DSC-BFS) framework, which also performs feature selection, was used to perform the verification. According to the authors, the discriminating capabilities of the employed features are related to the high verification performance achieved.

2.1.3 Hybrid WD-WI Approaches

Some authors use a combination of both WD and WI approaches. For example, (ESKANDER; SABOURIN; GRANGER, 2013) proposed a hybrid writer-independent-writer-dependent model. The aim of the authors was to take advantage of the positive character-

istics of each approach. In the scenario where only a few genuine signatures are available, they use the writer-independent classifier to perform the verification. On the other hand, the writer-dependent classifier is trained for a user when the number of genuine samples is above a defined threshold.

(HU; CHEN, 2013) proposed a HSV approach that combines multiple features with different classifications methods. It is worth noting that in the proposed approach, the verification task can be performed in both WI and WD forms. The multiple used features are based on local binary pattern (LBP), gray level co-occurrence matrix (GLCM) and histogram of oriented gradients (HOG). In turn, SVM classifiers are used as WD or Global Real Adaboost as WI classifiers to perform the verification. In the first mode, each WD SVM is trained using the feature vectors extracted from the reference signatures of the corresponding user, as positive samples, and random forgeries of each other writer, as negative samples. Differently, in the WI approach, a global Adaboost classifier is trained using genuine and random forgery signatures of writers that are not included in the test set. It is important to highlight that the use of all the features improves the overall verification performance.

(YILMAZ; YANIKOĞLU, 2016) also propose a hybrid approach using the main ideas from the WD and WI approaches, aiming to learn the importance of different dissimilarities, the writer-independent classifier is trained with dissimilarity vectors of query and reference signatures of all users. In its turn, the writer-dependent classifiers are trained separately for each user, to learn to differentiate genuine signatures and forgeries. The results are then combined using a score-level fusion of these complementary classifiers with different local features.

(ZHANG; LIU; CUI, 2016) proposed a multi-phase approach that combines unsupervised feature extraction and a hybrid WI-WD classification method to perform the signature verification. The feature extraction is based on Deep Convolutional Generative Adversarial Networks (DCGANs). If few samples per writer are available, a Gentle Adaboost is employed in WI way. Once enough samples are collected, then it operates in WD way. So, the more query samples are tested and enrolled into the system, the more accurate the system will be, theoretically. The authors defend the idea that in the long run the unsupervised feature learning of signature verification has the potential to outperform supervised training as it has access to more data.

2.1.4 Offline HSV datasets

Table 2 presents a summary of the most commonly used signature datasets, which are BRAZILIAN (FREITAS et al., 2000), CEDAR (KALERA; SRIHARI; XU, 2004), MCYT (ORTEGA-GARCIA et al., 2003) and GPDS-300 (VARGAS-BONILLA et al., 2007). In the BRAZILIAN, forgeries are available only for 60 users, that is why the number of users are segmented.

Table 2 – Summary of the used datasets.

Dataset Name	Users	Genuine signatures (per user)	Forgeries per user
BRAZILIAN	60 + 108	40	10 simple, 10 skilled
CEDAR	55	24	24
MCYT	75	15	15
GPDS-300	881	24	30

Table 10 summarizes the used Exploitation set ε for each dataset. The number of questioned signatures per writer is acquired according to the literature (HAFEMANN; SABOURIN; OLIVEIRA, 2017a).

Table 3 – Exploitation set ε

Dataset	#Samples	#questioned signatures (per writer)
BRAZILIAN	2400	10 genuine, 10 random, 10 simple, 10 skilled
CEDAR	1650	10 genuine, 10 skilled, 10 random
MCYT	2250	5 genuine, 15 skilled, 10 random
GPDS-300	9000	10 genuine, 10 skilled, 10 random

2.2 FEATURE REPRESENTATION

The *SigNet*, proposed by (HAFEMANN; SABOURIN; OLIVEIRA, 2017a), uses Deep Convolutional Neural Networks (DCNN) for learning the signature representations in a writer-independent way and, nowadays, represents a state of the art approach in this research area. This approach tries to build a new representation space in which different writers are clustered in separate regions, based on the most representative properties of the handwritten signatures. To achieve this, the DCNN is trained by minimizing the negative log likelihood of the correct writer given the signature image. Table 4 summarizes the DCNN architecture used by the *SigNet* model.

Table 4 – Summary of the *SigNet* layers

Layer	Size	Other Parameters
Input	1 x 150 x 220	
Convolution (C1)	96 x 11 x 11	Stride = 4, pad = 0
Pooling	96 x 3 x 3	Stride = 2
Convolution (C2)	256 x 5 x 5	Stride = 1, pad = 2
Pooling	256 x 3 x 3	Stride = 2
Convolution (C3)	384 x 3 x 3	Stride = 1, pad = 1
Convolution (C4)	384 x 3 x 3	Stride = 1, pad = 1
Convolution (C5)	256 x 3 x 3	Stride = 1, pad = 1
Pooling	256 x 3 x 3	Stride = 2
Fully Connected (FC6)	2048	
Fully Connected (FC7)	**2048**	
Fully Connected + Softmax ($P(\mathbf{y}\|X)$)	M	

In the paper by (HAFEMANN; SABOURIN; OLIVEIRA, 2017a), the authors present another DCNN architecture, called as *SigNet-f*, which uses skilled forgeries during the feature

learning process. Our option of using the *SigNet* is due to the fact that it is not reasonable to expect skilled forgeries to be available in the training phase for all users enrolled in the system.

In *SigNet* the features are learned from the development set of the GPDS dataset. For new writers (from the GPDS dataset itself or from another dataset), this approach is used to project the signature images onto the new representation space, by using feed-forward propagation until the FC7 layer, obtaining feature vectors with 2048 dimensions. Also, as a writer-independent approach, it has the advantage of not being specific for a particular set of writers, it can even be used to extract features for writers from other databases.

In their study, (HAFEMANN; SABOURIN; OLIVEIRA, 2017a) analysed the local structure of the learned feature space, by using the t-SNE algorithm in a subset containing 50 writers from the development set of the GPDS-300 dataset (referred to as the validation set for verification V_v). Figure 3 represents this analysis (HAFEMANN; SABOURIN; OLIVEIRA, 2017a).

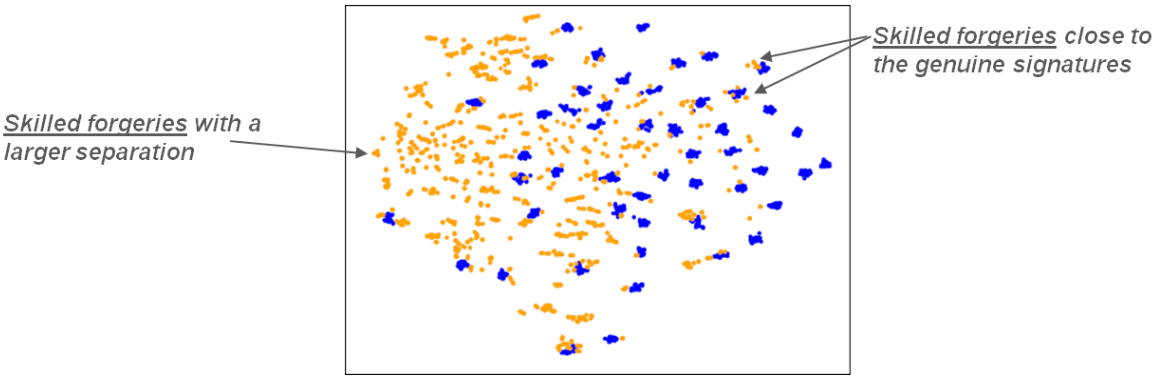

Fig. 3 – t-SNE 2D feature vector projections from the 50 writers of V_v. While blue points represent genuine signatures, orange points represent skilled forgeries

As depicted in Figure 3, in this feature space, for each writer, genuine signatures form compact clusters. According to (HOUMANI et al., 2011), the forgery quality measures the proximity of a forgery to a target signature. Thus, as highlighted, skilled forgeries come up with two different behaviors: (i) in some cases they have a larger separation from the genuine signatures. These forgeries are referred to as "bad quality skilled forgeries" in this thesis. (ii) For some writers, the skilled forgeries are closer to the genuine signatures - we call them "good quality skilled forgeries".

In this work, our original feature space is represented by the 2048 features obtained from the FC7 layer of the *SigNet* (HAFEMANN; SABOURIN; OLIVEIRA, 2017a) (no training or updating of *SigNet* has been carried out by us, the features for the considered datasets are available online[1]). This model was chosen mainly because of its behavior, characterized by different writers clustered in separate regions of the feature space.

[1] <http://en.etsmtl.ca/Unites-de-recherche/LIVIA/Recherche-et-innovation/Projets/Signature-Verification>

2.3 PROTOTYPE SELECTION (PS)

Prototype Selection (PS) approaches generally aim to obtain a representative training subset, with a lower number of samples as compared to the original one ($SelectedSubset \subseteq TrainingSet$) (GARCIA et al., 2012). Using the selected subset (from PS) can result in a similar or even higher classification accuracy for new incoming data (GARCIA et al., 2012).

According to Pekalska et al. (PEKALSKA; DUIN; PACLIK, 2006), prototype selection is an important aspect that should be considered for dissimilarity-based classification. In their paper, the authors showed that by using a few but well-chosen selected prototypes, it is possible to speed up the classifier training and still achieve a classification performance that is similar to or better than what is obtained by using all the training samples together. The authors also showed that, in general, a systematic prototypes selection approach works better than a random subsampling. To the best of our knowledge, this analysis has never been conducted specifically for the dichotomy transformation scenario.

According to (GARCIA et al., 2012), one segmentation when dealing with prototype selection techniques is related to the type of search. In this scenario, the concern is whether the PS approach seeks to keep the samples in the border region, samples that are far from the border, or some other set of samples. Thus, these approaches can be divided into: Condensation, Edition and Hybrid approaches.

- **Condensation Approaches:** These strategies aim to retain the samples which are closer to the decision boundaries, since internal samples do not affect the decision boundaries as much as border samples, and thus can be removed with relatively little effect on classification. One characteristic of this kind of approach is that the reduction capability is normally high due to the fact that there are fewer border points than internal points in most of the data.

- **Edition Approaches:** In contrast to condensation approaches, edition approaches try to remove border samples. The idea is to remove samples that do not agree with their neighbors, as they have a higher risk of being noise. Thus, the editing process occurs in regions of the space with a high degree of overlap between classes, aiming to obtain a smoother decision boundaries.

- **Hybrid Approaches:** The idea of this kind of approach is to combine the removal of both border and internal samples, based on the two previous strategies.

Both the papers by Garcia et al. (GARCIA et al., 2012) and Triguero et al. (Triguero et al., 2012) present a list of methods for each of these approaches.

2.4 TRANSFER LEARNING (TL)

Transfer Learning (TL) methods are based on the idea of utilizing the knowledge acquired from previously learned tasks and applying them to solve newer, related ones (SHAO; ZHU; LI, 2015). (PAN; YANG, 2010) present a formal definition for transfer learning. Given a source domain D_S and a learning task T_S, a target domain D_T and a learning task T_T, transfer learning aims to help improve the learning of the target predictive function $f_T(\cdot)$ in D_T using the knowledge obtained from D_S and T_S, where $D_S \neq D_T$, or $T_S \neq T_T$.

Following their notation, our context is related to the scenario where the target and the source domains are the same, i.e., $D_S = D_T$, and the learning tasks T_S and T_T are different. Specifically our case is that in which the conditional probability distributions of the domains are different, i.e., $P(Y_S|X_S) \neq P(Y_T|X_T)$, where Y_{S_i} and Y_{T_i} belong to the same label space formed by the positive and negative classes of the dichotomy transformation.

(PAN; YANG, 2010) suggest the following issues when dealing with transfer learning:

- What to transfer: the concern is related to which part of the knowledge may be common between the different domains, and so, may actually be useful to improve the performance in the target domain.

- How to transfer: methodologies need to be developed to deal with problems that may appear, such as the data distribution mismatch. Mining shared patterns from different domains, for instance, can significantly reduce the difference in the distribution between the target and the source domains

- When to transfer: considers in which situations TL should be used. When the domains are not related to each other, brute-force transfer may not succeed and/or even negatively affect the performance of learning in the target domain (situation knows as negative transfer).

2.5 INSTANCE HARDNESS (IH)

The Instance Hardness (IH) measure is used to identify hard to classify samples (SMITH; MARTINEZ; GIRAUD-CARRIER, 2014). According to the paper by (LORENA et al., 2019), an advantage of using the IH is to understand the difficulty of a problem at the instance level, rather than at the aggregated level with the entire dataset. Also, Smith et al. (SMITH; MARTINEZ; GIRAUD-CARRIER, 2014) argue that understanding why instances are misclassified can lead to the development of learning algorithms that tackles directly the causes of the misclassification.

For instance, in (SMITH; MARTINEZ; GIRAUD-CARRIER, 2014), the author uses the instance hardness to decrease the chance of overfitting the data and, thereby, obtain a more representative boundary of the data. For this, the authors remove the instances with

a high instance hardness (i.e., a high degree of overlap between classes) from the data sets before training the classifiers.

In the paper by (CRUZ et al., 2017), the authors used IH to identify the scenarios where an ensemble with dynamic selection techniques outperform the K-NN classifier.

In (WALMSLEY et al., 2018), the authors propose an ensemble generation method based on Bagging and instance hardness. The main idea is to remove outliers and noisy instances from the training set, maintaining instances that are close to border regions. Thus, the probability of an instance being picked to compose the bootstrapped training sets is defined to be inversely proportional to its hardness.

To empirically analyze hard-to-classify instances, the paper by (SMITH; MARTINEZ; GIRAUD-CARRIER, 2014) designed a set of seven heuristics (hardness heuristics).

- The k-Disagreeing Neighbors (kDN) gives the percentage of the k nearest neighbors that do not share the label of a given instance.
- The Disjunct Size (DS) corresponds to the size of a disjunct that covers an example divided by the largest disjunct produced. The disjuncts are obtained using a C4.5 decision tree learning algorithm.
- The Disjunct Class Percentage (DCP) can be computed by dividing the number of data points in a disjunct that belong to a same class by the total number of examples in the disjunct.
- The Class Likelihood (CL) global measure of overlap of a given instance belonging to a especific class.
- Class Likelihood Difference (CLD) is obtained by computing the difference between the class likelihood of an instance and the maximum likelihood for all of the other classes.
- The Minority Value (MV) is the ratio of the number of instances that shares the class value of a given instance to the number of instances in the majority class. This metric is used to obtain the skewness of the class to which an instance belongs.
- The Class Balance (CB) computes the balance between classes. If the data set is completely balanced the class balance value will be 0.

2.6 FEATURE SELECTION (FS)

In general, real-world machine learning problems involve a large number of features. However, not all features are essential and may have redundant or irrelevant information. In a classification algorithm, for instance, redundant features may reduce the performance of the classifier. Feature selection (FS) techniques deal with this problem by selecting only

a subset of relevant features from the original large set of features (XUE et al., 2015). The motivations for using this approach include: reduction of the computational complexity, reduction of dimensionality, removal of non-informative features, enhanced generalization power by reducing overfitting (CRUZ; SABOURIN; CAVALCANTI, 2017).

The feature selection task is becoming more challenging as the number of features and the complexity of the problems are increasing in many areas with the advances in the data collection techniques. An exhaustive search for the best subset of features is too costly and practically impossible for most problems. Some feature selection techniques, such as complete search, greedy search, heuristic search, and random search have been used (XUE et al., 2015). However, these techniques still suffer from problems of having a high computational cost and/or getting stuck in local optima (UNLER; MURAT, 2010) and, so, an efficient global search approach is need. The Particle Swarm Optimization (PSO) is well-known for it global search ability and have received a lot of attention in the feature selection scenario.

Based on the evaluation criteria, feature selection algorithms are generally classified into two categories: 1) filter approaches and 2) wrapper approaches (GUYON; ELISSEEFF, 2003). While filter methods evaluate features based on their intrinsic characteristics (independent of any classification algorithm), wrapper methods use the classification accuracy of a trained classifier to evaluate the feature subset. Although wrapper methods are typically more time consuming than filter methods, in general, they achieve better classification performance (TRAN; ZHANG; XUE, 2016).

In the paper by (TRAN; ZHANG; XUE, 2016), the authors developed an approach that combines wrapper and filter strategies and uses PSO. For this, the authors made changes in the fitness function and in the local search. A new hybrid fitness function, based on both classification accuracy and distance to promote higher discriminating feature subsets, is used to better evaluate candidate solutions. The local search is guided by a filter measure responsible for adding more relevant features to the current *pbest* of each solution.

In (CRUZ; SABOURIN; CAVALCANTI, 2017) the authors used a Binary PSO (BPSO) to select the relevant features in a meta-learning context. In this way, the meta-feature selection scheme based on BPSO is applied to optimize the performance of the meta-classifier in a wrapper mode.

The paper by (TRAN; XUE; ZHANG, 2018) proposes a variable-length PSO representation for feature selection. In this way, each particle from the swarm may have a different number of dimensions when compared to the others, which improves the performance of PSO. By sorting features in a descending order of importance, the proposed approach facilitate particles with a lower number of dimension to achieve better classification performance. Also, using the proposed variable-length mechanism, PSO can jump out of local optima, focusing its search on smaller and more fruitful area. Thus, this strategy enables PSO to reach better solutions in a shorter time.

2.6.1 Binary Particle Swarm Optimization (BPSO)

Particle Swarm Optimisation (PSO) (KENNEDY; EBERHART, 1995) is a global search technique that simulates the social behaviors of birds flocking. In PSO, a swarm consists of a set of candidate solutions (particles) moving in the solution search space. Each particle is encoded as its position ($\mathbf{p}_i$) and move through the space based on its velocity ($\mathbf{v}_i$), seeking to find better solutions (evaluated by a fitness function). During the search process, each particle has access to information about the best position found by it so far ($pBest$) and the best position found throughout the cluster ($gbest$).

In a context of feature selection, particle swarm optimization algorithms are used in their binary version - Binary Particle Swarm Optimization (BPSO) - and have been obtaining good results when compared to other optimization algorithms used for this task (CHUANG; TSAI; YANG, 2011). The transformation of the continuous search space into a binary space is conducted by using a transfer function, T (MIRJALILI; LEWIS, 2013).

Mirjalili et al. (MIRJALILI; LEWIS, 2013) state that an important aspect to obtain good performance on convergence is the choice of the well suited transfer function. According to the authors, in general, the V-Shaped transfer functions present better behavior both in terms of avoiding local minima and convergence speed (MIRJALILI; LEWIS, 2013). Also, in Cruz et al. (CRUZ; SABOURIN; CAVALCANTI, 2017), the V-Shaped function presented the best overall performance.

For a formal definition, given a binary search space with D dimensions and a swarm with N particles, the i-th particle of the swarm can be represented by a D-dimensional vector $\mathbf{p}_i = [p_{i1}; p_{i2}; ...; p_{iD}]$, which corresponds to the position of the particle in space. In this work context, each dimension p_{id} represents a single feature and value "1" means that the respective feature is selected and "0" otherwise. The particle velocity consists of $\mathbf{v}_i = [v_{i1}; v_{i2}; ...; v_{iD}]$; the best position found by the particle as $\mathbf{pBest}_i = [pBest_{i1}; pBest_{i2}; ...; pBest_{iD}]$ and the best position obtained by the swarm as $\mathbf{gBest} = [gBest_1; gBest_2; ...; gBest_D]$. Then, for each iteration, the update of the velocity and the position are computed, respectively, by equations 2.1 and 2.2 (ZHANG; XIONG; ZHANG, 2013).

$$\begin{aligned}\mathbf{v}_i(t+1) = w \cdot \mathbf{v}_i(t) + c_1 \cdot rand \cdot (\mathbf{pBest}_i - \mathbf{p}_i(t)) \\ + c_2 \cdot Rand \cdot (\mathbf{gBest}_i - \mathbf{p}_i(t))\end{aligned} \tag{2.1}$$

$$\mathbf{p}_i(t+1) = \begin{cases} \mathbf{p}_i(t)^{-1} & If \ \ rand_p < T(\mathbf{v}_i(t+1)) \\ \mathbf{p}_i(t) & If \ \ rand_p \geq T(\mathbf{v}_i(t+1)) \end{cases} \tag{2.2}$$

where, c_1 and c_2 represent acceleration factors and are positive constants; $rand$, $Rand$ and $rand_p$ are random variables with uniform distribution within the interval $[0, 1]$, and w is the weight of inertia. In the velocity equation, the first factor represents inertia, the second factor the cognitive component and the third factor the social component.

As we are dealing with a binary search space, updating the position of a particle means switching between selecting the feature (“1”) or not (“0”). The particle velocity is responsible for driving this, thus, the higher the velocity, the higher the probability of the particle to change its position.

The good overall performance in previous works (MIRJALILI; LEWIS, 2013) motivated our option to use the V-Shaped transfer function in this work. It can be computed through equation 2.3.

$$T(x) = \left| \frac{2}{\pi} arctan(\frac{\pi}{2} x) \right| \tag{2.3}$$

An important aspect is that this transfer function encourage particles to stay in their current positions when their velocity values are low or switch to their complements when the velocity values are high.

3 PROPOSED METHOD

We listed the challenges in dealing with the HSV problem in Section 1. Among them: (C_1) the high number of writers (classes), (C_2) the high-dimensional feature space, (C_3) small number of training samples per writer with high intra-class variability, (C_4) the heavily imbalanced class distributions, (C_5) no skill forgeries during training, even though they are the most similar to genuine signatures, in general, they are not available for training purposes in real HSV applications, (C_6) new incoming writers.

In this section we provide a deep analysis on the writer-independent (WI) dichotomy transformation applied to the offline handwritten signature verification problem. Having a better understanding of the generated dissimilarity space allowed us to make choices of methods that best fit our problem.

3.1 WRITER-INDEPENDENT DICHOTOMY TRANSFORMATION FOR HANDLING HSV DATA DIFFICULTIES

The Dichotomy Transformation (DT) approach (CHA; SRIHARI, 2000), allows to transform a multi-class pattern recognition problem into a 2-class problem. In this approach, a dissimilarity (distance) measure is used to determine whether a given reference signature and a questioned signature as belonging to the same writer or not (ESKANDER; SABOURIN; GRANGER, 2013).

Formally, let $\mathbf{x}_q$ and $\mathbf{x}_r$ be two feature vectors in the feature space, the dissimilarity vector resulting from the Dichotomy Transformation, $\mathbf{u}$, is computed by equation 3.1:

$$\mathbf{u}(\mathbf{x}_q, \mathbf{x}_r) = \begin{bmatrix} |x_{q1} - x_{r1}| \\ |x_{q2} - x_{r2}| \\ \vdots \\ |x_{qn} - x_{rn}| \end{bmatrix} \tag{3.1}$$

where $|\cdot|$ represents the absolute value of the difference, x_{qi} and x_{ri} are the i-th features of the signatures $\mathbf{x}_q$ and $\mathbf{x}_r$ respectively, and n is the number of features. Hence, each dimension of the $\mathbf{u}$ vector is equal to the distance between the corresponding dimensions of the vectors $\mathbf{x}_q$ and $\mathbf{x}_r$, and therefore all these vectors have the same dimensionality (BERTOLINI; OLIVEIRA; SABOURIN, 2016).

As mentioned, regardless of the number of writers, after applying DT, only two classes are present in the dissimilarity space:

- The *within/positive class* w_+: the intraclass dissimilarity vectors, i.e., obtained when the questioned and the reference signatures belong to the same writer.

- The *between/negative class* w_{-}: the interclass dissimilarity vectors, i.e., obtained when the questioned and the reference signatures belong to different writers.

Once the data is transposed into the dissimilarity space, a 2-class classifier (known as dichotomizer) is trained and used to perform the verification task. A common practice for WI systems is to use disjoint subsets of writers to train the classifier and to perform the verification. In general, the training set is known as the development set D and the test set as exploitation set ε (CHA; SRIHARI, 2000).

The Dichotomy Transformation has already been used in various contexts, such as: bird species identification (ZOTTESSO et al., 2018), forest species recognition (MARTINS et al., 2015), writer identification (BERTOLINI; OLIVEIRA; SABOURIN, 2016) and also for handwritten signature verification (RIVARD; GRANGER; SABOURIN, 2013; ESKANDER; SABOURIN; GRANGER, 2013; SOUZA; OLIVEIRA; SABOURIN, 2018).

Based on the DT definition we can highlight the following points: (C_1) first of all, the DT reduces the high number of classes (writers) to a 2-class problem, and only one model is trained to perform the verification for all writers from the dissimilarity space (DS) generated by the dichotomy transformation (ESKANDER; SABOURIN; GRANGER, 2013). (C_6) The WI verification only depends on the reference signature used as input to the classifier; it means that the WI framework is scalable and can easily manage new incoming writers without requiring additional training or updating of the model (unlike the WD approach, where a new classifier needs to be trained). In this way, a WI classifier trained in one dataset can be used to verify signatures from other datasets in a transfer learning task. In this scenario, the different datasets would represent samples that belong to the same domain (signature representations in DS). As defined before, given that the development set D and the exploitation set ε are disjoint, by default this approach already operates by using transfer learning.

An important property of the dichotomy transformation is its ability to increase the number of samples in the dissimilarity space since it is composed of each pairwise comparisons of signatures. Thus, if M writers provide a set of R reference signatures each, Equation 3.1 generates up to $\binom{MR}{2}$ different distances vectors. Of these, $M\binom{R}{2}$ are from the positive class and $\binom{M}{2}R^2$ belong to the negative class (RIVARD; GRANGER; SABOURIN, 2013). Therefore, even with a small number of reference signatures per writer, DT can generate a large amount of samples in DS.

In this way, the model can handle the small number of samples per class. Also, by increasing the number of samples, the model may be able to obtain sufficient information to capture the full range of signature variations, reducing the effects of the intra-class variability (HAFEMANN; SABOURIN; OLIVEIRA, 2017b) (C_3). Besides, by generating the same number of samples for both the positive class (questioned signatures are the genuine signatures from the writers) and the negative class (questioned signatures are the random forgeries), the model can manage the dataset imbalance (C_4).

However, many of the samples generated by DT in the WI-HSV scenario represent redundant information and therefore have little importance for training purposes. In this way, prototype selection (PS) techniques can be used in the dissimilarity space to reduce the complexity and the training time of the classifier used without degrading its generalization (GARCIA et al., 2012).

Another aspect is faced when writers have more than one reference signature. In this case, the pairwise comparison of DT is applied considering the questioned signature and each of the references, producing a set of dissimilarity vectors $\{\mathbf{u}_r\}_1^R$, where R is the number of reference signatures belonging to the writer. Thus, the dichotomizer evaluates each dissimilarity vector individually and produces a set of partial decisions $\{f(\mathbf{u}_r)\}_1^R$ (RIVARD; GRANGER; SABOURIN, 2013). The final decision about the questioned signature is based on the fusion of all partial decisions by a function $g(\cdot)$ and depends on the output of the dichotomizer. For discrete output classifiers, the majority vote can be used; whereas for distance or probability outputs, the max, mean, median, min and sum functions may be applied (RIVARD; GRANGER; SABOURIN, 2013).

Finally, one possible drawback of DT is that, perfectly grouped writers in the feature space may not be perfectly separated in the dissimilarity space (CHA; SRIHARI, 2000). Thus, the greater the dispersion between sample distributions among the writers, the less the dichotomizer is able to detect real differences between similar signatures (RIVARD; GRANGER; SABOURIN, 2013).

Summarizing, based on the main properties of the WI dichotomy transformation, this approach can handle some data difficulties of the HSV problems that WD systems are not capable of. Other characteristics of DT can be found in (CHA; SRIHARI, 2000; RIVARD; GRANGER; SABOURIN, 2013).

To facilitate the understanding of DT, Figure 4 (left) depicts a synthetic 2D feature space with synthetic data (containing genuine signatures and skilled forgeries from 3 different writers); on the right the respective dichotomy transformation is shown. The skilled forgeries in the feature space for each writer are presented in red with the same marker. These data were generated based on what was observed in Figure 3. The reader should keep in mind that although the negative samples in the dissimilarity space are represented by different colors (red for the ones generated by the skilled forgeries and green for the random forgeries), they are part of the same class. This separation was made to support further discussions that will be held later.

Signatures that belong to the same writer are close to each other in the feature space. Hence, they will form a cluster located close to the origin in DS. The quality of a forgery can be measured by its proximity to a target signature (HOUMANI et al., 2011); this proximity should be considered in the feature space. When transposed to the DS, it is expected that, while bad quality skilled forgeries generate negative samples more distant to the origin, good quality skilled forgeries generate samples closer to the origin, and may even be

within the positive cluster.

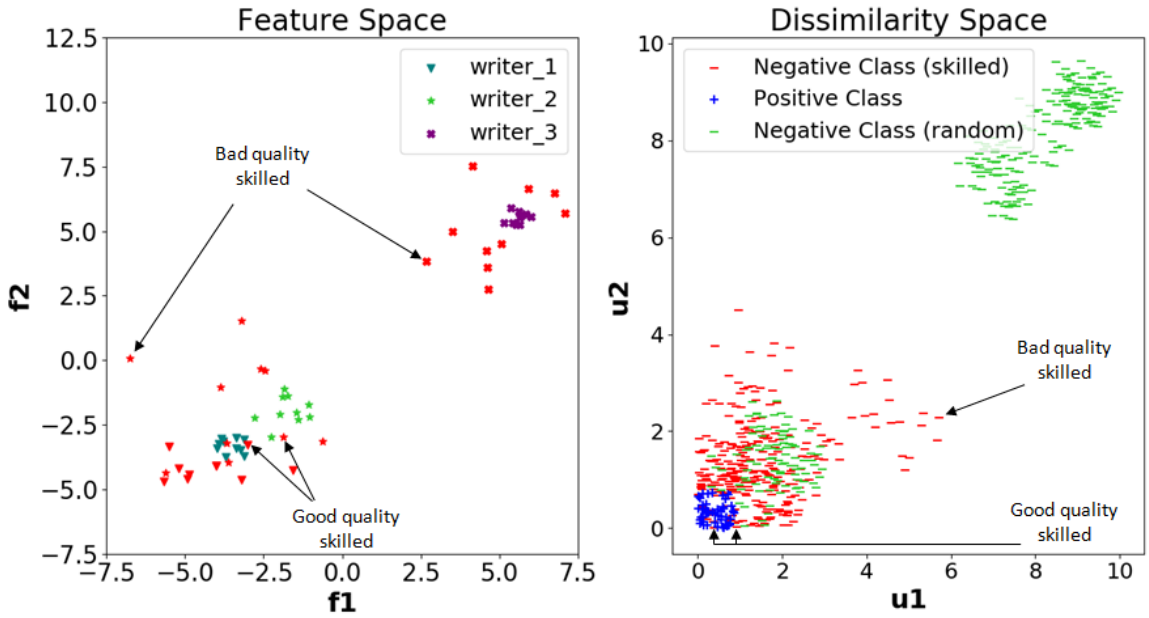

Fig. 4 – On the left, the feature space containing genuine signatures and skilled forgeries from 3 different writers (the skilled forgeries for each writer are presented in red with the same marker). On the right, the dissimilarity space generated after applying the dichotomy transformation.

3.1.1 Lessons learned

In the handwritten signature verification problem, the WI framework based on the dichotomy transformation (DT) is scalable, adaptable and presents the benefit of being able to handle some of the challenges faced when dealing with the HSV problem. Among them, (C_1) the high number of writers (classes), (C_3) the small number of training samples per writer with high intra-class variability and (C_4) the heavily imbalanced class distributions.

Another advantage of the WI framework is that it can easily manage new incoming writers (C_6), and may even be used in a transfer learning context since the different datasets would represent samples that belong to the same domain (signature representations in the dissimilarity space). However with different acquisition protocol (scanner, writing space, writing tool etc). Therefore, a single model already trained can be used to verify the signatures of new incoming writers without any further transfer adaptation.

This analysis already answers the research question 1 presented in Chapter 1.

3.2 SYSTEM OVERVIEW

Figure 5 depicts a block diagram containing the overview of the proposed approach. The top part of Figure 5 contains the training phase. The first step is to obtain the feature

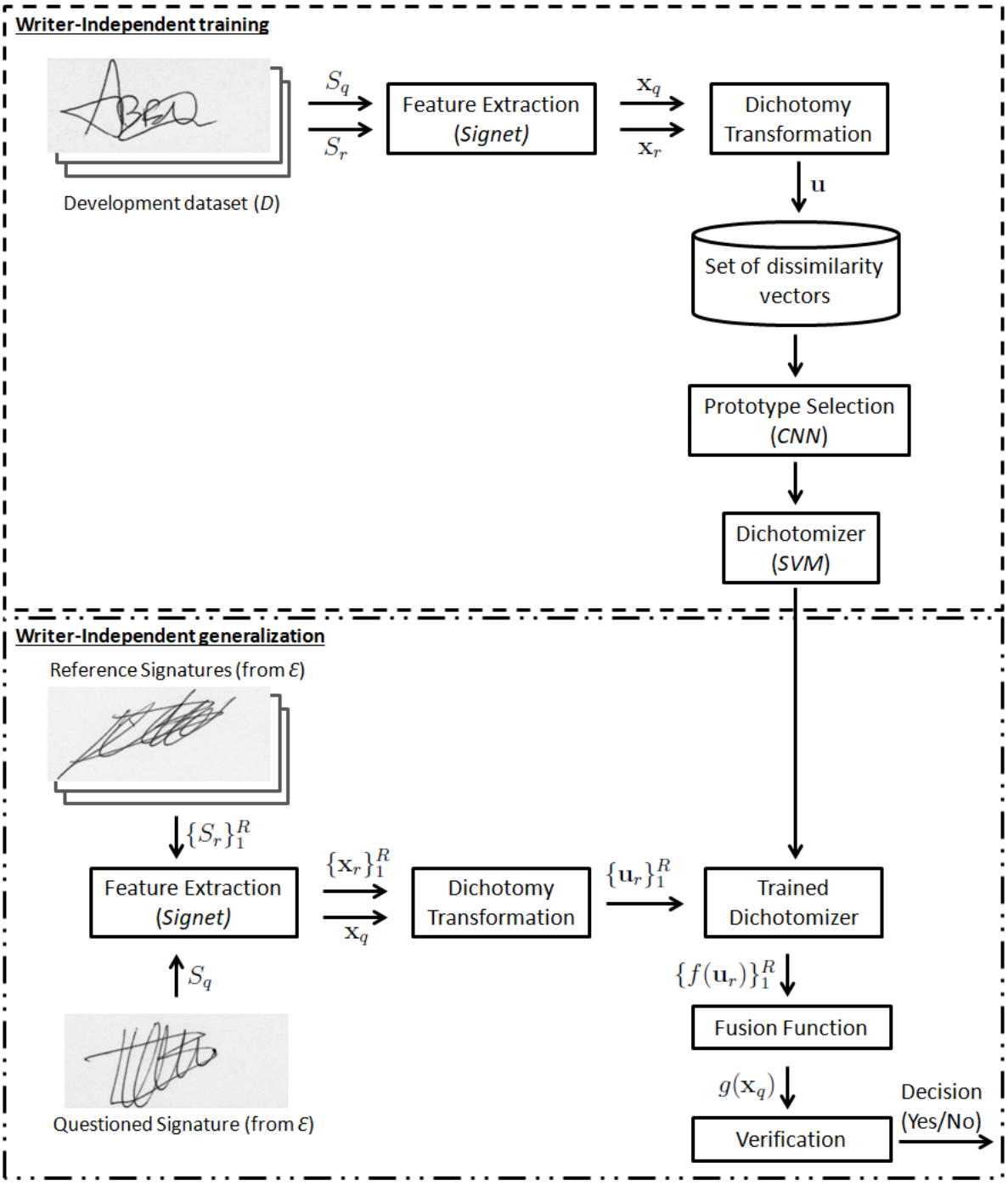

Fig. 5 – Block diagram containing the overview of the proposed approach.

vectors $\mathbf{x}_q$ and $\mathbf{x}_r$ extracted, respectively, from the images of the questioned signature S_q and the reference signature S_r belonging to the writers of the development dataset (D). This feature extraction is performed by using *SigNet* (model is available online[1]). Next, the dichotomy transformation is applied to obtain the dissimilarity vector $\mathbf{u}$. After obtaining the set of dissimilarity vectors for all considered signatures of D, the Condensed Nearest Neighbors (CNN) is applied to discard redundant samples and then the dichotomizer is

[1] <https://github.com/luizgh/sigver_wiwd>

trained with the selected samples. The SVM with RBF kernel was chosen as dichotomizer because it is considered one of the best classification methods for both WD and WI signature verification tasks (HAFEMANN; SABOURIN; OLIVEIRA, 2017b).

The generalization phase of the proposed approach is presented in the bottom part of Figure 5. Again, the first step is to extract the feature vectors through *SigNet*. One difference from the training phase is that, here, a set of reference signatures $\{S_r\}_1^R$ is considered for each questioned signature S_q, $\{S_r\}_1^R$ and S_q are obtained from the exploitation dataset (ε). Consequently a set of reference signatures, $\{\mathbf{x}_r\}_1^R$, is considered in the dichotomy transformation. In this way, DT is applied considering the feature vector of the writer's questioned signature $\mathbf{x}_q$ and the features vector set of his/her reference signatures $\{\mathbf{x}_r\}_1^R$ and produces the set of dissimilarity vectors $\{\mathbf{u}_r\}_1^R$. Next, the dichotomizer evaluates each dissimilarity vector individually and outputs a set of partial decisions $\{f(\mathbf{u}_r)\}_1^R$. The final decision of the approach about the questioned signature is based on the fusion of all partial decisions by a function $g(\mathbf{x}_q)$.

Our approach is centered on the dichotomy transformation. Thus, it presents the advantages and suffer the same weaknesses as this transformation (which were discussed and analysed in section 3.1).

3.2.1 Feature Representation

It is important to mention that DT has already been used in the handwritten signature verification scenario (RIVARD; GRANGER; SABOURIN, 2013; ESKANDER; SABOURIN; GRANGER, 2013), but using older feature representations. An important aspect of this transformation is the need for a good feature representation, as the one used in this paper. The motivation for this statement is as follows: (i) signatures that are close in the feature space will be close to the origin in the dissimilarity space. This behavior is expected for genuine signatures. (ii) the further away two signatures are in the feature space, the farther the vector resulting from the dichotomy transformation will be from the origin. It is expected to find this second behavior for the forgeries (CHA; SRIHARI, 2000). To complete the reasoning, as depicted in Figure 16, this scenario can actually be found in the feature space from *SigNet* (HAFEMANN; SABOURIN; OLIVEIRA, 2017a), as different writers are clustered in separate regions. This feature representation is discussed in section 2.2.

Another aspect is that, regardless of the signature image, *Signet* will generate feature vectors containing 2048 dimensions. This fact, facilitates the use of this feature representation in a context of transfer learning.

3.2.2 Fusion function and the number of references

SVM with RBF kernel was chosen as dichotomizer because it is considered one of the best classification methods for both WD and WI signature verification tasks (HAFEMANN;

SABOURIN; OLIVEIRA, 2017b). In the experiments we show the needing for a strong discriminant classifier that can model complex distributions. That's why the SVM with RBF kernel is a good choice.

The signed distance of the samples to the classifier's hyperplane are used as classifiers output (HAFEMANN; SABOURIN; OLIVEIRA, 2017a). So, in the experiments we analyse which partial decisions fusion functions is the best (functions MAX, MEAN, MEDIAN and MIN are tested). We also investigate the influence of the number of signatures used in the reference set.

3.2.3 Prototype Selection

Since each sample generated by the DT is formed by the distance of each pair of signatures, this approach is able to increase the number of samples in the WI-HSV scenario. However, many of these samples are redundant and have little influence during the training of the verification model. Thus, the use of prototype selection (PS) techniques in the dissimilarity space may allow the reduction of the complexity and the training time of the used classifier without degrade its results.

Considering the dissimilarity space behavior, if two samples are far in the feature space the resulted dissimilarity vector will also be far from the origin after applying the dichotomy transformation. The more distant from origin the minor the influence of the sample during the training of the WI-classifier, since the sample will be far from the border region in the dissimilarity space. In this scenario, a Condensation Approach should be used to retain the samples which are closer to the decision boundaries.

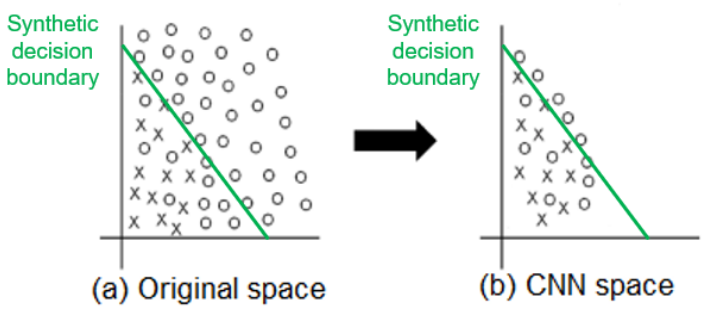

Fig. 6 – The Condensed Nearest Neighbors (CNN) method maintains the samples located in the decision boundaries.

In this work, the classical Condensed Nearest Neighbors is used as systematic prototypes selection method. This approach maintains the instances that are misclassified by a 1-NN classifier (1-nearest neighbor classifier), discarding them otherwise (HART, 1968). Its goal is to reduce the dataset size by removing redundant instances, maintaining the samples in the decision boundaries (GARCIA et al., 2012). This behavior is depicted in

Figure 6. In the left the original space, on the right, the samples kept after CNN being applied.

3.2.4 Transfer Learning

As the verification only depends on the input reference signature, by using the DT in a writer-independent approach, the dichotomizer can verify signatures of writers for whom the classifier was not trained (transfer learning). The used feature representation already keeps the same number of features for all the considered datasets. Thus, in the transfer learning scenarios, the same normalization from the training set is used for the other datasets (so the data is on the same scale). No further transfer adaptation is needed.

3.2.5 Instance Hardness

We propose to use an analysis based on the Instance Hardness of the samples in the dissimilarity space to obtain a better understanding of the space. The kDisagreeing Neighbors (kDN) measure is used herein to estimate IH. It represents the percentage of the K nearest neighbors that do not share the label of a target instance. This metric was chosen because it is able to capture the occurrence of class overlap and is also correlated with the frequency of a given instance being misclassified (SMITH; MARTINEZ; GIRAUD-CARRIER, 2014). In a more formal definition, the kDN measure, $kDN(x_q)$, of a query instance x_q, whose K nearest neighbors are denoted by $KNN(x_q)$, is defined as:

$$kDN(x_q) = \frac{|x_k : x_k \in KNN(x_q) \wedge label(x_k) \neq label(x_q)|}{K} \tag{3.2}$$

where x_k represents a neighborhood instance and, $label(x_q)$ and $label(x_k)$ represent the class labels of the instances x_q and x_k respectively (SMITH; MARTINEZ; GIRAUD-CARRIER, 2014).

We also use the IH to characterize the good and the bad quality when considering the skilled forgeries. In the experiments we show that class overlap may not occur when considering only genuine signatures and random forgeries. For this reason, this characterization of good and bad quality was only conducted based on the skilled forgeries. Figure 7 depicts examples of good and bad skilled forgeries at the image level, for the MCYT dataset. On the left, the genuine signature used as a reference is shown; the skilled forgeries are shown on the right. Forgery index represents its index (recall, for the MCYT dataset, each writer has 15 skilled forgeries). It is expected that good quality skilled forgeries be more similar to the genuine signature than the bad ones. We are using the K-Nearest Neighbors (KNN) limit to characterize the "bad" quality skilled forgeries ($IH <= 0.5$) and the "good" quality skilled forgeries ($IH > 0.5$).

As previously presented, it is expected that the negative samples from the good skilled forgeries be close to the DS origin (as depicted in figure 4). Therefore, these negative

samples may have more neighbors belonging to the positive class, i.e., higher IH values. On the other hand, as the negative samples from bad skilled forgeries are more distant to the origin, they may have more neighbors belonging to the negative class, i.e., lower IH values. These aspects can also be seen in Figure 7. Further discussions about these aspects are done in Section 4.4 and in Appendix A.

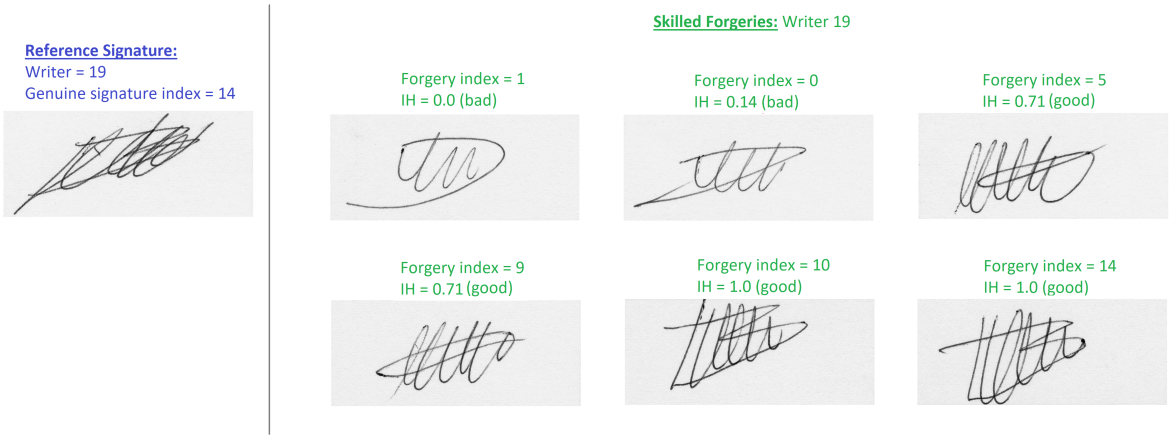

Fig. 7 – Good and bad skilled forgeries at image level

3.3 SYSTEM OVERVIEW WITH FEATURE SELECTION

One more step can be added to complement our initial approach, namely, the application of feature selection in the dissimilarity space (as depicted in Figure 8).

The option of using a Binary Particle Swarm Optimization (BPSO) algorithm for feature selection in this work comes from the good results it has obtained when compared to other optimization algorithms used for this task (CHUANG; TSAI; YANG, 2011).

In this binary swarm optimization scenario, we propose to use a BPSO-based feature selection for WI handwritten signature verification in a wrapper mode. Wrapper methods consider the selection as a search problem, where different combinations are prepared, evaluated and compared. Then, a predictive model is used to evaluate a combination of features and assign a score based on model accuracy (RADTKE; WONG; SABOURIN, 2006).

To decrease the chance of overfitting, we propose to use a global validation strategy, where the validation of the candidate solutions is executed in all iterations of the optimization process and an external archive is responsible to store the best validated solutions (RADTKE; WONG; SABOURIN, 2006).

3.3.1 Variation on BPSO

Research has shown that the three parameters w, c_1 and c_2 have a significant impact on the algorithm performance (HASSAN et al., 2005). In a variation of PSO, the Improved Self-Adaptive Particle Swarm Optimization Algorithm (IDPSO) (ZHANG; XIONG; ZHANG,

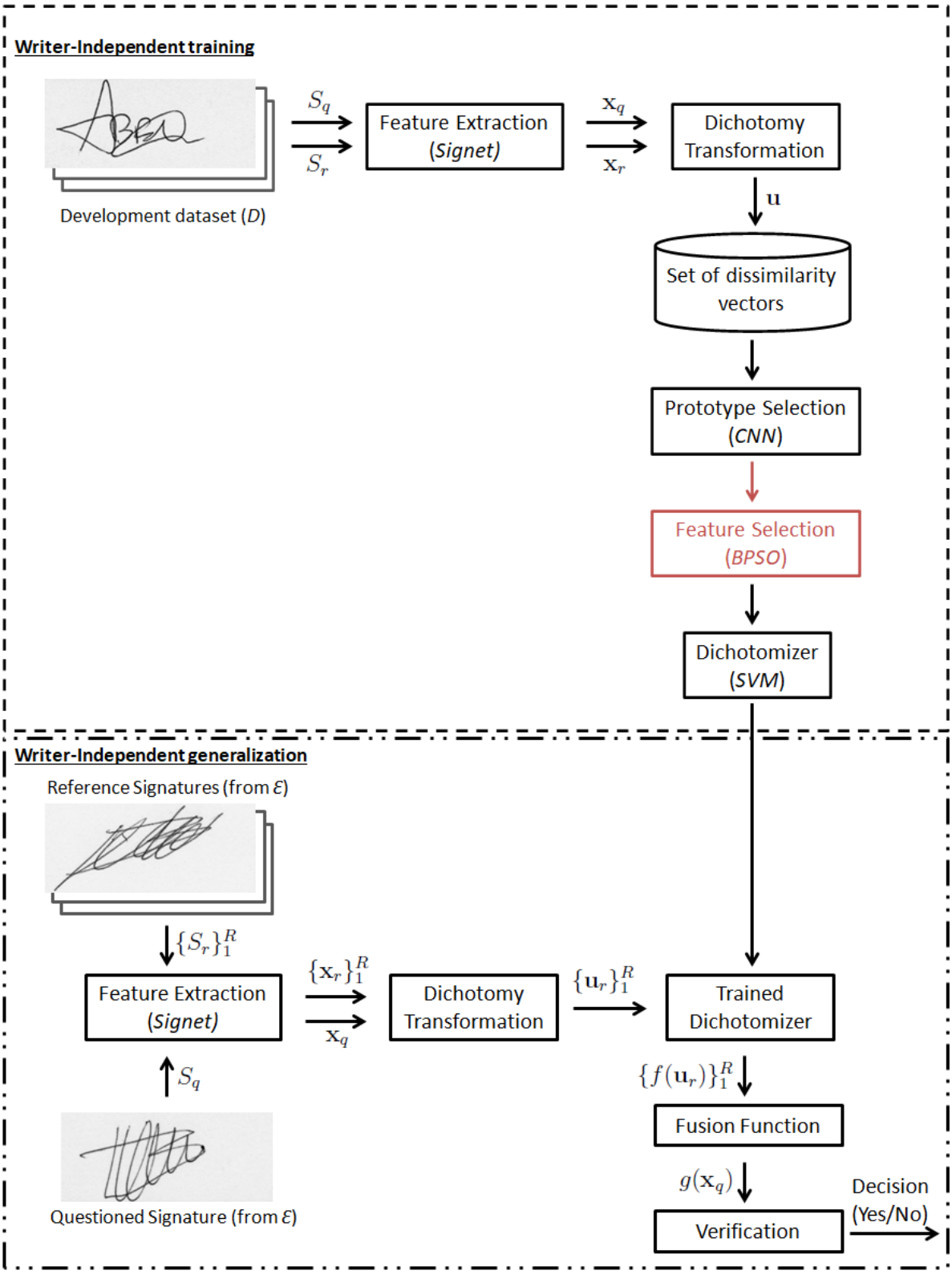

Fig. 8 – Block diagram containing the overview of the proposed approach **with feature selection**.

2013), the algorithm itself adjusts w, c_1 and c_2 dynamically over iterations, promoting global search in the beginning and local search in the final iterations. The dynamic ad-

justment of these parameters occurs through a detection function $\varphi(t)$, which is defined by Eq. 3.3:

$$\varphi(t) = |(\mathbf{gBest} - \mathbf{p}_i(t-1))/(\mathbf{pBest}_i - \mathbf{p}_i(t-1))| \tag{3.3}$$

where $|(\mathbf{gBest} - \mathbf{p}_i(t-1))|$ represents the Euclidean distance between the best position found by the swarm and the previous position found by the i-th particle. $|(\mathbf{pBest}_i - \mathbf{p}_i(t-1))|$ represents the Euclidean distance between the best position found by the i-th particle and its previous position.

Once the detection function is computed, the values of c_1 and c_2 are dynamically updated (respectively, Eqs. 3.4 and 3.5). The inertia is computed based on $\varphi(t)$ and on a sigmoid function through Eq. 3.6 (ZHANG; XIONG; ZHANG, 2013):

$$c_1 = c_1 \cdot \varphi(t)^{-1} \tag{3.4}$$

$$c_2 = c_2 \cdot \varphi(t) \tag{3.5}$$

$$w(t) = \frac{w_{inicial} - w_{final}}{1 + e^{\varphi(t)\cdot(t-((1+ln(\varphi(t)))\cdot k_{max})/\mu)}} + w_{final} \tag{3.6}$$

where $w_{inicial}$ and w_{final} respectively present the initial and the final values of the inertia w (values in the range $0 < w < 2$). K_{max} is the maximum number of iterations used in the algorithm, t is the current iteration of the algorithm, $\varphi(t)$ is the detection function and μ is an adjustment factor (ZHANG; XIONG; ZHANG, 2013).

By using this dynamic update, the IDPSO presents the following behavior:

- In initial iterations: $\varphi(t) \geq 1$ and so the value of c_1 is reduced and value of c_2 is increased. With this, the algorithm improves the exchange of information and cooperation of the particles, emphasizing the global search in the whole space.
- In final iterations: $\varphi(t) < 1$ and so so the value of c_1 is increased and value of c_2 is reduced. Thus, the algorithm improves the influence of the particle itself, emphasizing local search capability (refinement of the solution found)

Thus, IDPSO displays all PSO advantages and also improves both the ability of global search in initial iterations and local search in final iterations. Another advantage is that IDPSO is not dependent on the parameters w, c_1 and c_2 (ZHANG; XIONG; ZHANG, 2013). Algorithm 1 presents IDPSO Pseudocode.

3.3.2 Fitness function

The wrapper-based optimization is conducted by minimizing of the Equal Error Rate (EER) of the SVM. The EER metric is the error obtained when False Rejection Rate

Algorithm 1: IDPSO Pseudocode.

```
1  begin
2      Generate an initial population of particles (swarm): S
3      Randomly initialize the initial position (p) and velocity (v) of each particle i that belongs
         to the swarm S.
4      for each particle i of S do
5          Compute fitness f_i through the chosen fitness function
6          Compute the best particle position i so far: pBest_i
7      end
8      Select the particle with best fitness of the swarm: gBest
9      for each particle i of S do
10         Compute detection function φ(t). Eq. 3.3
11         Update inertia. Eq. 3.6
12         Update the variables c_1 and c_2. Eq. 3.4 and Eq. 3.5
13         Update particle velocity: Eq. 2.1
14         Update particle position: Eq. 2.2
15     end
16     if stopping criteria is not reached then
17         Return to line 4.
18 end
```

(FRR) is equal to False Acceptance Rate (FAR) (SOUZA; OLIVEIRA; SABOURIN, 2018). The user threshold (considering just the genuine signatures and the skilled forgeries) was employed (SOUZA; OLIVEIRA; SABOURIN, 2018). As mentioned, the motivation for using Support Vector Machines (SVM) as the classifier is because it is one of the most effective classifiers for both writer-dependent (WD) and writer-independent (WI) signature verification tasks (HAFEMANN; SABOURIN; OLIVEIRA, 2017b).

3.3.3 Overfitting control

In the feature selection scenario, overfitting occurs when the optimized feature set memorizes the training set instead of producing a general model. Hence, it may fail to generalize well to unseen data. In the wrapper-based approach, the swarm optimization process becomes another learning process and may be subject to overfitting. To decrease the chance of overfitting, a validation procedure can be used during the optimization process in order to select solutions with good generalization power.

According to Santos et al. (SANTOS; SABOURIN; MAUPIN, 2009), one possible validation strategy is to validate final candidate solutions on another set of unknown observations – the selection set. By using this approach, the optimization routine produces better results than selecting solutions based solely on the accuracy of the optimization set alone. However, this strategy has the disadvantage that the solution is validated only once, after the optimization process is completed.

Another approach is the global validation strategy (RADTKE; WONG; SABOURIN, 2006), where the validation of the candidate solutions are executed in all iterations of the optimization process. This can be accomplished by storing the best validated solutions in an

external (auxiliary) archive.

The studies by Radtke et al. (RADTKE; WONG; SABOURIN, 2006) and Santos et al. (SANTOS; SABOURIN; MAUPIN, 2009) formulated the problem of classifier ensemble selection as an optimization problem and applied these strategies to control the overfitting. In both scenarios, the global validation strategy was able to detect overfitting and outperformed the other overfitting control methods. So, in this work, we use this approach since it has shown better results in the literature.

Algorithm 2: Pseudocode of the global validation strategy to control overfitting.

Result: External archive A
Creates initial population $P(1)$ with N individuals
Replaces optimization set by the selection set for objective function evaluation
Calculate objective functions for all solutions in $P(t)$
$A = \emptyset$
$t = 1$
while $t <$ *maximum iterations* **do**
    Evolve $P(t)$ to $P(t+1)$
    Validate $P(t+1)$ with the selection set
    Update the external archive A with the individuals from $P(t+1)$ based on their fitness from the validation process
    t=t+1
end

Algorithm 2 presents the pseudo-code for the global validation strategy. As can be seen, an empty external archive S is created at the beginning and updated at each iteration according to the validated solutions. During this routine, the Optimization set is temporarily replaced by the Selection set to evaluate the fitness function. At each iteration, all the best solutions previously found are grouped with the population of the new swarm and then ranked. Finally, the external archive maintains the N best candidate solutions.

Figure 9 depicts the global validation strategy overview. The Optimization set is used to guide the search during the iterations of the BPSO. In turns, the Selection set is used in the validation stage for any of the methods used to control the overfitting.

3.3.4 Limitations of the proposed approach

The error in the experiments is computed considering an user threshold. So, although the proposed model being writer-independent, the decision is carried out in at the user's context, not using a global threshold to compute the error.

The user threshold was used to compare the results with those found in the literature under the same conditions.

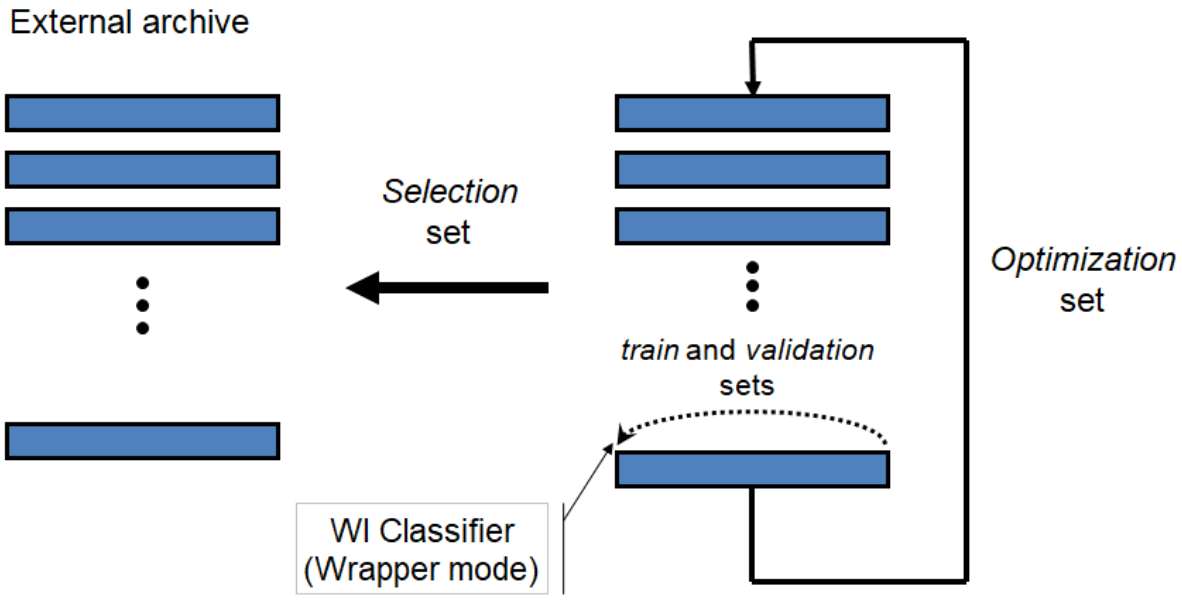

Fig. 9 – Global validation strategy overview.

4 EXPERIMENTS

In this chapter we describe the experiments performed in order to evaluate the proposed method discussed in Chapter 3. The dataset used in the experiments are described and the segmentation into training, validation and test datasets is explained. Also, the experimental setup is discussed and we present and analyze the main results obtained in these experiments and how they answer the research questions formulated in Chapter 1.

4.1 DATASETS

The experiments are carried out using the GPDS, BRAZILIAN, MCYT and CEDAR datasets, which are summarized in Table 2.

Table 5 – Summary of the used datasets.

Dataset Name	Users	Genuine signatures (per user)	Forgeries per user
BRAZILIAN	60 + 108	40	10 simple, 10 skilled
CEDAR	55	24	24
MCYT	75	15	15
GPDS-300	881	24	30

To enable comparison with other works, the GPDS-300 segmentation was used. In this case, the first 300 writers from the GPDS-300 dataset form the exploitation set ε and the development set D is composed by the remaining 581 writers (HAFEMANN; SABOURIN; OLIVEIRA, 2017a). It is worth noting that these subsets are disjoint, hence both of them are composed of different writers.

The training set is generated by using a subset of 14 genuine signatures for each writer from D (RIVARD; GRANGER; SABOURIN, 2013; ESKANDER; SABOURIN; GRANGER, 2013). Samples belonging to the positive class are generated by applying the DT on the genuine signatures from every writer in D, as in Table 6. To have an equivalent number of counterexamples, the negative samples are generated by using 13 genuine signatures (references signatures) against one selected from a genuine signature of 7 different writers (7 random forgeries), as in Table 6. Thus, the same number of samples for both positive and negative classes are generated to be part of the training set.

Table 6 – GPDS-300 dataset: Development set D

Training set (14 signatures per writer)	
Negative Class	Positive Class
Pairwise comparisons among 13 signatures per writer and 7 random signatures of other writers	Pairwise comparisons among the 14 signatures per writer
$581 \cdot 13 \cdot 7 = 52,871$ negative samples	$581 \cdot 14 \cdot 13/2 = 52,871$ positive samples

In this study, the IH analyses are performed considering the neighborhood of the GPDS-300 training set (after applying CNN preprocessing). So, to compute the IH value,

each test sample is considered alone with the whole training set. Thus, we can observe the behavior of the test samples from different datasets in relation to the same training set neighborhood and, consequently, obtain a better understanding in a transfer learning scenario. The motivation for using the GPDS-300 base training set is as follows: (i) GPDS-300 has the largest training set, and (ii) as explained in section 2.2, the features of the other datasets are obtained from a Deep Convolutional Neural Networks (DCNN) trained using the GPDS-300 dataset.

The same methodology was used for the BRAZILIAN dataset, and the division is summarized in Table 7. For CEDAR and MCYT datasets, we used a 5x2 fold cross-validation (i.e., 5 runs using 2-fold CV each). Hence, as the MCYT dataset has 75 writers, each fold would have 37 or 38 writers. For the training folds, from the 15 genuine signatures of each writer in D, 10 signatures are randomly selected to generate the training set (Table 9). For the CEDAR dataset, the 55 writers were split into 27 or 28 writers per fold. For the training folds, the 24 genuine signatures of each writer in D, 14 signatures are randomly selected to generate the training set (Table 8). The other fold is used for testing in both scenarios.

In its turn, in the transfer learning scenario, the whole set of writers are used to obtain the development sets, but we keep the number of genuine signatures and random forgeries.

Table 7 – BRAZILIAN dataset: Development set D

Training set	
Positive Class	Negative Class
Distances between the 30 signatures for each writer (D)	Distances between the 29 signatures for each writer and 15 random signatures from other writers
$108 \cdot 30 \cdot 29/2 = 46,980$ samples	$108 \cdot 29 \cdot 15 = 46,980$ samples

Table 8 – CEDAR dataset: Development set D

Training set	
Positive Class	Negative Class
Distances between the 14 signatures for each writer (D)	Distances between the 13 signatures for each writer and 7 random signatures from other writers
(27 *or* 28) $\cdot$ 14 $\cdot$ 13/2 samples	(27 *or* 28) $\cdot$ 13 $\cdot$ 7 samples

Table 9 – MCYT dataset: Development set D

Training set	
Positive Class	Negative Class
Distances between the 10 signatures for each writer (D)	Distances between the 9 signatures for each writer and 5 random signatures from other writers
(37 *or* 38) $\cdot$ 10 $\cdot$ 9/2 samples	(37 *or* 38) $\cdot$ 9 $\cdot$ 5 samples

Considering that each dataset has a different number of writers and signature per writers and to be able to compare the results with the state-of-the-art the testing set

is acquired using a methodology similar to that described in (HAFEMANN; SABOURIN; OLIVEIRA, 2017a). The BRAZILIAN dataset contains simple and skilled forgeries for the first 60 writers. Thus only these writers were used. For the MCYT and the CEDAR datasets all the writers were used. Table 10 summarizes the used Exploitation set ε for each dataset.

Table 10 – Exploitation set ε

Dataset	#Samples	#questioned signatures (per writer)
BRAZILIAN	2400	10 genuine, 10 random, 10 simple, 10 skilled
CEDAR	1650	10 genuine, 10 skilled, 10 random
MCYT	2250	5 genuine, 15 skilled, 10 random
GPDS-300	9000	10 genuine, 10 skilled, 10 random

4.1.1 General experimental setup

For all sections of the experiments, as first step, the distance vectors $\mathbf{u}$ (in the dissimilarity space) are standardized (zero mean and unit variance). In the transfer learning scenarios, the same normalization from the training set is used for the other datasets (so the data is on the same scale).

In this study, the SVM is used as writer-independent classifier with the following settings: RBF kernel, $\gamma = 2^{-11}$ and $C = 1.0$ (C and γ were selected based on a grid search: $C_{grid} = \{0.0001, 0.001, 0.01, 0.1, 1, 10, 100\}$ and $\gamma_{grid} = \{2^{-11}, 0.0001, 0.001, 0.01, 0.1, 1, 10, 100\}$). The signed distance of the samples to the classifier's hyperplane are used as classifiers output (HAFEMANN; SABOURIN; OLIVEIRA, 2017a).

The Equal Error Rate (EER) metric, using user thresholds (considering just the genuine signatures and the skilled forgeries) was used in the evaluation of the verification models (HAFEMANN; SABOURIN; OLIVEIRA, 2017a). EER is the error obtained when $FRR = FAR$, where (i) FRR (False Rejection Rate), represents the percentage of genuine signatures that are rejected by the system, and (ii) FAR (False Acceptance Rate), represents the percentage of forgeries that are accepted (HAFEMANN; SABOURIN; OLIVEIRA, 2017a).

All data were randomly selected, and a different SVM was trained for each replication (ten replications were performed for each experimental configuration). To evaluate the effectiveness of the results, we conducted the Wilcoxon paired signed-rank test with a 5% level of significance to confirm whether the two methods were significantly different in terms of EER.

4.2 DEEP CONVOLUTIONAL NEURAL NETWORK (DCNN) FEATURES FOR WI HSV

The main objective of this section is to investigate whether *SigNet* (HAFEMANN; SABOURIN; OLIVEIRA, 2017a) can also lead to good results in a writer-independent HSV

context. To this end, it is proposed the use of the dichotomy transformation (RIVARD; GRANGER; SABOURIN, 2013) combined with an SVM as a writer-independent classifier to perform the signature verification.

The results are organized as follows: (i) initially, the number of signatures in the reference set R is fixed and an analysis of which fusion function is the best is carried out (functions max, mean, median and min are tested), (ii) In the sequence, the analysis about the influence of the number of signatures used in the reference set R is presented. (iii) Finally, the comparison with the state-of-the-art for the used datasets is discussed (as the present work uses *SigNet* features, Hafemann et al. results are also presented using only these features (HAFEMANN; SABOURIN; OLIVEIRA, 2017a)).

In this section we answer the research questions 2 and 3 presented in Chapter 1.

4.2.1 Detailed experimental setup

In the experiments reported in this Section, we use the Equal Error Rate (EER) metric, using both global and user thresholds (considering just the genuine signatures and the skilled forgeries) for the evaluation of the verification models (HAFEMANN; SABOURIN; OLIVEIRA, 2017a).

Some analyses are also carried out considering the Average Error Rate (AER) metric, which is the average error considering FRR, FAR_{random}, FAR_{simple}, $FAR_{skilled}$. The $AER_{genuine+skilled}$ metric, which is the average error considering just FRR and $FAR_{skilled}$, is also used.

4.2.2 Results and discussion

4.2.2.1 BRAZILIAN dataset: Fusion function analysis

To measure the impact of the fusion function, in this section the number of references per writer is fixed in 30 (highest number of references). The tested fusion functions are: (i) mean, (ii) max, (iii) median and (iv) min.

Figure 10 depicts the boxplots for AER and $AER_{genuine+skilled}$ for the tested functions (max, mean, median, min). As can be seen, the max function obtained the best results for both metrics (this settings will be referenced as SVM_{max}). In the opposite direction, the min function had the worst results in both cases.

The Wilcoxon paired signed-rank test with 5% significance level for both AER and $AER_{genuine+skilled}$ metrics shows that the max function outperforms the other functions with statistical relevance.

For the EER metric, the Wilcoxon paired signed-rank test also showed that the max function is statistically better when compared to the median and the min. However, there is no statistical difference to the mean function (this behavior can be observed in Figure 11).

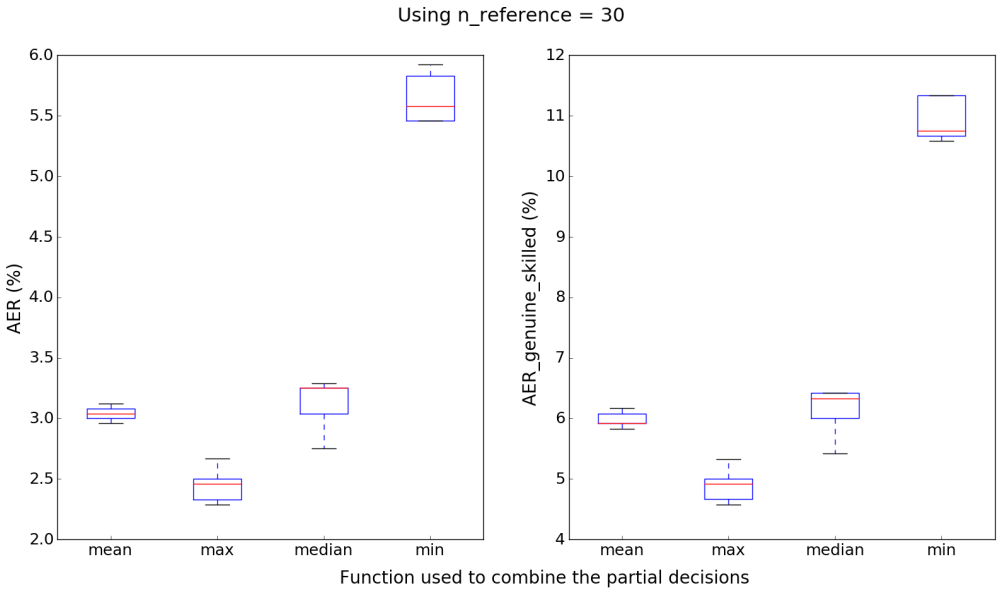

Fig. 10 – Boxplots for AER (left) and $AER_{genuine+skilled}$ (right) metrics on the BRAZILIAN dataset, using 30 references per writer.

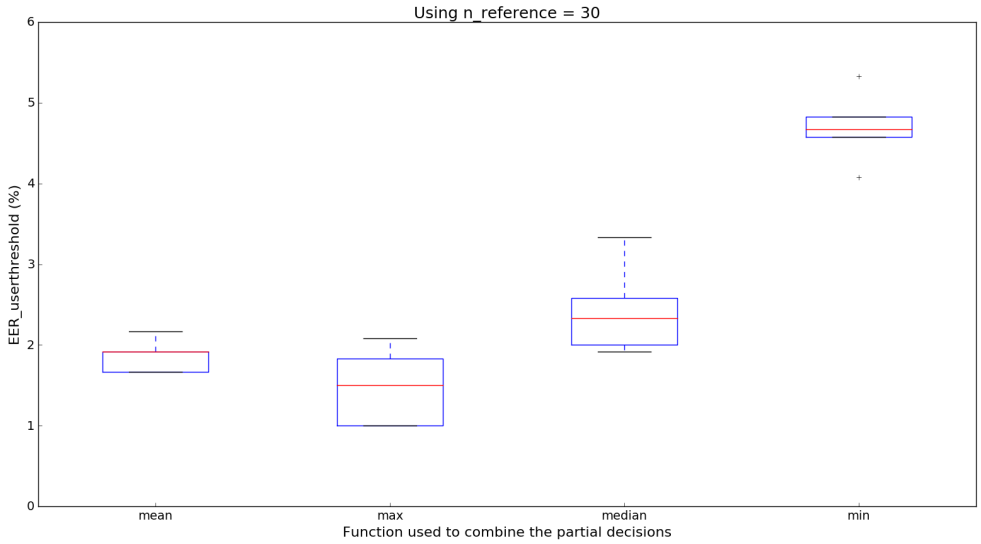

Fig. 11 – Boxplots for EER (user threshold) metric on the BRAZILIAN dataset, using 30 references per writer.

4.2.2.2 BRAZILIAN dataset: Number of reference signatures analysis

In the previous section it was shown that the max obtained better results when compared to other functions. In this section, to measure the impact of the reference set cardinality, the max function was fixed and the number of references per writer varies. To this end, reference subsets containing [1, 5, 10, 15, 20, 25, 30] randomly selected signatures are used as references for the verification task.

Figure 12 depicts the boxplots for AER and $AER_{genuine+skilled}$ for different number

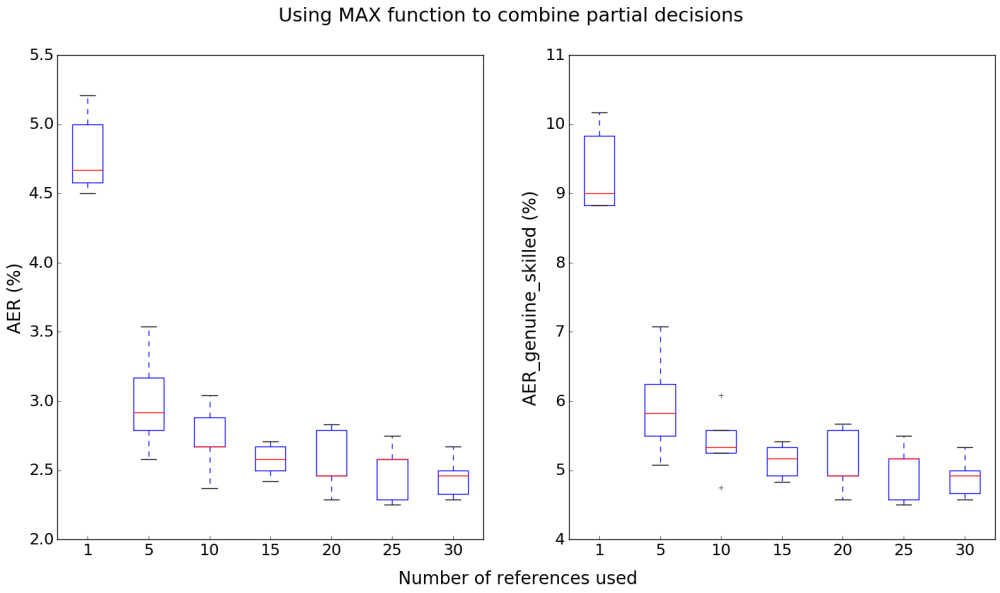

Fig. 12 – Boxplots for AER (left) and $AER_{genuine+skilled}$ (right) metrics on the BRAZILIAN dataset, using max function.

of reference signatures. As can be seen, using more references per writer produces better results (the worst cases are with number of references = 1 and = 5). However, the variation among results decreases as the number of references increases. For instance, the experiments using 15 or 20 references present similar results; the same can be observed with 25 and 30.

The Wilcoxon paired signed-rank test with 5% significance level for both metrics, using n_reference = 30 as baseline, shows that results are statistically better only when compared with the cases where number of references = 1 and = 5. There is no statistical difference to n = 10, 15, 20 or 25.

4.2.2.3 BRAZILIAN dataset: Comparison with the state-of-the-art

Table 11 – Comparison of EER with the state-of-the-art on the BRAZILIAN dataset, using max function (errors in %).

Type	Reference	#references	EER
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2016)	15	4.17
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	5	2.92 (0.44)
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	15	2.07 (0.63)
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	30	2.01 (0.43)
WI	SVM_{max} (using a global threshold)	5	5.95 (0.68)
WI	SVM_{max} (using a global threshold)	15	5.13 (0.23)
WI	SVM_{max} (using a global threshold)	30	4.90 (0.27)
WI	SVM_{max} (using an user threshold)	5	2.58 (0.72)
WI	SVM_{max} (using an user threshold)	15	1.70 (0.40)
WI	SVM_{max} **(using an user threshold)**	30	**1.47 (0.36)**

As can be observed in Table 11, the results of SVM_{max} (using an user threshold)

outperforms the other models. Also, the SVM_{max} achieved better results when compared both to (HAFEMANN; SABOURIN; OLIVEIRA, 2016) and (HAFEMANN; SABOURIN; OLIVEIRA, 2017a) for the EER metric, considering models with the same number of references. It worth to notice that the proposed method performs writer-independent verification and both Hafemann's models operate in a writer-dependent way and, even so, our WI approach was able to improve the results.

4.2.2.4 GPDS-300 dataset

For the GPDS-300 datasets, the same fusion functions were tested. As the highest value for the number of references is 12, reference subsets containing [1, 2, 3, 4, 5, 10, 12] randomly selected signatures were tested as references for the verification.

As in the BRAZILIAN dataset, in the GPDS-300 dataset the best results are obtained using the max function with the highest value for the number of references (in this case, n_reference = 12) for both global and user threshold scenarios. Figures 13 and 14 depict, respectively, the boxplots for AER and $AER_{genuine+skilled}$ metrics (i) varying the fusion functions and (ii) varying the number of references, for the GPDS-300 dataset.

For the GPDS-300 dataset, performing the Wilcoxon paired signed-rank test for both metrics: (i) max function outperforms the other functions with statistical relevance. (ii) Using n_reference = 12 is statistically better when compared to the other cases.

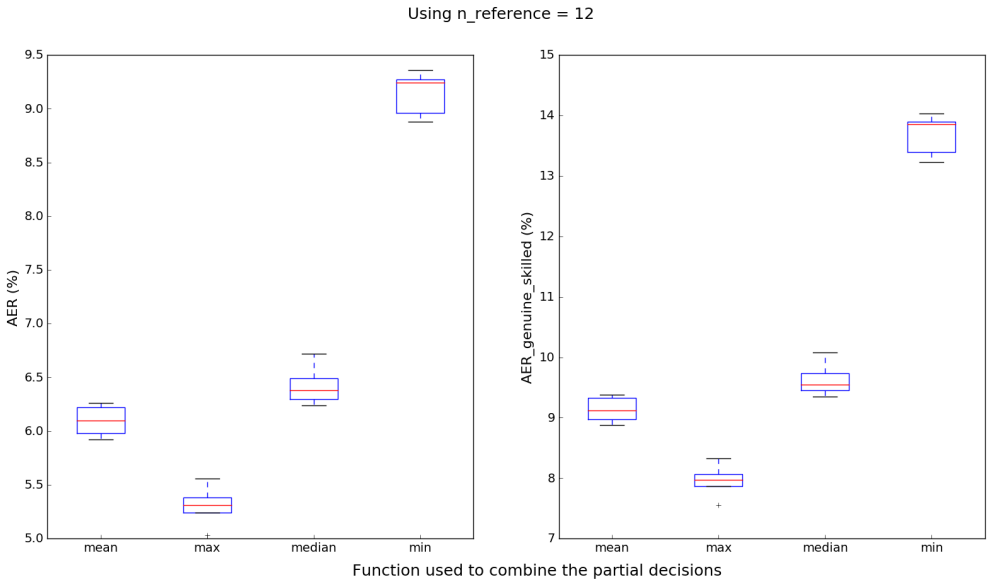

Fig. 13 – Boxplots for AER (left) and $AER_{genuine+skilled}$ (right) metrics on the GPDS-300 dataset, using n_reference = 12.

Table 12 presents the results when SVM_{max} is used and compares the obtained results with those from Table 7 of Hafemann et al. paper (HAFEMANN; SABOURIN; OLIVEIRA, 2017a), for the GPDS-300 dataset.

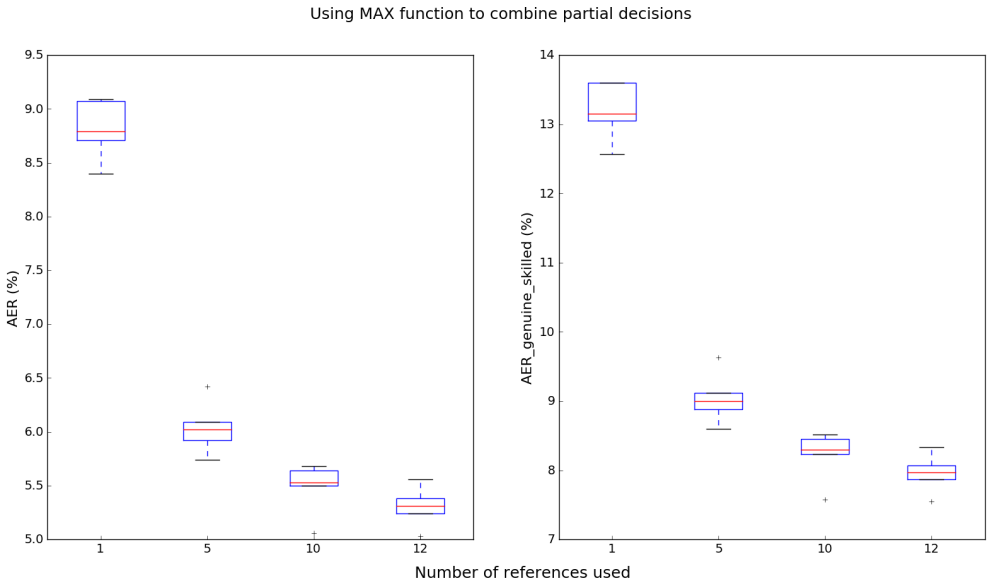

Fig. 14 – Boxplots for AER (left) and $AER_{genuine+skilled}$ (right) metrics on the GPDS-300 dataset, using max function.

Table 12 – Comparison of EER with the state-of-the-art on the GPDS-300 dataset, using Max function (errors in %).

Type	Reference	#samples	EER
WD	(SOLEIMANI; ARAABI; FOULADI, 2016)	10	20.94
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2016)	12	12.83
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	5	3.92 (0.18)
WD	**(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)**	12	**3.15 (0.18)**
WI	SVM_{max} (using a global threshold)	5	9.05 (0.34)
WI	SVM_{max} (using a global threshold)	12	7.96 (0.26)
WI	SVM_{max} (using an user threshold)	5	4.40 (0.34)
WI	SVM_{max} (using an user threshold)	12	3.69 (0.18)

As presented in Table 12, SVM_{max} was able to outperform (SOLEIMANI; ARAABI; FOULADI, 2016) and (HAFEMANN; SABOURIN; OLIVEIRA, 2016) using both global and user thresholds. However, the proposed approach using an user threshold obtained slightly inferior results in comparison with the WD model from (HAFEMANN; SABOURIN; OLIVEIRA, 2017a) for the EER metric.

4.2.2.5 Dynamic reference selection through MAX funcion

We are using the signed distance of a sample to the classifier's hyperplane as classifiers output. An important aspect related to the signed distance is that it indicates in which side of the hyperplane generated by the classifier the sample is located and how far it is from this hyperplane. Figure 15 depicts this property. Given the dissimilarity space and the blue line representing a decision hyperplane with the left side as its positive side (because it is closer to the origin), then, for each fusion function, the distance used for the final decision would be:

- MAX: the distance from the sample farthest from the hyperplane on the positive side;
- MIN: the distance from the sample farthest from the hyperplane on the negative side;
- MEAN and MEDIAN: respectively, the mean and median of all distances.

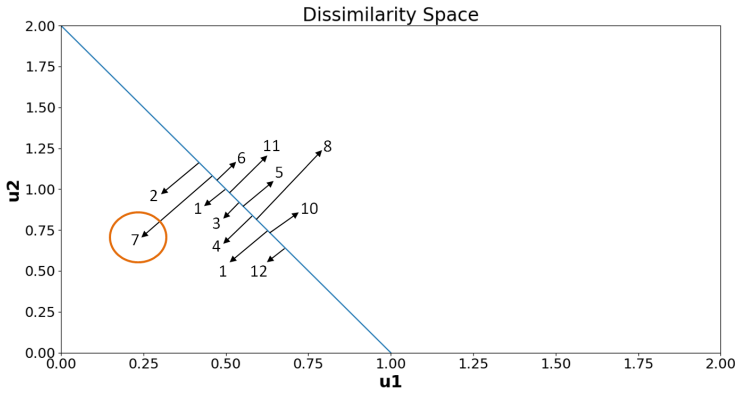

Fig. 15 – Dissimilarity space with the highlight on the selected reference, when MAX is used as a fusion function.

As we are in the dissimilarity space, the sample farthest from the hyperplane on the positive side represents the one that is closest to the DS origin, and, hence, the one generated by the DT of the reference signature and the questioned signature that are closest in the feature space. This happens when applying the MAX as fusion function and then, in the scenario of Figure 15, sample 7 would be the one used to perform the verification.

On the other hand, the sample farthest from the hyperplane on the negative side represents the one that is further away from the DS origin, i.e., the one generated by the DT of the reference signature and the questioned signature that are farther apart from each other in the feature space. This represents the scenario of MIN as fusion function and therefore, in Figure 15, sample 8 would be the one used to perform the verification. In MEAN and MEDIAN, there is no specific sample selected since the mean and median of all distances are respectively used in each case.

Thus, when we apply the MAX as fusion function, the approach dynamically selects the sample closest to the origin in the dissimilarity space. Hence, it dynamically selects the reference (from the set of references) that is most similar to the questioned signature and uses it to perform the verification.

4.2.3 Lessons learned

In this section, the results of the experiments showed that, in general, for the tested datasets, the best results are obtained using the MAX as fusion function (research question 3) with the highest number of references (research question 2).

Moreover, in the global threshold scenario, the proposed approach was able to outperform (HAFEMANN; SABOURIN; OLIVEIRA, 2017a) in the BRAZILIAN dataset. For the user threshold scenario, the proposed approach was able to obtain performance comparable to (HAFEMANN; SABOURIN; OLIVEIRA, 2017a). This was the case even with the proposed method performing writer-independent verification and (HAFEMANN; SABOURIN; OLIVEIRA, 2017a) operating in a writer-dependent way. In the BRAZILIAN dataset our proposed approach was slightly superior and in the GPDS-300 dataset slightly inferior. However, for both datasets, the proposed approach was able to outperform other methods from the literature that use WD classification and the WI dissimilarity representation with different features and more complex classification architectures (for instance, ensembles of classifiers).

We also showed that, by using MAX as fusion function, the approach dynamically selects the reference (from the set of references) that is most similar to the questioned signature and uses it to perform the verification.

4.3 PROTOTYPE SELECTION AND TRANSFER LEARNING

The objective of the experiments is to analyze whether: (i) prototype selection preprocessing techniques can be used without degrading the performance of the classifier; (ii) preprocessing based on a systematic prototype selection technique is better than a random selection for the WI-HSV problem; and (iii) WI-SVM trained in the GPDS-300 dataset can be used to verify signatures in the other datasets (in a transfer learning approach).

The aim was also to explain these objectives based on the main characteristics of the dissimilarity space resulting from the dichotomy transformation for WI-HSV. The instance hardness distribution of genuine signatures, random forgeries and skilled forgeries, was used to this end.

In this section we answer the research questions 4 and 5 presented in Chapter 1.

4.3.1 Detailed experimental setup

In the experiments reported in this section, the classical Condensed Nearest Neighbors (CNN) approach is used for systematic prototype selection. This approach maintains the instances that are misclassified by a 1-NN classifier (1-nearest neighbor classifier), discarding them otherwise (HART, 1968). CNN was chosen because its goal is to reduce the dataset size by removing redundant instances, maintaining the samples in the decision boundaries (GARCIA et al., 2012).

As presented in Section 4.2, for the tested datasets, the best results are generally obtained using the highest number of references and max as the fusion function. Therefore, only this approach is considered in this Section.

For the instance hardness (IH) analysis, $K = 7$ is used for the estimation of the kDN (CRUZ et al., 2017).

4.3.2 Using Prototype Selection

The following experiments evaluate the application of prototype selection before training the SVM. The $\%\_SVM$ represents the models with uniform random subsampling of the training set. We use 1.0%, 5.0% and 10.0% of the original training set. The Condensed Nearest Neighbors is referred to as CNN_SVM in the tables.

4.3.2.1 GPDS-300 dataset

Table 13 presents a comparative analysis of the results obtained by the SVMs (with and without prototype selection) versus those obtained with state of the art models, considering the EER metric. Tables 14 and 15 respectively present the comparative analysis of the number of samples and the number of support vectors (SV) obtained by the SVMs (with and without prototype selection), for the GPDS-300 dataset.

Table 13 – Comparison of EER with the state of the art in the GPDS-300 dataset, using max function (errors in %)

Type	Model	#references	EER
WD	(SOLEIMANI; ARAABI; FOULADI, 2016)	10	20.94
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2016)	12	12.83
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	5	3.92 (0.18)
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	12	3.15 (0.18)
WI	SVM_{max}	12	3.69 (0.18)
WI	$1\%\_SVM_{max}$	12	3.54 (0.26)
WI	$5\%\_SVM_{max}$	12	3.62 (0.32)
WI	$10\%\_SVM_{max}$	12	3.48 (0.12)
WI	CNN_SVM_{max}	12	3.47 (0.15)

Table 14 – Comparison of the number of training samples in the GPDS-300 dataset

Model	#Positive Samples	#Negative Samples	#Retained Samples (%)
SVM	52871	52871	100.00 (0.00)
$1\%\_SVM$	531.70 (17.04)	526.30 (17.04)	1.00 (0.00)
$5\%\_SVM$	2648.10 (24.78)	2639.90 (24.78)	5.00 (0.00)
$10\%\_SVM$	5289.30 (31.69)	5285.70 (31.69)	10.00 (0.00)
CNN_SVM	345.90 (15.25)	4437.80 (125.11)	4.52 (0.13)

As presented in Tables 13 and 14, the use of the prototype selection methods allows the SVM to be trained with a much smaller number of samples, while keeping performance in terms of EER. This also results in a large reduction in the number in the support vectors

Table 15 – Comparison of the number of support vectors (SV) in the GPDS-300 dataset

Model	#SV	#Positive SV	#Negative SV
SVM	3398.40 (95.01)	1640.30 (45.90)	1758.10 (53.46)
$1\%\_SVM$	194.60 (8.92)	78.70 (4.61)	115.90 (9.87)
$5\%\_SVM$	481.30 (16.78)	208.50 (12.15)	272.80 (11.39)
$10\%\_SVM$	720.90 (23.64)	309.80 (9.35)	411.10 (16.88)
CNN_SVM	928.20 (28.44)	312.90 (12.32)	615.30 (19.34)

used by the SVM (Table 15), which in turn reduces the complexity and computational cost of training a SVM in the offline WI-HSV context.

In Table 13, a simple random subsampling with 1% of the training samples provides similar results to what is obtained with the SVM trained with the complete training set. This shows that the samples resulting from the dichotomy transformation are redundant for this dataset.

For Table 13, considering the WD model from Hafemann et al. (HAFEMANN; SABOURIN; OLIVEIRA, 2017a) for the GPDS-300 dataset, both the models, with and without preprocessing, obtained comparable results for the EER metric, even operating in a writer-independent fashion. When compared to the other models, the proposed approach obtained better results.

Given these results, we can see that for the GPDS-300 dataset, the dichotomy transformation was able to increase the number of samples in the WI-HSV scenario, and yet many of them were redundant. This therefore means that the use of prototype selection in the dissimilarity space allowed a reduction of the complexity of the classifier used without degrading its results.

4.3.2.2 BRAZILIAN dataset

Tables 16, 17 and 18 respectively present, a comparative analysis of the classification metrics, the number of samples and the number of support vectors (SV) obtained by the SVMs (with and without prototype selection) in the BRAZILIAN dataset.

Table 16 – Comparison of EER with the state of the art in the BRAZILIAN dataset, using max function (errors in %)

Type	Model	#references	EER
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2016)	15	4.17
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	5	2.92 (0.44)
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	15	2.07 (0.63)
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	30	2.01 (0.43)
WI	SVM_{max}	30	1.47 (0.36)
WI	$1\%\_SVM_{max}$	30	1.21 (0.45)
WI	$5\%\_SVM_{max}$	30	1.19 (0.42)
WI	$10\%\_SVM_{max}$	30	1.23 (0.51)
WI	CNN_SVM_{max}	30	1.26 (0.33)

Much as in the GPDS-300 scenario, Tables 16 and 17 show that a simple random

subsampling with 1% of the samples maintains similar results as those obtained with the SVM trained with the complete training set.

Once again, this demonstrates that the samples resulting from the dichotomy transformation are redundant for this database. The data from Brazilian dataset are probably more redundant when compared to what we have in the GPDS-300 dataset. This can be observed from the greater reduction secured by the CNN approach: while 4.52% of the samples are needed to represent the border region in the GPDS-300, only 1.47% are needed for the BRAZILIAN dataset.

Table 17 – Comparison of the number of training samples in the BRAZILIAN dataset

Model	#Positive Samples	#Negative Samples	#Retained Samples (%)
SVM	46980	46980	100.00 (0.00)
$1\%\_random$	474.20 (14.60)	465.80 (14.60)	1.00 (0.00)
$5\%\_random$	2336.50 (35.63)	2361.50 (35.63)	5.00 (0.00)
$10\%\_random$	4681.60 (38.21)	4714.40 (38.21)	10.00 (0.00)
CNN_SVM	379.30 (41.22)	1005.10 (33.65)	1.47 (0.07)

Table 18 – Comparison of the number of support vectors (SV) in the BRAZILIAN dataset

Model	#SV	#Positive SV	#Negative SV
SVM	3368.80 (72.96)	1627.40 (40.70)	1741.40 (41.29)
$1\%\_SVM$	259.70 (17.25)	92.70 (7.44)	167.00 (11.09)
$5\%\_SVM$	688.00 (33.79)	267.80 (13.67)	420.20 (28.61)
$10\%\_SVM$	1014.20 (43.57)	420.20 (14.23)	594.00 (31.38)
CNN_SVM	658.40 (34.84)	261.00 (18.77)	397.40 (20.93)

Still in Table 16, for this dataset, even operating in a writer-independent fashion, both the models with and without preprocessing (prototype selection) obtained better results considering the EER metric, when compared to the other WD models.

4.3.2.3 MCYT dataset

Tables 19, 20 and 21 respectively present a comparative analysis on the classification metrics, the number of samples and the number of support vectors (SV) obtained by the SVMs (with and without prototype selection) in the MCYT dataset.

As presented in Table 19, unlike with the CNN, the use of random subsampling resulted in the degradation of the performance of the classifier. Used as the prototype selection method, the Condensed Nearest Neighbors provided results comparable to those obtained with the SVM trained with all the data; additionally the CNN allowed the SVM to be trained with only about 8% of the training samples (as presented in Table 20). This also resulted in an almost 28% reduction in the number of the support vectors used by the SVM (Table 21). Unlike with random subsampling, using the Condensed Nearest Neighbors allowed more attention to be paid to border samples, which removed the need to store more instances than were necessary for an accurate generalization.

Table 19 – Comparison of EER with the state of the art in the MCYT dataset, using max function (errors in %)

Type	Model	#references	EER
WD	(GILPEREZ et al., 2008)	10	6.44
WD	(WEN et al., 2009)	5	15.02
WD	(VARGAS et al., 2011)	10	7.08
WD	(OOI et al., 2016)	10	9.87
WD	(SOLEIMANI; ARAABI; FOULADI, 2016)	10	9.86
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	10	2.87 (0.42)
WI	SVM_{max}	10	2.73 (0.20)
WI	$1\%\_SVM_{max}$	10	3.67 (0.11)
WI	$5\%\_SVM_{max}$	10	3.27 (0.26)
WI	$10\%\_SVM_{max}$	10	3.19 (0.20)
WI	CNN_SVM_{max}	10	2.99 (0.16)

Table 20 – Comparison of the number of training samples in the MCYT dataset

Model	#Positive Samples	#Negative samples	#Retained Samples (%)
SVM	1687.50 (22.50)	1687.50 (22.50)	100.00 (0.00)
$1\%\_random$	16.78 (3.29)	17.72 (3.13)	1.00 (0.00)
$5\%\_random$	83.22 (6.19)	85.78 (6.00)	5.00 (0.00)
$10\%\_random$	167.78 (9.02)	169.72 (9.51)	10.00 (0.00)
CNN_SVM	38.29 (5.19)	224.93 (21.08)	7.80 (0.75)

Table 21 – Comparison of the number of support vectors (SV) in the MCYT dataset

Model	#SV	#Positive SV	#Negative SV
SVM	567.77 (38.61)	251.29 (18.93)	316.48 (22.14)
$1\%\_SVM$	32.47 (1.66)	14.78 (2.32)	17.69 (3.09)
$5\%\_SVM$	105.63 (5.61)	42.54 (3.57)	63.09 (4.56)
$10\%\_SVM$	161.89 (9.79)	64.27 (4.42)	97.62 (7.47)
CNN_SVM	160.49 (15.66)	38.22 (5.14)	122.27 (12.53)

Also in Table 19, for the MCYT dataset, when compared to the other models, the proposed approach obtained better results for the EER metric. The only exception was for the comparison with the model proposed in Hafemann et al. (HAFEMANN; SABOURIN; OLIVEIRA, 2017a).

4.3.2.4 CEDAR dataset

Tables 22, 23 and 24 respectively present a comparative analysis on the classification metrics, the number of samples and the number of support vectors (SV) obtained by the SVMs (with and without prototype selection) in the CEDAR dataset.

In Table 22, for the CEDAR dataset while the use of random subsampling resulted in the degradation of the model, using the CNN did not affect the performance of the WI classifier. Used as the prototype selection method, the Condensed Nearest Neighbors provided results comparable to those obtained with the SVM trained with all the data; additionally the CNN allowed the SVM to be trained with only about 3% of the training samples (Table 23). This also results in an almost 18% reduction in the number of the

Table 22 – Comparison of EER with the state of the art in the CEDAR dataset, using max function (errors in %)

Type	Model	#references	EER
WI	(KUMAR et al., 2010)	1	11.81
WI	(KUMAR; SHARMA; CHANDA, 2012)	1	8.33
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	12	4.76 (0.36)
WI	SVM_{max}	12	5.78 (0.38)
WI	$1\%\_SVM_{max}$	12	7.22 (0.27)
WI	$5\%\_SVM_{max}$	12	6.45 (0.23)
WI	$10\%\_SVM_{max}$	12	6.02 (0.32)
WI	CNN_SVM_{max}	12	5.86 (0.50)

support vectors used by the SVM (Table 24).

Table 23 – Comparison of the number of training samples in the CEDAR dataset

Model	#Positive Samples	#Negative Samples	#Retained Samples (%)
SVM	2502.50 (45.50)	2502.50 (45.50)	100.00 (0.00)
$1\%\_random$	24.81 (3.48)	25.69 (3.50)	1.00 (0.00)
$5\%\_random$	124.31 (7.85)	126.19 (7.67)	5.00 (0.00)
$10\%\_random$	251.07 (12.57)	249.93 (11.72)	910.00 (0.00)
CNN_SVM	30.78 (7.43)	115.13 (18.26)	2.91 (0.49)

Table 24 – Comparison of the number of support vectors (SV) in the CEDAR dataset

Model	#SV	#Positive SV	#Negative SV
SVM	676.37 (64.57)	390.30 (35.63)	286.07 (32.22)
$1\%\_SVM$	39.46 (2.97)	14.45 (2.03)	25.01 (2.98)
$5\%\_SVM$	117.60 (10.30)	40.63 (4.33)	76.97 (7.32)
$10\%\_SVM$	181.44 (13.98)	65.09 (5.68)	116.35 (10.16)
CNN_SVM	119.75 (19.37)	30.66 (7.29)	89.09 (13.57)

Still in Table 22, for this dataset, the proposed approach obtained worse results when compared to the model proposed by Hafemann et al. (HAFEMANN; SABOURIN; OLIVEIRA, 2017a) and better results in the comparison with the others WI classifiers. However, the comparative results were obtained by a model using just one reference signature.

Given the above results for the tested datasets, the dichotomy transformation was thus able to increase the number of samples in the offline WI-HSV scenario; however, many of these samples are redundant. Using prototype selection in the dissimilarity space allowed a reduction of the complexity of the classifier used without degrading its performance. Furthermore, using a systematic PS, such as the CNN, allows more attention to be paid to border samples. Consequently, prototype selection may thus be used without degrading the performance of the WI classifier, while removing the need to store more instances than are necessary for an accurate generalization.

Unlike with the CEDAR dataset, the models with and without preprocessing for the other datasets obtained results comparable to those of the WD models for the EER metric, even operating in a writer-independent fashion.

4.3.3 Using Transfer Learning and Prototype Selection

For the BRAZILIAN, MCYT and CEDAR datasets, in addition to investigating the use of prototype selection, we also analyze if a WI-SVM trained in the GPDS-300 can be used to verify signatures from other datasets, akin to a transfer learning (PAN; YANG, 2010). The associated results are presented in Table 25.

In our scenario, no adaptation is required in the classifier or in the features. We only get the WI-SVM trained in the GPDS-300 and use it to verify signatures in the other datasets ($GPDS_{max}$ results in Table 25). On the other hand, SVM_{max} results in this table were obtained by training and testing the classifier on the same dataset. It should be recalled that: (i) all datasets have the same number of features; (ii) the features used for all datasets are based on the Convolutional Neural Network trained in the GPDS-300, and (iii) the same normalization/standardization is used, and therefore, all data are within the same interval.

Table 25 – Comparison of EER for the models with and without transfer learning on the BRAZILIAN, MCYT and CEDAR datasets, using max function (errors in %)

Dataset	Model	#references	EER
BRAZILIAN	SVM_{max}	30	1.47 (0.36)
	CNN_SVM_{max}	30	1.26 (0.33)
	$GPDS_{max}$	30	1.35 (0.40)
	CNN_GPDS_{max}	30	1.11 (0.37)
MCYT	SVM_{max}	10	2.73 (0.20)
	CNN_SVM_{max}	10	2.99 (0.16)
	$GPDS_{max}$	10	2.97 (0.20)
	CNN_GPDS_{max}	10	2.89 (0.13)
CEDAR	SVM_{max}	12	5.78 (0.38)
	CNN_SVM_{max}	12	5.86 (0.50)
	$GPDS_{max}$	12	3.42 (0.28)
	CNN_GPDS_{max}	12	3.32 (0.22)

In Table 25, for the BRAZILIAN and MCYT datasets, the WI-SVM trained in the GPDS-300 obtained results comparable to the SVMs being trained and tested on their own dataset, for the EER metric. More interesting results are presented for the CEDAR dataset, since the WI-SVM was trained in another dataset, and still obtained better results than the classifiers trained and tested on the same dataset. These results also show that using the CNN for transfer learning (CNN_GPDS_{max}) slightly improved the results versus the case with transfer learning without PS ($GPDS_{max}$).

4.3.4 Instance hardness analysis

Herein, we analyze the results obtained by using the instance hardness measure.

Hafemann et al. (HAFEMANN; SABOURIN; OLIVEIRA, 2017a) performed an analysis to examine the local structure of the learned feature space (WD), using the t-SNE algorithm in a subset of the development set of the GPDS-300 dataset, called the validation set for

verification V_v. Figure 16 herein is the same as Fig. 5 (b) in their paper (HAFEMANN; SABOURIN; OLIVEIRA, 2017a), and is going to be used here to describe our feature space (as we are using the same features).

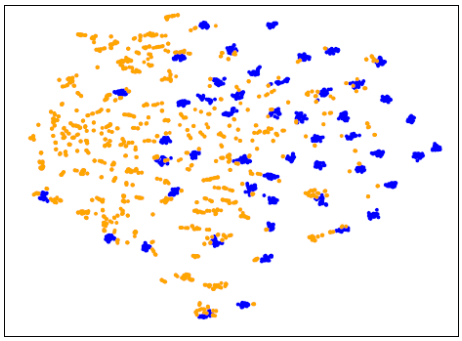

Fig. 16 – t-SNE 2D projections of the feature vectors from the 50 users in the validation set for verification V_v. The blue points represent genuine signatures and the orange ones represent skilled forgeries

As can be seen in Figure 16, (i) genuine signatures from different users are clustered and occupy different regions of the feature space; (ii) for some writers, the model achieves a good separation between skilled forgeries and genuine signatures, but this is not the case for all writers, and (iii) some writers still have skilled forgeries which are close to genuine signatures.

With regard to the dissimilarity space representation: (i) signatures that are close in the feature space will be close to the origin in the dissimilarity space, and (ii) the further away two signatures are in the feature space, the farther the vector resulting from the dichotomy transformation will be from the origin (CHA; SRIHARI, 2000). Based on the feature space shown in Figure 16, it is expected that the resulting dissimilarity space will have the following characteristics:

- F_1: Since genuine signatures from the writers form dense clusters in the feature space, positive samples will be close to the origin, forming a dense cluster in the dissimilarity space.
- F_2: As random forgeries are genuine signatures from other writers and different writers occupy different regions of the feature space, negative samples from random forgeries will be far from the origin of the dissimilarity space.
- F_3: For the writers with a larger separation between skilled forgeries and genuine signatures, negative samples will be far from the origin in the dissimilarity space.
- F_4: For the writers that have skilled forgeries close to the genuine signatures, negative samples will be closer to the origin in the dissimilarity space (when compared to the other negative samples), and may even be within the space occupied by the positive samples.

To show that this behavior is actually present, we analyze the instance hardness of the samples in the dissimilarity space using the kDN metric (Eq. 3.2) in the validation set for verification V_v. A methodology similar to the one applied to obtain the exploitation dataset (section 4.1) is used here to obtain the dissimilarity space: (i) the reference set R is composed of just 1 (one) randomly selected genuine signature from each writer of the V_v set, and (ii) the questioned set Q is composed of 10 of the remaining genuine signatures and 10 skilled forgeries from each writer, plus 10 random forgeries, each one selected from a genuine signature of 10 different writers.

Figures 17, 18 and 19 present, for the GPDS-300 dataset, the histograms of the instance hardness considering: (i) all the data, (ii) just positive samples and negative samples from random forgeries, and (iii) just positive samples and negative samples from skilled forgeries, respectively.

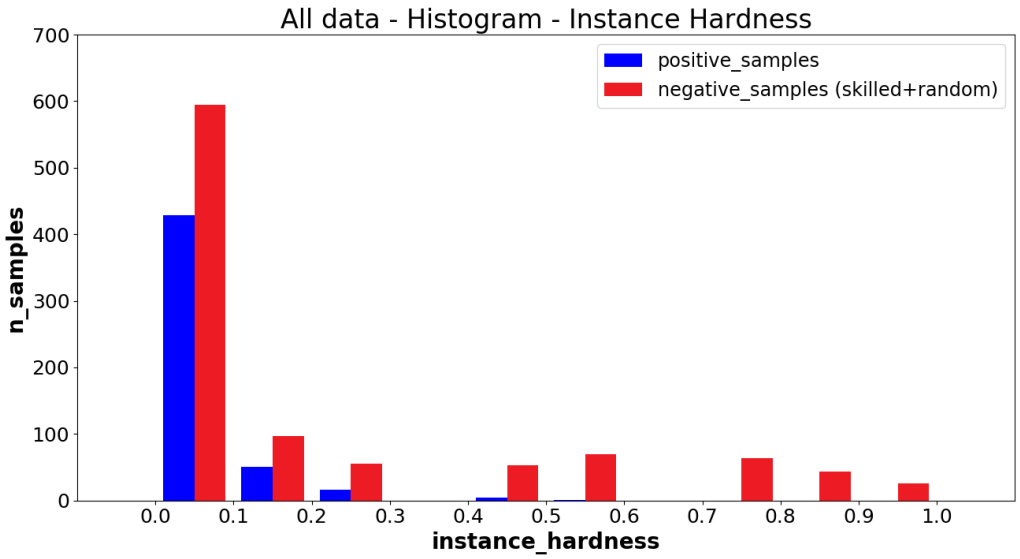

Fig. 17 – Instance hardness considering all selected data from the V_v segmentation of GPDS-300 dataset

As can be seen in Figure 17, for almost all the positive samples, $IH < 0.3$. So, in the dissimilarity space, since we are considering the kDN with $K = 7$, at least 5 of the 7 neighbors of the positive samples are from the positive class itself (F_1).

As shown in Figure 18, the following points can be seen when considering just the positive samples and the negative samples from the random forgeries: (i) The IH of all the positive samples are in the IH = 0.0 bin. Hence, for this scenario, all the neighbors of the positive samples are from the positive class itself. For this to occur, the positive samples should be concentrated in a dense region of the dissimilarity space, and no positive samples can go to the negative side of the space. Additionally, no negative sample is within the positive cluster (F_1). (ii) The IH of the negative samples are arranged along the histogram, and so the negative samples (random) should be in a sparse region of the dissimilarity

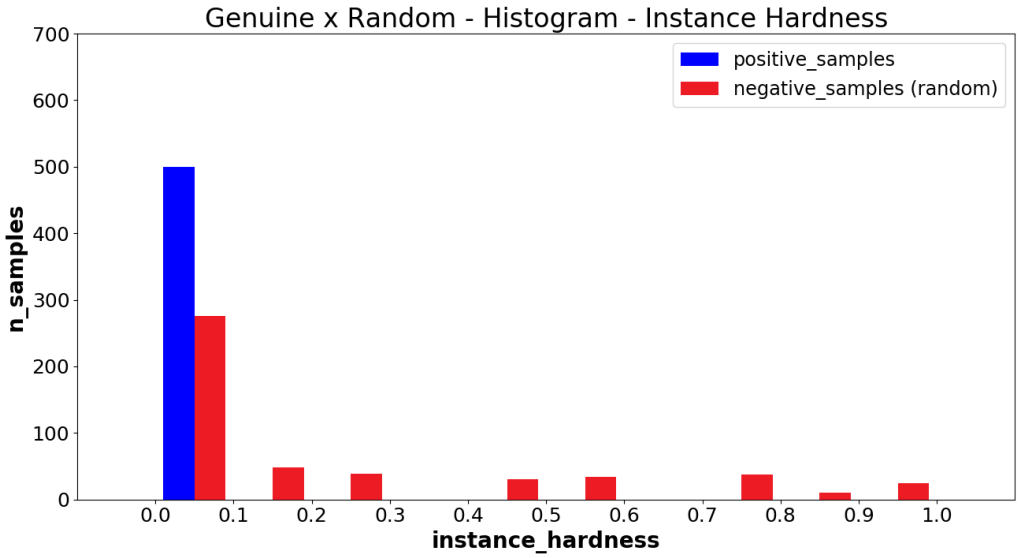

Fig. 18 – Instance hardness considering only the positive samples and negative samples (random forgeries) from the V_v segmentation of GPDS-300 dataset

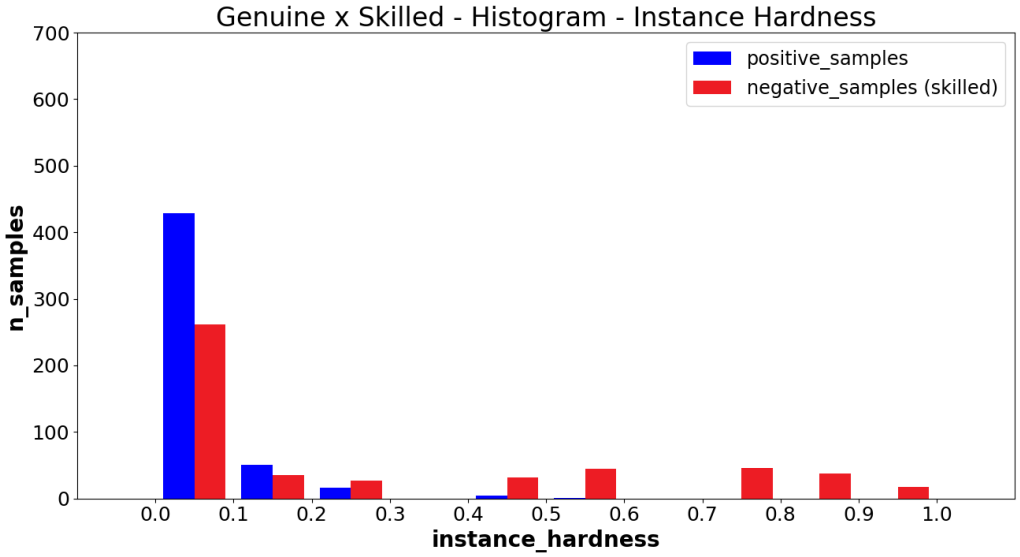

Fig. 19 – Instance hardness considering only the positive samples and negative samples (skilled forgeries) from the V_v segmentation of GPDS-300 dataset

space, and some samples should be in a region closer to the dense positive region of the space, since some samples have IH = 1.0 (F_2).

It is worth noting that the representation was able to actually separate positive samples from negative (random) ones, as all the positive samples were in the IH = 0.0 bin, i.e., there was no class overlap.

As can be seen in Figure 19, all the positive samples with $IH \neq 0.0$ from Figure 17 are derived from skilled forgeries. Thus, here, unlike in the negative samples (random) scenario, there should be class overlapping in the dissimilarity space (F_4). This behavior

is expected, since, in theory, the skilled forgeries are more similar to the genuine ones, when compared to random forgeries.

The following aspects must also be highlighted: (i) as the positive samples are concentrated on the left side and the negative samples (skilled) are arranged along the histogram, the negative samples should be more sparse than the positive ones in the dissimilarity space (F_3), and (ii) as the negative samples have samples with higher IH, the overlap of the classes should be in the positive region of the dissimilarity space (F_4).

If this same methodology is applied to the rest of the Development dataset (i.e., for the other 531 writers), the data will have a similar IH behavior with a larger number of samples. Making a uniform random selection to pick up the same number of samples as in V_v and performing the Kolmogorov-Smirnov test with a 5% level of significance, we see that both scenarios are drawn from the same continuous distribution in all scenarios. Therefore, the validation set for verification is representative of the Development set.

Generally speaking, positive samples are located in a dense cluster close to the origin and the negative samples are scattered throughout the dissimilarity space. Moreover, the clusters are disjointed, with a small overlap area, based on the concentration of the IH with low values. Considering that hard to classify samples are in the border region, the use of a condensation PS technique such as CNN has been shown to produce good experimental results because it retains samples in the decision boundaries (GARCIA et al., 2012). This IH analysis is also in line with the findings from the previous section regarding the use of transfer learning.

4.3.5 Lessons learned

In this section, we evaluated the use of prototype selection in a WI-SVM approach applied to the dissimilarity space resulting from dichotomy transformation.

The experimental results showed that, in the transfer learning scenario, with the features used, a WI-SVM trained in the GPDS-300 can be employed to verify signatures in the other datasets without any further transfer adaptation in the WI-HSV context and still obtain similar results when compared to both WD and WI classifiers trained and tested in their own datasets (research question 5).

Additionally, dichotomy transformation is able to increase the number of samples in the offline WI-HSV scenario, but many of the samples are redundant. By using prototype selection, it is possible to discard redundant training samples and still achieve a classification performance that is similar to or better than what is obtained by using all the training samples. Even being a classic and simple technique, the Condensed Nearest Neighbors (HART, 1968) applied systematically was able to select fewer prototypes and still maintain high performance levels when compared to both the SVM trained with the complete original training set and the random subsampling approach (research question 4).

Analyses performed using the IH measure have shown that, in general, positive samples are located in a dense cluster close to the origin, and negative ones are scattered throughout the dissimilarity space generated by the dichotomy transformation.

4.4 INSTANCE HARDNESS ANALYSES

The objectives of these experiments are: (i) analyse the accuracy as a function of the IH in the GPDS-300 dataset and in transfer learning; (ii) characterize "good" and "bad" quality skilled forgeries; (iii) extend the transfer learning analysis from Section 4.3, by training and testing the models in all the considered datasets (not just the model trained at GPDS).

In this section we answer the research questions 5 and 6 presented in Chapter 1.

4.4.1 Detailed experimental setup

The experiments in this section consider the training set after the Condensed Nearest Neighbors (CNN) preprocessing (SOUZA et al., 2019b; SOUZA et al., 2019a). Also, we fixed MAX as the fusion function and the highest number of references, since this results in better performance.

For the instance hardness (IH) analysis, $K = 7$ is used for the estimation of the kDN (CRUZ et al., 2017).

4.4.2 Results and discussion

4.4.2.1 Comparison with the state of the art

In this section we present the results on the GPDS-300 exploitation set, comparing the results with the state-of-the-art.

Table 26 contains both the comparison with the state of the art methods for the GPDS-300 dataset and also the results obtained by the WI-SVMs (with and without the CNN prototype selection).

In general, our WI approach obtains low EER that outperforms almost all other methods (eight of fourteen models), being comparable to (HAFEMANN; SABOURIN; OLIVEIRA, 2017a) and (HAFEMANN; OLIVEIRA; SABOURIN, 2018). It is overpassed only by the models reported in (HAFEMANN; OLIVEIRA; SABOURIN, 2018) (fine-tuned), (YILMAZ; OZTURK, 2018) and (ZOIS et al., 2019). Although these models presented the best results, they are writer-dependent (WD); thus, our approach offers the advantage of being much more scalable, since only one classifier is used, while theirs requires 300. Compared to the other WI models, our approach was able to outperform almost them all, except the model proposed by (ZOIS; ALEXANDRIDIS; ECONOMOU, 2019). It is worth noting that there is still room for improvement in our approach, such as, using ensemble or feature selection, which are approaches used in the paper by (ZOIS; ALEXANDRIDIS; ECONOMOU, 2019).

Table 26 – Comparison of EER with the state of the art in the GPDS-300 dataset (errors in %)

Type	HSV Approach	#Ref	#Models	EER
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2016)	12	300	12.83
WD	(SOLEIMANI; ARAABI; FOULADI, 2016)	10	300	20.94
WD	(ZOIS; ALEWIJNSE; ECONOMOU, 2016)	5	300	5.48
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	5	300	3.92 (0.18)
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	12	300	3.15 (0.18)
WD	(SERDOUK; NEMMOUR; CHIBANI, 2017)	10	300	9.30
WD	(HAFEMANN; OLIVEIRA; SABOURIN, 2018)	12	300	3.15 (0.14)
WD	(HAFEMANN; OLIVEIRA; SABOURIN, 2018) (fine-tuned)	12	300	0.41 (0.05)
WD	(YILMAZ; OZTURK, 2018)	12	300	0.88 (0.36)
WD	(ZOIS et al., 2019)	12	300	0.70
WI	(KUMAR; SHARMA; CHANDA, 2012)	1	1	13.76
WI	(ESKANDER; SABOURIN; GRANGER, 2013)	1	1	17.82
WI	(DUTTA; PAL; LLADOS, 2016)	N/A	1	11.21
WI	(HAMADENE; CHIBANI, 2016)	5	1	18.42
WI	(ZOIS; ALEXANDRIDIS; ECONOMOU, 2019)	5	1	3.06
WI	SVM_{max}	12	1	3.69 (0.18)
WI	CNN_SVM_{max}	12	1	3.47 (0.15)

4.4.2.2 Extended instance hardness analysis

In this section, some error evaluations considering the instance hardness are presented. For the skilled forgeries, some analysis considering the good and bad quality skilled forgeries were also carried out.

Unlike previous section (Section 4.3), where the instance hardness was computed considering the test set, in this section, the IH analyses are performed considering the neighborhood in the training set itself. By fixing the neighborhood in the training set itself, we can extrapolate the analysis to all the considered datasets (since for all of them, we are using the same classifier trained in the GPDS-300 dataset). Thus, to compute the IH value, each test sample is considered alone with the whole training set. Hence, in Equation 3.2, the query instance, x_q, is a test sample and the K nearest neighbors, $KNN(x_q)$, belong to the training set.

We also extended the IH analyses to have a better understanding of the decision boundary (class overlap region). To this end, we present the relationship of IH values and the accuracy (%) of the model when the user threshold of EER is used as the decision threshold.

Tables 27, 28 and 29 present the relationship of the IH and the accuracy (%) of the model when the user threshold of EER is used as decision threshold, respectively for the positive samples, negative samples from the random forgeries and negative samples from the skilled forgeries (for the GPDS-300 dataset). In the tables, the first column lists the IH values ($K = 7$) and the second column, the number of samples for the respective IH value. The other columns represent the accuracy (%) when considering the CNN-SVM and using, respectively, one (R1), five ($R5_{max}$), and twelve ($R12_{max}$) reference signatures.

First of all, we analyse the number of samples per IH value (second column). As can be seen in Table 27, positive samples presented a major concentration in the $IH = 0.0$ bin and almost all of them had $IH <= 0.14$, which shows that the positive samples form a compact cluster.

On the other hand, the negative samples were distributed along the IH values (Tables 28 and 29); this indicates that they are more sparsely distributed in the dissimilarity space than the positive samples. As the negative samples present higher IH values, including $IH = 1.0$, there may be an overlap of the classes, i.e., negative samples located inside the positive region of the dissimilarity space (all the negative sample neighbors belong to the positive class). These aspects are illustrated in the right part of Figure 4.

Moreover, the dashed line represents the limit where a kNN with $K = 7$ classifier performs the correct classification (that is, most of the neighborhood belong to the test samples belong to the correct class). As can be seen, a kNN classifier would obtain good results for the positive samples (due to the dense positive cluster), but it does not perform very well for the negative samples. For the negative samples, the high dimensionality, the data sparsity, the class overlap, and the presence of negative samples in the positive region of the dissimilarity space indicate the need for a strong discriminant classifier that can model complex distributions. That is why a kNN classifier fails on the classification. However, the CNN-SVM with *RBF* kernel can deal with it and obtains better results even operating with one reference (R1 columns).

Table 27 – Relationship between IH and accuracy (%) for the **positive samples**, for the GPDS-300 dataset

IH	#Samples	R1	$R5_{max}$	$R12_{max}$
0.00	**2330**	**95.02**	**96.26**	**97.03**
0.14	591	90.18	94.07	94.75
0.28	69	71.01	84.05	88.40
0.42	6	66.66	83.33	100.00
0.57	3	0.00	33.33	66.66
0.71	1	100.00	100.00	100.00
0.85	0	-	-	-
1.00	0	-	-	-

Table 28 – Relationship between IH and accuracy (%) for the **negative samples from the random forgeries**, for the GPDS-300 dataset

IH	#Samples	R1	$R5_{max}$	$R12_{max}$
0.00	498	100.00	100.00	100.00
0.14	488	100.00	100.00	100.00
0.28	461	100.00	100.00	100.00
0.42	415	100.00	100.00	100.00
0.57	418	100.00	100.00	100.00
0.71	323	99.38	99.69	99.69
0.85	276	99.27	100.00	100.00
1.00	121	99.17	100.00	100.00

Table 29 – Relationship between IH and accuracy (%) for the **negative samples from the skilled forgeries**, for the GPDS-300 dataset

IH	#Samples	R1	$R5_{max}$	$R12_{max}$
0.00	420	100.00	100.00	100.00
0.14	284	100.00	100.00	100.00
0.28	219	100.00	100.00	100.00
0.42	208	100.00	100.00	99.51
0.57	239	99.58	97.90	99.16
0.71	348	95.86	97.70	97.98
0.85	562	90.92	93.41	94.30
1.00	**720**	**81.52**	**88.05**	**90.69**

The overlap in the positive region of the DS and the necessity of a more complex decision boundary can also be observed in the first row of Table 27 and the last row of Table 29 (as highlighted). From Table 27, first line, notice that all neighborhood instances from the positive samples belong to the positive class itself ($IH = 0.0$). Even so, the classifier did not achieve a perfect classification. In the same way, from Table 29, the classifier can correctly classify most of the negative (skilled) samples presenting the neighborhood formed by the positive class ($IH = 1.0$).

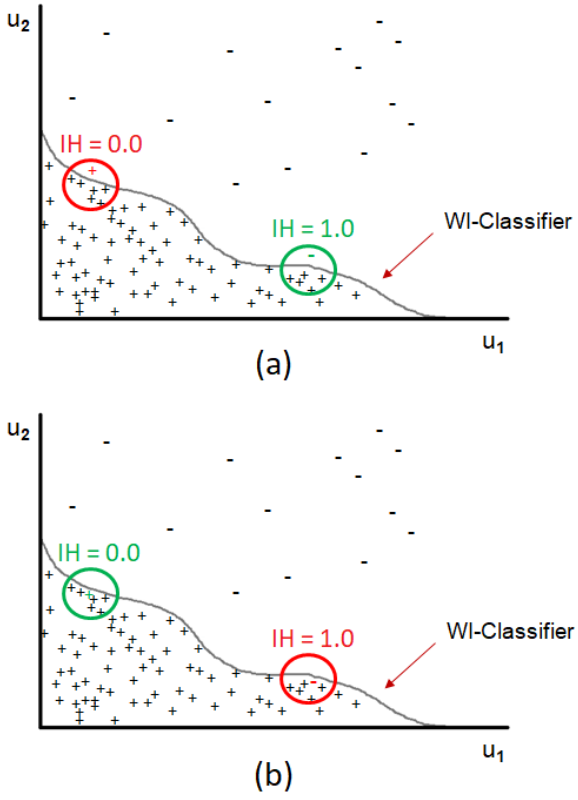

Fig. 20 – Synthetic decision frontiers: in (a) the scenario where the seven neighbors belong to the positive class and the model was able to correctly classify the sample from the negative class ($IH = 1.0$) but wrongly classified the positive test sample. In (b) the opposite scenario.

Figure 20 depicts this behavior in a synthetic representation. Considering the presented WI-classifier decision frontier, (a) ilustrattes the cases where the seven neighbors belong to the positive class the model was able to correctly classify the sample from the negative class ($IH = 1.0$) but wrongly classified the positive test sample ($IH = 0.0$). Figure 20 (b) illustrates the opposite scenario. In Appendix A, more analyses on this topic are carried out.

Specifically for the negative (skilled) forgeries, if we consider this same kNN limit to characterize the "bad quality skilled forgeries" ($IH <= 0.5$) and the "good quality skilled forgeries" ($IH > 0.5$), we can see that the CNN-SVM has an almost perfect performance for the bad quality skilled forgeries, independently of the number of references used (see the first four lines of Table 29). However, the higher the number of references used, the better the verification for the good quality skilled forgeries (the last four lines of Table 29).

From Table 28, the negative (random) samples are arranged along the IH values. This indicates that these samples are located in a sparse region of the dissimilarity space and some samples are closer to the region of the compact positive cluster in the space, because of the $IH = 1.0$ samples. However, the positive and the negative (random) sets may be disjoint, as the classifier presents an almost perfect verification performance. These aspects can also be seen in the right part of Figure 4.

4.4.2.3 Extended transfer learning analysis

In Section 4.3, we experimentally showed that a WI-SVM trained in the GPDS-300 can be employed to verify signatures in the other datasets without any further transfer adaptation in the WI-HSV context and still obtain similar results when compared to both WD and WI classifiers trained and tested in their own datasets.

In addition of using the CNN-SVM trained in the GPDS-300 dataset (referred to as CNN-SVM$_{gpds}$ in this section), we also extend the transfer learning analysis by training WI-SVM models in BRAZILIAN (CNN-SVM$_{brazilian}$), CEDAR (CNN-SVM$_{cedar}$) and MCYT (CNN-SVM$_{mcyt}$) datasets and testing them in the other datasets. Table 30 presents the comparison of these models.

As can be seen in Table 30, for the proposed approach, transfer learning models were able to outperform the models trained and tested in the own datasets for the CEDAR dataset and obtained similar results for the MCYT and GPDS-300 datasets. This shows that the proposed approach can actually be used in a transfer learning context, reinforcing the scalability and adaptability of the WI systems.

In the paper by (ZOIS; ALEXANDRIDIS; ECONOMOU, 2019), the authors also used a transfer learning methodology. In this scenario, our approach obtains comparable results, being better in the MCYT dataset and worse when CEDAR and GPDS-300 datasets are considered.

Table 30 – Comparison of EER for both scenarios where models are trained and tested in their own datasets and transfer learning, for the considered datasets (errors in %). CNN-SVM$_{brazilian}$, CNN-SVM$_{cedar}$, CNN-SVM$_{mcyt}$ and CNN-SVM$_{gpds}$ are respectively the models trained in the BRAZILIAN, CEDAR, MCYT and GPDS-300 datasets. The models from (ZOIS; ALEXANDRIDIS; ECONOMOU, 2019) follow the same terminology, so, $P_{2AD-cedar}$, $P_{2AD-mcyt}$, $P_{2AD-gpds}$ are respectively the models trained in the CEDAR, MCYT and GPDS-300 datasets.

Model	#Ref	$EER_{BRAZILIAN}$	EER_{CEDAR}	EER_{MCYT}	$EER_{GPDS-300}$
$P_{2AD-cedar}$	5	-	3.1	3.7	3.3
$P_{2AD-mcyt}$	5	-	2.9	4.6	**2.9**
$P_{2AD-gpds}$	5	-	**2.8**	3.4	3.7
CNN-SVM$_{brazilian}$	30	1.26 (0.33)	**3.12 (0.41)**	6.57 (0.33)	7.35 (0.34)
CNN-SVM$_{cedar}$	12	**0.72 (0.14)**	5.86 (0.50)	4.22 (0.77)	5.42 (0.26)
CNN-SVM$_{mcyt}$	12	1.16 (0.29)	4.21 (0.37)	2.99 (0.16)	3.57 (0.10)
CNN-SVM$_{gpds}$	10	1.11 (0.37)	3.32 (0.22)	**2.89 (0.13)**	**3.47 (0.15)**

Tables 31, 32 and 33 show some state-of-the-art results as well as the results obtained by the CNN-SVM$_{gpds}$ model (respectively, BRAZILIAN, CEDAR and MCYT as testing datasets), since the IH analysis will be performed through it (the motivation of using this model was presented in section 4.1).

Table 31 – Comparison of EER with the state of the art in the BRAZILIAN dataset (errors in %)

Type	HSV Approach	#Ref	#Models	EER
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2016)	15	60	4.17
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	5	60	2.92 (0.44)
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	15	60	2.07 (0.63)
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	30	60	2.01 (0.43)
WD	(HAFEMANN; OLIVEIRA; SABOURIN, 2018)	15	60	1.33 (0.65)
WD	(HAFEMANN; OLIVEIRA; SABOURIN, 2018) (finetuned)	15	60	1.35 (0.60)
WI	CNN-SVM$_{gpds}$	30	1	1.11 (0.37)

From Tables 31, 32 and 33, even operating in a transfer learning scenario our approach was able to obtain low verification errors that are at least comparable to the models derived from other state-of-the-art methods. For the BRAZILIAN dataset, our approach was able to outperform the state of the art methods. When compared to the WD models, our approach outperforms seven out of fourteen methods in CEDAR and is overpassed by only one of sixteen models in MCYT dataset. Still, our approach has the advantage of being scalable and using only one classifier to perform the verification. For the WI scenario, in the CEDAR dataset our approach presents better results than six of the nine models. When considering the MCYT dataset, our approach outperformed the results by (ZOIS; ALEXANDRIDIS; ECONOMOU, 2019).

It is worth noting that when our WI-classifier is used in the transfer learning scenario, it never had access to data from other datasets different from the one in which it was trained. Thus, combining DT and the used features representation allowed the model

Table 32 – Comparison of EER with the state of the art in the CEDAR dataset (errors in %)

Type	HSV Approach	#Ref	#Models	EER
WD	(BHARATHI; SHEKAR, 2013)	12	55	7.84
WD	(GANAPATHI; NADARAJAN, 2013)	14	55	6.01
WD	(SHEKAR; BHARATHI; PILAR, 2013)	16	55	9.58
WD	(OKAWA, 2016)	16	55	1.60
WD	(NEW..., 2016)	16	55	3.52
WD	(ZOIS; ALEWIJNSE; ECONOMOU, 2016)	5	55	4.12
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	12	55	4.76 (0.36)
WD	(ZOIS; THEODORAKOPOULOS; ECONOMOU, 2017)	5	55	2.07
WD	(HAFEMANN; OLIVEIRA; SABOURIN, 2018)	10	55	3.60 (1.26)
WD	(HAFEMANN; OLIVEIRA; SABOURIN, 2018) (fine-tuned)	10	55	2.33 (0.88)
WD	(OKAWA, 2018)	16	55	1.00
WD	(TSOUROUNIS et al., 2018)	5	55	2.82
WD	(ZOIS et al., 2018)	5	55	2.30
WD	(ZOIS et al., 2019)	10	55	0.79
WI	(KALERA; SRIHARI; XU, 2004)	16	1	21.9
WI	(CHEN; SRIHARI, 2006)	16	1	7.90
WI	(KUMAR et al., 2010)	1	1	11.81
WI	(KUMAR; SHARMA; CHANDA, 2012)	1	1	8.33
WI	(KUMAR; PUHAN, 2014)	16	1	6.02
WI	(GUERBAI; CHIBANI; HADJADJI, 2015)	12	1	5.60
WI	(DUTTA; PAL; LLADOS, 2016)	N/A	1	0.00
WI	(HAMADENE; CHIBANI, 2016)	5	1	2.11
WI	(ZOIS; ALEXANDRIDIS; ECONOMOU, 2019)	5	1	2.90
WI	CNN-SVM$_{gpds}$	12	1	3.32 (0.22)

Table 33 – Comparison of EER with the state of the art in the MCYT dataset (errors in %)

Type	HSV Approach	#Ref	#Models	EER
WD	(FIERREZ-AGUILAR et al., 2004)	10	75	9.28
WD	(ALONSO-FERNANDEZ et al., 2007)	5	75	22.4
WD	(GILPEREZ et al., 2008)	10	75	6.44
WD	(WEN et al., 2009)	5	75	15.02
WD	(VARGAS et al., 2011)	10	75	7.08
WD	(OOI et al., 2016)	10	75	9.87
WD	(SOLEIMANI; ARAABI; FOULADI, 2016)	10	75	9.86
WD	(ZOIS; ALEWIJNSE; ECONOMOU, 2016)	5	75	6.02
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	10	75	2.87 (0.42)
WD	(SERDOUK; NEMMOUR; CHIBANI, 2017)	10	75	18.15
WD	(ZOIS; THEODORAKOPOULOS; ECONOMOU, 2017)	5	75	3.97
WD	(HAFEMANN; OLIVEIRA; SABOURIN, 2018)	10	75	3.64 (1.04)
WD	(HAFEMANN; OLIVEIRA; SABOURIN, 2018) (fine-tuned)	10	75	3.40 (1.08)
WD	(OKAWA, 2018)	10	75	6.40
WD	(ZOIS et al., 2018)	5	75	3.52
WD	(ZOIS et al., 2019)	10	75	1.37
WI	(ZOIS; ALEXANDRIDIS; ECONOMOU, 2019)	5	1	3.50
WI	CNN-SVM$_{gpds}$	10	1	2.89 (0.13)

to remove the bias from signature acquisition protocols of the different datasets (e.g., scanner, writing space, type of writing tool).

4.4.2.4 IH analysis in transfer learning

Tables 34, 35 and 36 present the relationship of IH and the accuracy (%) of the model when the user threshold of EER is used as decision threshold, respectively for the positive samples, negative samples from the random forgeries and negative samples from the skilled forgeries, for the MCYT dataset. In these tables, the first column represents the possible IH values ($K = 7$) and the second column shows the number of samples for the respective IH value. The other columns present the accuracy (%) when using one (R1), five ($R5_{max}$) and ten ($R10_{max}$) references to perform the verification.

Table 34 – Relationship between IH and accuracy (%) for the **positive samples**, for the MCYT dataset

IH	#Samples	R1	$R5_{max}$	$R10_{max}$
0.00	**357**	**94.95**	**97.19**	**97.75**
0.14	16	62.50	100.00	100.00
0.28	1	0.00	100.00	100.00
0.42	0	-	-	-
0.57	0	-	-	-
0.71	1	0.00	100.00	100.00
0.85	0	-	-	-
1.00	0	-	-	-

Table 35 – Relationship between IH and accuracy (%) for the **negative samples from the random forgeries**, for the MCYT dataset

IH	#Samples	R1	$R5_{max}$	$R10_{max}$
0.00	9	100.00	100.00	100.00
0.14	51	100.00	100.00	100.00
0.28	63	100.00	100.00	100.00
0.42	94	100.00	100.00	100.00
0.57	109	100.00	100.00	100.00
0.71	123	100.00	100.00	100.00
0.85	160	99.37	100.00	100.00
1.00	141	100.00	100.00	100.00

Table 36 – Relationship between IH and accuracy (%) for the **negative samples from the skilled forgeries**, for the MCYT dataset

IH	#Samples	R1	$R5_{max}$	$R10_{max}$
0.00	0	-	-	-
0.14	2	100.00	100.00	100.00
0.28	9	100.00	100.00	100.00
0.42	22	100.00	100.00	100.00
0.57	34	97.05	100.00	100.00
0.71	101	99.00	99.00	99.00
0.85	255	96.47	98.43	98.43
1.00	**702**	**88.31**	**95.86**	**96.29**

As discussed before, we can consider that the dissimilarity space from different datasets as samples that belong to the same domain (signature representations in DS). Even the

data here presenting different concentration of samples per IH value, the used CNN-SVM presented similar behavior in the error analysis, when compared to the GDPS-300 scenario. From the first row of Table 34, the classifier did not achieve a perfect classification, even the positive samples presenting all their neighbor from the positive class ($IH = 0.0$). From the last row of Table 36, the classifier was able to correctly classify most of the negative (skilled), even the samples presenting the neighborhood formed by the positive class ($IH = 1.0$). Thus, this confirms the overlap in the positive region of the DS and the need for a more complex decision boundary.

From Table 35, the negative (random) samples are arranged along the IH values. This indicates the sparsity of these data in the dissimilarity space and that some samples are located closer to the compact positive region of the space, because of the samples with $IH = 1.0$. However, there is no overlap in the positive region, as the model achieved a perfect verification performance.

As can be seen in Appendix B and C, the WI approach also presented similar behavior, respectively, for the BRAZILIAN and CEDAR datasets. Thus, in all the scenarios, positive samples form a dense cluster (almost all positive samples have $IH \leq 0.14$), and the negative samples are scattered throughout space. The negative (random) samples may be disjoint to the positive set. The negative samples formed by the "good quality skilled forgeries" overlap the positive region of the DS, resulting in the need for a classifier with complex decision boundary.

4.4.3 Lessons learned

The experimental evaluations carried out in this section, were based on both the EER and IH metric, which allowed us to understand the difficulty of the HSV problem at the instance level.

Here we extended the transfer learning analysis from Section 4.3, using models trained and tested in all databases considered. The results obtained consolidate those found previously. This reaffirms the ability to use the proposed approach in a context of transfer learrning. Therefore, a single model already trained can be used to verify the signatures of new incoming writers without any further transfer adaptation (research question 5).

The reported IH analysis showed that the samples belonging to the positive class form a compact cluster located close to the origin and the negative samples are sparsely distributed in the dissimilarity space generated by the dichotomy transformation. In addition, based on the IH analysis, the overlap between positive and negative (skilled) samples is still present, so feature selection could be applied in the dissimilarity space in the attempt to separate these sets of samples.

Furthermore, we were able to characterize the "good" and "bad" quality skilled forgeries using the IH analysis and also the frontier between the hard to classify samples, which are genuine signatures and "good" skilled forgeries close to the frontier (research

question 6). And so, having a good feature representation of the signatures, like the one used in this study (characterized by different writers clustered in separate regions of the feature space) is very important for DT. The greater the separation between writers in the feature space, the smaller the overlap between the positive and negative classes in the dissimilarity space.

4.5 FEATURE SELECTION

These experiments aim to investigate both the use of feature selection through BPSO and the effectiveness of the overfitting control strategy. Thus, the experiments are conducted for four different situations: (i) the model without feature selection (i.e., the 2048 features are used); (ii) the model with feature selection using just the optimization and no validation stage; (iii) the model with feature selection using the validation stage, where candidate solutions are validated at the last iteration using the Selection set; (iv) the model with feature selection using the external archive, where candidate solutions are validated at all iterations using the Selection set (global validation).

The same analysis is also carried out to check whether the space generated by the feature selection can be used in a transfer learning context.

In this section we answer the research question 7 and 8 presented in Chapter 1.

4.5.1 Specifics in the dataset for this section

In the experiments the whole set of steps are carried out using GPDS-300 dataset, specifically in the GPDS-300 stratification (HAFEMANN; SABOURIN; OLIVEIRA, 2017b). MCYT and CEDAR datasets are considered only for test purpose on the transfer learning scenario. Figure 21 depicts the segmentation of the writers on the GPDS-300 dataset.

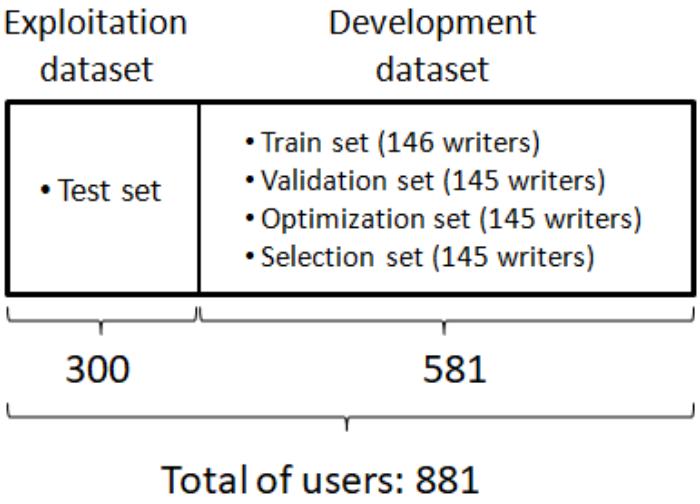

Fig. 21 – GPDS-300 dataset segmentation

As can be seen in Figure 21, (i) the Exploitation set, where the tested set is acquired, is composed of writers 1 to 300. (ii) The Development set is formed by the other 581 writers, from these: 146 writers are randomly selected to compose the train set ($train$), another 145 for the validation set (Val), another 145 for the optimization set (Opt) and the remaining 145 for the selection set (Sel).

As in Section 4.4, we use the highest value for the number of references, i.e., 12 references per writer, and the Max function as the partial decisions. In the training step (training and validation sets), the model uses genuine signatures and random forgeries. For each writer, 10 genuine signatures and 10 random forgeries are used as questioned signatures to obtain respectively the positive samples and the negative samples. In its turn, during optimization (optimization and selection sets), the proposed approach needs genuine signatures and skilled forgeries. As mentioned, the fitness function minimizes the EER with user threshold considering only genuine signatures and skilled forgeries. In this case, for each writer, 10 genuine signatures and 10 skilled forgeries are used. These operations are performed in the space with reduced samples, i.e., after prototype selection through Condensed Nearest Neighbors (CNN).

4.5.2 Detailed experimental setup

The experiments in this section consider the training set after the Condensed Nearest Neighbors (CNN) preprocessing.

The IDPSO parameters are set to their default values, as presented by Zhang et al. (ZHANG; XIONG; ZHANG, 2013). The population size is equal to 20, the acceleration constants are set to $c_1 = c_2 = 2.0$; $w_{inicial} = 0.9$, $w_{final} = 0.4$ and $\mu = 100$. The maximum number of iterations was set to 40. In Figure 22 column (a), we can see that 40 iterations were enough for the swarm to converge.

4.5.3 Results and discussions

Table 37 presents the results obtained by the following models: (i) without feature selection; (ii) with feature selection using just the optimization and no validation stage; (iii) with feature selection using the validation stage, where candidate solutions are validated at the last iteration using the Selection set; (iv) with feature selection using the external archive, where candidate solutions are validated at all iterations using the Selection set (global validation).

As can be seen, when the BPSO feature selection is used without validation the overfitting actually happened, and the lack of generalization power resulted in a worse EER when compared to the scenario without feature selection, 3.76 against 3.47.

In terms of validation strategy, results indicate that not using a validation stage is worse than using validation at the last iteration, which in turn is worse than using the global validation strategy. Thus, by using the global validation strategy it is possible to

Table 37 – Comparison of EER considering the presented models, in the GPDS-300 dataset (errors in %)

Approach	#features	EER
No feature selection	2048	3.47 (0.15)
Feature selection and no validation	1124	3.76 (0.07)
Feature selection and last iteration validation	1120	3.64 (0.08)
Feature selection and global validation (external archive)	1140	**3.46 (0.08)**

control the overfitting of the model and, thereby, improve the performance of the BPSO-based feature selection approach.

Another aspect that can be observed is the presence of redundant features in the dissimilarity space generated by the dichotomy transformation. To this end, notice that the model with feature selection and external archive uses only almost 55% of the total number of features and still manages to obtain similar EER when compared to the model trained with all the 2048 features.

Table 38 contains the comparison of the presented models with the state of the art methods for the GPDS-300 dataset. "(SOUZA et al., 2020) - Section 4.4" represents the WI-SVM trained in the original feature space, and, $SVM_{global-validation}$, the model with feature selection and global validation.

Table 38 – Comparison of EER with the state of the art, in the GPDS-300 dataset (errors in %)

Type	HSV Approach	#Ref	#Models	EER
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2016)	12	300	12.83
WD	(SOLEIMANI; ARAABI; FOULADI, 2016)	10	300	20.94
WD	(ZOIS; ALEWIJNSE; ECONOMOU, 2016)	5	300	5.48
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	5	300	3.92 (0.18)
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	12	300	3.15 (0.18)
WD	(SERDOUK; NEMMOUR; CHIBANI, 2017)	10	300	9.30
WD	(HAFEMANN; OLIVEIRA; SABOURIN, 2018)	12	300	3.15 (0.14)
WD	(HAFEMANN; OLIVEIRA; SABOURIN, 2018) (fine-tuned)	12	300	0.41 (0.05)
WD	(YILMAZ; OZTURK, 2018)	12	300	0.88 (0.36)
WD	(ZOIS et al., 2019)	12	300	0.70
WI	(KUMAR; SHARMA; CHANDA, 2012)	1	1	13.76
WI	(ESKANDER; SABOURIN; GRANGER, 2013)	1	1	17.82
WI	(DUTTA; PAL; LLADOS, 2016)	N/A	1	11.21
WI	(HAMADENE; CHIBANI, 2016)	5	1	18.42
WI	(ZOIS; ALEXANDRIDIS; ECONOMOU, 2019)	5	1	3.06
WI	(SOUZA et al., 2020) - Section 4.4	12	1	3.47 (0.15)
WI	CNN-SVM$_{global-validation}$	12	1	3.46 (0.08)

In general, our CNN-SVM$_{global-validation}$ approach obtains low EER. In the WI scenario, it was able to outperform almost all the other methods. However, it presents similar

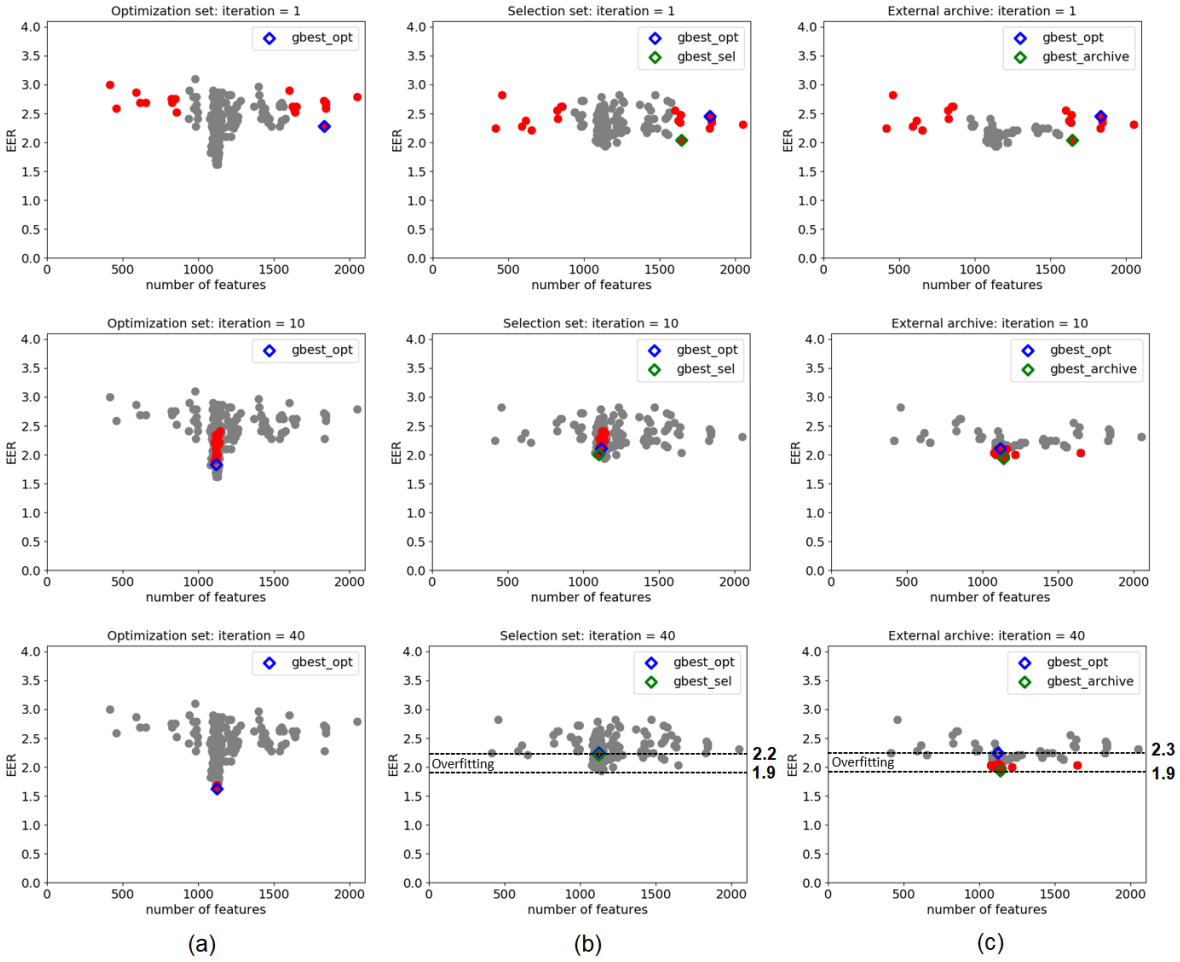

Fig. 22 – At first column (a) swarm behavior on the optimization set; in the second column (b) the swarm behavior when projected on the selection set; and in the third column (c) the swarm behavior in the external archive.

results to (SOUZA et al., 2020) and is worse when compared to the model proposed by (ZOIS; ALEXANDRIDIS; ECONOMOU, 2019).

In the comparison with WD models, was outperformed only by Hafemann et al. (HAFEMANN; OLIVEIRA; SABOURIN, 2018) (fine-tuned), Yilmaz and Ozturk (YILMAZ; OZTURK, 2018) and Zois et al. (ZOIS et al., 2019), being better or comparable than the other method of Table 38. It is important to point out that, as a WI model, our approach has greater scalability than these other models, since only one classifier is needed to perform signature verification.

4.5.4 Overfitting analysis

Figure 22 depicts the convergence of the swarm (iterations 1, 10 and 40 are presented). Gray dots represent the whole set of candidate solutions, considering all iterations. Red dots represent the particles in the respective iteration. Blue diamond represents the information of the best solution (*gBest*) from the optimization set. The green diamond

represents the information of the best solution found in the selection set.

We initialize the particles randomly in the intervals between [500, 1000] and [1500, 2048] dimensions. The objective was to extend the search space as much as possible. However, the particles soon converge into the space containing around half the maximum number of features (i.e., 1024).

The first column details algorithm convergence during the optimization process, considering the own optimization set. In the second column, the same solutions are projected on the validation objective function space, i.e., on the selection set (corresponding to the validation at the last iteration strategy). Finally, the third column simulates the convergence in the external archive obtained by validating all the candidate solutions on a selection set at each generation t (which corresponds to the global validation strategy).

As can be seen in the second column of Figure 22, considering the iteration 40, the overfitting actually happened when solutions are validated only at the last iteration. We can get an estimated overfitting of about 0.3 of EER, when compared to the lowest error rate in the external archive. The second column also indicates that some candidate solutions that perform well in the selection set are discarded by the algorithm. This observation also confirms the needing of a validation stage at each iteration t.

As depicted in the third column of Figure 22, considering iteration 40, we can also see the overfitting happening when solutions do not use any validation stage. The best candidate solution from the optimization set is almost 0.4 overfitted when compared to the lowest error rate in the external archive.

4.5.5 Transfer learning analysis

In Section 4.3, we experimentally showed that a WI-SVM trained in the GPDS-300 can be employed to verify signatures in the other datasets without any further transfer adaptation. Herein, we investigate whether the space generated by the feature selection approach can also be used in a transfer learning context.

Table 39 shows the results obtained when the models from Table 37, trained in the GPDS-300 dataset, are used to perform the verification in the CEDAR and MCYT databases.

Table 39 – Comparison of EER considering the presented models, in a transfer learning conxtext in the CEDAR and MCYT datasets (errors in %)

Approach	#features	EER_{CEDAR}	EER_{MCYT}
No feature selection	2048	3.32 (0.22)	2.89 (0.13)
Feature selection and no validation	1124	4.00 (0.17)	2.69 (0.13)
Feature selection and last iteration validation	1120	3.98 (0.25)	2.56 (0.05)
Feature selection and global validation (external archive)	1140	**3.27 (0.22)**	**2.48 (0.23)**

As can be seen, in terms of validation strategy, results indicate that not using a validation stage is worse than using validation at the last iteration, which in turn is worse

than using the global validation strategy (external archive). Thus, by using the global validation strategy it is possible to control the overfitting of the model and, thereby, improve the performance of the BPSO-based feature selection approach. This, behavior similar to that found in the GPDS-300 database (Table 37).

In the CEDAR dataset, the model with all the features obtained the better results, except the model with global validation strategy. In its turn, for the MCYT dataset, all models with feature selection obtained better results when compared to the one using the whole set of features, the model with global validation being the best. Recall that these models with feature selection use only almost 55% of the total number of features.

Table 40 – Comparison of EER with the state of the art in the CEDAR dataset (errors in %)

Type	HSV Approach	#Ref	#Models	EER
WD	(BHARATHI; SHEKAR, 2013)	12	55	7.84
WD	(GANAPATHI; NADARAJAN, 2013)	14	55	6.01
WD	(SHEKAR; BHARATHI; PILAR, 2013)	16	55	9.58
WD	(OKAWA, 2016)	16	55	1.60
WD	(NEW..., 2016)	16	55	3.52
WD	(ZOIS; ALEWIJNSE; ECONOMOU, 2016)	5	55	4.12
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	12	55	4.76 (0.36)
WD	(ZOIS; THEODORAKOPOULOS; ECONOMOU, 2017)	5	55	2.07
WD	(HAFEMANN; OLIVEIRA; SABOURIN, 2018)	10	55	3.60 (1.26)
WD	(HAFEMANN; OLIVEIRA; SABOURIN, 2018) (fine-tuned)	10	55	2.33 (0.88)
WD	(OKAWA, 2018)	16	55	1.00
WD	(TSOUROUNIS et al., 2018)	5	55	2.82
WD	(ZOIS et al., 2018)	5	55	2.30
WD	(ZOIS et al., 2019)	10	55	0.79
WI	(KALERA; SRIHARI; XU, 2004)	16	1	21.9
WI	(CHEN; SRIHARI, 2006)	16	1	7.90
WI	(KUMAR et al., 2010)	1	1	11.81
WI	(KUMAR; SHARMA; CHANDA, 2012)	1	1	8.33
WI	(KUMAR; PUHAN, 2014)	16	1	6.02
WI	(GUERBAI; CHIBANI; HADJADJI, 2015)	12	1	5.60
WI	(DUTTA; PAL; LLADOS, 2016)	N/A	1	0.00
WI	(HAMADENE; CHIBANI, 2016)	5	1	2.11
WI	(ZOIS; ALEXANDRIDIS; ECONOMOU, 2019)	5	1	2.90
WI	(SOUZA et al., 2020) - Section 4.4	12	1	3.32 (0.22)
WI	CNN-SVM$_{global-validation}$	12	1	3.27 (0.22)

From Tables 40 and 41, even operating in a transfer learning scenario the CNN-SVM$_{global-validation}$ model was able to obtain low verification errors, comparable to the other state of the art models. In the WD scenario, CNN-SVM$_{global-validation}$ outperforms half of the listed methods in CEDAR and is overpassed by only one method in MCYT dataset. Still, our approach has the advantage of being adaptable (since it is being used in a transfer learning context) and using only one classifier to perform the verification. For the WI comparison, in the CEDAR dataset our approach presents better results than seven of the ten models. When considering the MCYT dataset, our approach outperformed the results by both (SOUZA et al., 2020) and (ZOIS; ALEXANDRIDIS; ECONOMOU, 2019).

Table 41 – Comparison of EER with the state of the art in the MCYT dataset (errors in %)

Type	HSV Approach	#Ref	#Models	EER
WD	(FIERREZ-AGUILAR et al., 2004)	10	75	9.28
WD	(ALONSO-FERNANDEZ et al., 2007)	5	75	22.4
WD	(GILPEREZ et al., 2008)	10	75	6.44
WD	(WEN et al., 2009)	5	75	15.02
WD	(VARGAS et al., 2011)	10	75	7.08
WD	(OOI et al., 2016)	10	75	9.87
WD	(SOLEIMANI; ARAABI; FOULADI, 2016)	10	75	9.86
WD	(ZOIS; ALEWIJNSE; ECONOMOU, 2016)	5	75	6.02
WD	(HAFEMANN; SABOURIN; OLIVEIRA, 2017a)	10	75	2.87 (0.42)
WD	(SERDOUK; NEMMOUR; CHIBANI, 2017)	10	75	18.15
WD	(ZOIS; THEODORAKOPOULOS; ECONOMOU, 2017)	5	75	3.97
WD	(HAFEMANN; OLIVEIRA; SABOURIN, 2018)	10	75	3.64 (1.04)
WD	(HAFEMANN; OLIVEIRA; SABOURIN, 2018) (fine-tuned)	10	75	3.40 (1.08)
WD	(OKAWA, 2018)	10	75	6.40
WD	(ZOIS et al., 2018)	5	75	3.52
WD	(ZOIS et al., 2019)	10	75	1.37
WI	(ZOIS; ALEXANDRIDIS; ECONOMOU, 2019)	5	1	3.50
WI	(SOUZA et al., 2020) - Section 4.4	10	1	2.89 (0.13)
WI	CNN-SVM$_{global-validation}$	10	1	2.48 (0.23)

4.5.6 Instance hardness analysis

In this section, a comparative study based on the IH metric is performed to better understand the data difficulty in the optimized dissimilarity space (resulted from the model with feature selection and external archive, i.e., 1140 features) and the original dissimilarity space (no feature selection, i.e., 2048 features).

A methodology similar to that used in the previous section (Section 4.4) was used. Thus, to compute the IH value, each test sample is considered alone with the whole training set. Hence, in Equation 3.2, the query instance, x_q, is a test sample and the K nearest neighbors, $KNN(x_q)$, belong to the training set.

Tables 42, 43 and 44 present the relationship of the IH and the accuracy (%) of the model when the user threshold of EER is used as decision threshold, respectively for the positive samples, negative samples from the random forgeries and negative samples from the skilled forgeries (for the GPDS-300 dataset). In the tables, the first column lists the IH values ($K = 7$), the second and third columns contain respectively the number of samples and the accuracy (%) when considering the model using twelve reference signatures, for each IH value in the original space. The fourth and fifth columns, the same information in the optimized space.

Tables 45, 46 and 47 present the relationship of IH and the accuracy (%) of the model when the user threshold of EER is used as decision threshold, respectively for the positive samples, negative samples from the random forgeries and negative samples from the skilled forgeries, for the MCYT dataset. They follow the same structure as the previous tables. These evaluations consider the CNN-SVM trained using the training set of GPDS-300

Table 42 – Relationship between IH and accuracy (%) for the **positive samples**, for the GPDS-300 dataset

	Original Space		**Optimized Space**	
IH	#Samples	$R12_{max}$	#Samples	$R12_{max}$
0.00	2330	97.03	2387	96.19
0.14	591	94.75	443	93.45
0.28	69	88.40	117	87.18
0.42	6	100.00	37	97.30
0.57	3	66.66	10	100.00
0.71	1	100.00	4	75.00
0.85	0	-	2	100.00
1.00	0	-	0	-

Table 43 – Relationship between IH and accuracy (%) for the **negative samples from the random forgeries**, for the GPDS-300 dataset

	Original Space		**Optimized Space**	
IH	#Samples	$R12_{max}$	#Samples	$R12_{max}$
0.00	498	100.00	436	100.00
0.14	488	100.00	537	100.00
0.28	461	100.00	519	100.00
0.42	415	100.00	441	100.00
0.57	418	100.00	423	100.00
0.71	323	99.69	349	100.00
0.85	276	100.00	199	100.00
1.00	121	100.00	96	100.00

Table 44 – Relationship between IH and accuracy (%) for the **negative samples from the skilled forgeries**, for the GPDS-300 dataset

	Original Space		**Optimized Space**	
IH	#Samples	$R12_{max}$	#Samples	$R12_{max}$
0.00	420	100.00	316	100.00
0.14	284	100.00	270	100.00
0.28	219	100.00	233	100.00
0.42	208	100.00	270	100.00
0.57	239	99.58	294	100.00
0.71	348	95.86	376	97.61
0.85	562	90.92	533	95.50
1.00	720	81.52	708	89.12

dataset in both original and optimized space.

As can be seen from the presented tables, even with different number of features in each space, in general, for both the GPDS-300 and MCYT datasets, there is no significant difference between the frequencies of the number of samples for each IH value, for both spaces. This analysis can be observed for positive, negative (random) or for negative (skilled) samples. As can be seen in Appendix D, this approach also presented similar behavior for the CEDAR datasets.

In this context, given that the frequency of negative (skilled) samples has not changed, there has been no change in the number of "good" quality skilled forgeries. Thus, for the considered datasets, the proposed feature selection technique was able to reduce the

Table 45 – Relationship between IH and accuracy (%) for the **positive samples**, for the MCYT dataset

	Original Space		**Optimized Space**	
IH	#Samples	$R10_{max}$	#Samples	$R10_{max}$
0.00	357	97.75	347	97.12
0.14	16	100.00	22	90.91
0.28	1	100.00	5	80.00
0.42	0	-	1	100.00
0.57	0	-	0	-
0.71	1	100.00	0	-
0.85	0	-	0	-
1.00	0	-	0	-

Table 46 – Relationship between IH and accuracy (%) for the **negative samples from the random forgeries**, for the MCYT dataset

	Original Space		**Optimized Space**	
IH	#Samples	$R10_{max}$	#Samples	$R10_{max}$
0.00	9	100.00	14	100.00
0.14	51	100.00	29	100.00
0.28	63	100.00	75	100.00
0.42	94	100.00	76	100.00
0.57	109	100.00	100	100.00
0.71	123	100.00	129	100.00
0.85	160	100.00	176	100.00
1.00	141	100.00	151	100.00

Table 47 – Relationship between IH and accuracy (%) for the **negative samples from the skilled forgeries**, for the MCYT dataset

	Original Space		**Optimized Space**	
IH	#Samples	$R10_{max}$	#Samples	$R10_{max}$
0.00	0	-	1	100.00
0.14	2	100.00	0	-
0.28	9	100.00	5	100.00
0.42	22	100.00	28	100.00
0.57	34	100.00	65	98.46
0.71	101	99.00	121	99.17
0.85	255	98.43	269	97.77
1.00	702	96.29	636	95.44

redundancy of the features, but did not result in a space with greater separation between the positive and negative samples. This may explain the similar results presented by the model in the original space (models from Section 4.3) and in the optimized space (CNN-SVM$_{global-validation}$), as presented in Tables 38, 40 and 41.

4.5.7 Lessons learned

In this section, we evaluated the use of BPSO-based feature selection for offline writer-independent handwritten signature verification. The optimization was conducted based on the minimization of the Equal Error Rate (EER) of the SVM in a wrapper mode.

Experimental results showed the presence of overfitting when no validation is used,

given that the lack of generalization power resulted in a worse EER when compared to the scenario without feature selection.

Results also showed that not using a validation stage is worse than using validation at the last iteration, which in turn is worse than using the global validation strategy. Thus, by using the global validation strategy it is possible to control the overfitting of the model and, thereby, improve the performance of the BPSO-based feature selection approach (research question 8).

Another aspect that can be observed is the presence of redundant features in the dissimilarity space generated by the dichotomy transformation, since the models with a validation stage managed to obtain a better EER using only almost 55% of the total number of features (research question 7). However, it did not result in a space with greater separation between the positive and negative samples, as presented in the IH analysis.

Finally, the experiments demonstrated that the space generated after feature selection can actually be used in a transfer learning context.

5 CONCLUSION

In this study, we proposed and investigated a novel writer-independent (WI) framework for handwritten signatures verification (HSV). The proposed method is based on a deep understanding of the dichotomy transformation (DT) applied in a WI framework for handwritten signatures verification. The experimental evaluations, carried out in four datasets, were based on both the EER and IH metric, which allowed us to understand the difficulty of the HSV problem at the instance level.

Understanding of why the instances are misclassified (IH analysis) led us to the development of an approach well suited to our WI-HSV scenario, by directly addressing the causes of the misclassification.

The first choice was which features representation to use. An important aspect of DT is the needing for a good feature representation, the *Signet*, used in this paper, is well adapted to this scenario as different writers are clustered in separate regions of the feature space. Another aspect is that, regardless of the signature image, *Signet* will generate feature vectors containing 2048 dimensions. This fact, facilitates the use of this feature representation in a context of transfer learning.

We also showed that, by using the highest number of references and MAX as fusion function, the approach dynamically selects the reference (from the set of references) that is most similar to the questioned signature and uses it to perform the verification.

Another aspect in this work was the use of the classical Condensed Nearest Neighbors as systematic prototypes selection. This approach maintains the instances that are misclassified by a 1-NN classifier, discarding them otherwise. Its goal is to reduce the dataset size by removing redundant instances, maintaining the samples in the decision boundaries.

As the verification only depends on the input reference signature, WI systems have the advantages of being scalable and adaptable. By using the DT in a writer-independent approach, the dichotomizer (classifier) can verify signatures of writers for whom the classifier was not trained in a transfer learning context. Therefore, a single model already trained can be used to verify the signatures of new incoming writers without any further transfer adaptation.

The option of using a Binary Particle Swarm Optimization algorithm for feature selection, was that it has obtained good results in different classification problems when compared to other optimization algorithms used for this task. In this binary swarm optimization scenario, we propose to use a BPSO-based feature selection for WI handwritten signature verification in a wrapper mode. To decrease the chance of overfitting, we proposed a global validation strategy, where the validation of the candidate solutions is executed in all iterations of the optimization process and an external archive is responsible

to store the best validated solutions

Also, the reported IH analysis showed that the samples belonging to the positive class form a compact cluster located close to the origin and the negative samples are sparsely distributed in the dissimilarity space generated by the dichotomy transformation. Furthermore, we were able to characterize "good" and "bad" quality skilled forgeries using the IH analysis and also the frontier between the hard to classify samples, which are genuine signatures and "good" skilled forgeries close to the frontier.

To conclude, in Chapter 1 the following research questions were presented:

1. How the writer-independent dichotomy transformation can handle the HSV data difficulties?

    - In Section 3.1.1 we present how the proposed WI approach handle the challenges faced when dealing with the HSV problem. Among them, (C_1) the high number of writers (classes), (C_3) the small number of training samples per writer with high intra-class variability, (C_4) the heavily imbalanced class distributions and manage new incoming writers (C_6)

2. Does the number of reference signatures used influence the ability to verify signatures in the writer-independent model?

    - By using the highest number of available reference signatures, the WI model achieves better verification performance (a more detailed answer is presented in Section 4.2). For instance, for the GPDS-300 dataset we have an average error (when considering a global threshold) $AER_{genuine+skilled} = 8.0$ for the model using 12 references against $AER_{genuine+skilled} = 13.1$, when using only one reference.

3. What is the best fusion function to be used to combine partial decisions in a scenario with multiple reference signatures?

    - The best results are obtained using the MAX as fusion function (a more detailed answer is presented in Section 4.2). For the GPDS-300 dataset, we have an average error (when considering a global threshold) $AER_{genuine+skilled} = 8.0$ for the model using MAX as fusion function against $AER_{genuine+skilled} = 13.6$, when using only one reference.

4. Does the dissimilarity space generated by the dichotomy transformation have samples with redundant information, i.e., with little importance for training purposes? Can we use prototype selection methods for eliminating redundant training data?

    - The Condensed Nearest Neighbors (CNN) applied systematically is able to select fewer prototypes and still maintain high performance levels when compared to the SVM trained with the complete original training set (a more

detailed answer is presented in Section 4.3). For the GPDS-300 dataset, the CNN_SVM was trained with almost 5% of the total number of samples and obtained $EER = 3.47$, agains an $EER = 3.69$ from the SVM trained using the whole training set.

5. Can the writer-independent approach be used in the context of transfer learning and still obtain good verification performance?

   - A WI-classifier trained in one dataset can be employed to verify signatures in the other datasets without any further transfer adaptation in the WI-HSV context (a more detailed answer is presented in Section 4.3). For instance, the SVM trained in the GPDS-300 dataset and used to verify signatures from the CEDAR dataset obtained an $EER = 3.42$ while the SVM trained and tested in the own CEDAR obtained an $EER = 5.78$.

6. Can skilled forgeries be characterized as having "good" or "bad" quality based on the measure of instance hardness?

   - In the skilled forgeries scenarios, we can consider the same kNN limit to characterize the "bad quality skilled forgeries" ($IH <= 0.5$) and the "good" quality skilled forgeries ($IH > 0.5$) (a more detailed answer is presented in Section 4.4). Bad quality skilled forgeries are located far from the positive cluster ($IH = 0.0$) and are easily classified, the good quality ones are difficult to classify and can even located within the positive class cluster ($IH = 1.0$).

7. Does the generated dissimilarity space have redundant features?

   - There are redundant features in the dissimilarity space generated by the dichotomy transformation, since the models with a global validation stage managed to obtain a better verification performance using only almost 55% of the total number of features (a more detailed answer is presented in Section 4.5).

8. Can overfitting control improve the performance of the optimization in the feature selection scenario?

   - By using the global validation with external archive strategy it is possible to control the overfitting and, thereby, improve the performance of the BPSO-based feature selection approach (a more detailed answer is presented in Section 4.5). For the GPDS-300 dataset, not using an overfitting control strategy resulted in $EER = 3.76$ against $EER = 3.46$ obtained by the model using the proposed global strategy with external archive.

5.1 FUTURE WORKS

The DT characteristics and the analyses reported in this study serve as motivation for future works aiming at improving the discrimination between genuine signatures and forgeries, focusing mainly on discriminating between good quality skilled forgeries. Some suggestions for future work include:

- A cascade classification structure, with a rejection mechanism;
- Ensemble learning and dynamic selection adapted to work on the Dissimilarity Space;
- The use of multiobjective algorithms for feature selection, in order to minimize both the EER and the number of features during the optimization;
- The use of feature selection combined with prototype selection in a single optimization process, to deal with the redundancy in the dissimilarity at both the feature and the sample levels.

REFERENCES

ALONSO-FERNANDEZ, F.; FAIRHURST, M. C.; FIERREZ, J.; ORTEGA-GARCIA, J. Automatic measures for predicting performance in off-line signature. In: *2007 IEEE International Conference on Image Processing*. [S.l.: s.n.], 2007. v. 1, p. I – 369–I – 372. ISSN 2381-8549.

BATISTA, L.; GRANGER, E.; SABOURIN, R. Dynamic selection of generative–discriminative ensembles for off-line signature verification. *Pattern Recognition*, v. 45, n. 4, p. 1326 – 1340, 2012. ISSN 0031-3203.

BERTOLINI, D.; OLIVEIRA, L.; JUSTINO, E.; SABOURIN, R. Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers. *Pattern Recognition*, v. 43, n. 1, p. 387 – 396, 2010. ISSN 0031-3203.

BERTOLINI, D.; OLIVEIRA, L. S.; SABOURIN, R. Multi-script writer identification using dissimilarity. In: IEEE. *2016 23rd International Conference on Pattern Recognition (ICPR)*. [S.l.], 2016. p. 3025–3030.

BHARATHI, R.; SHEKAR, B. Off-line signature verification based on chain code histogram and support vector machine. In: IEEE. *Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on*. [S.l.], 2013. p. 2063–2068.

BOUAMRA, W.; DJEDDI, C.; NINI, B.; DIAZ, M.; SIDDIQI, I. Towards the design of an offline signature verifier based on a small number of genuine samples for training. *Expert Systems with Applications*, Elsevier, v. 107, p. 182–195, 2018.

CHA, S.-H.; SRIHARI, S. N. Writer identification: statistical analysis and dichotomizer. In: SPRINGER. *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. [S.l.], 2000. p. 123–132.

CHEN, S.; SRIHARI, S. A new off-line signature verification method based on graph. In: *18th International Conference on Pattern Recognition (ICPR'06)*. [S.l.: s.n.], 2006. v. 2, p. 869–872. ISSN 1051-4651.

CHUANG, L.-Y.; TSAI, S.-W.; YANG, C.-H. Improved binary particle swarm optimization using catfish effect for feature selection. *Expert Systems with Applications*, v. 38, n. 10, p. 12699 – 12707, 2011. ISSN 0957-4174.

CRUZ, R. M.; SABOURIN, R.; CAVALCANTI, G. D. Dynamic classifier selection: Recent advances and perspectives. *Information Fusion*, Elsevier, v. 41, p. 195–216, 2017.

CRUZ, R. M.; ZAKANE, H. H.; SABOURIN, R.; CAVALCANTI, G. D. Dynamic ensemble selection vs k-nn: Why and when dynamic selection obtains higher classification performance? In: *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*. [S.l.: s.n.], 2017. p. 1–6. ISSN 2154-512X.

DUTTA, A.; PAL, U.; LLADOS, J. Compact correlated features for writer independent signature verification. In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. [S.l.: s.n.], 2016. p. 3422–3427. ISSN null.

ESKANDER, G. S.; SABOURIN, R.; GRANGER, E. Hybrid writer-independent–writer-dependent offline signature verification system. *IET biometrics*, IET, v. 2, n. 4, p. 169–181, 2013. ISSN 2047-4938.

FIERREZ-AGUILAR, J.; ALONSO-HERMIRA, N.; MORENO-MARQUEZ, G.; ORTEGA-GARCIA, J. An off-line signature verification system based on fusion of local and global information. In: SPRINGER. *International Workshop on Biometric Authentication*. [S.l.], 2004. p. 295–306.

FREITAS, C.; BORTOLOZZI, F.; SABOURIN, R.; FACON, J. Bases de dados de cheques bancarios brasileiros. September 2000.

GANAPATHI, G.; NADARAJAN, R. A fuzzy hybrid framework for offline signature verification. In: SPRINGER. *International Conference on Pattern Recognition and Machine Intelligence*. [S.l.], 2013. p. 121–127.

GARCIA, S.; DERRAC, J.; CANO, J.; HERRERA, F. Prototype selection for nearest neighbor classification: Taxonomy and empirical study. *IEEE T PATTERN ANAL*, IEEE, v. 34, n. 3, p. 417–435, 2012.

GILPEREZ, A.; ALONSO-FERNANDEZ, F.; PECHARROMAN, S.; FIERREZ, J.; ORTEGA-GARCIA, J. Off-line signature verification using contour features. In: CENPARMI, CONCORDIA UNIVERSITY. *11th International Conference on Frontiers in Handwriting Recognition, Montreal, Quebec-Canada, August 19-21, 2008*. [S.l.], 2008.

GUERBAI, Y.; CHIBANI, Y.; HADJADJI, B. The effective use of the one-class svm classifier for handwritten signature verification based on writer-independent parameters. *Pattern Recognition*, v. 48, n. 1, p. 103 – 113, 2015. ISSN 0031-3203.

GURU, D.; MANJUNATHA, K.; MANJUNATH, S.; SOMASHEKARA, M. Interval valued symbolic representation of writer dependent features for online signature verification. *Expert Systems with Applications*, v. 80, p. 232 – 243, 2017. ISSN 0957-4174.

GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. *Journal of machine learning research*, v. 3, n. Mar, p. 1157–1182, 2003.

HAFEMANN, L. G.; OLIVEIRA, L. S.; SABOURIN, R. Fixed-sized representation learning from offline handwritten signatures of different sizes. *International Journal on Document Analysis and Recognition (IJDAR)*, v. 21, n. 3, p. 219–232, Sep 2018. ISSN 1433-2825.

HAFEMANN, L. G.; SABOURIN, R.; OLIVEIRA, L. S. Writer-independent feature learning for offline signature verification using deep convolutional neural networks. In: *2016 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2016. p. 2576–2583. ISSN 2161-4407.

HAFEMANN, L. G.; SABOURIN, R.; OLIVEIRA, L. S. Learning features for offline handwritten signature verification using deep convolutional neural networks. *Pattern Recognition*, v. 70, p. 163 – 176, 2017. ISSN 0031-3203.

HAFEMANN, L. G.; SABOURIN, R.; OLIVEIRA, L. S. Offline handwritten signature verification — literature review. In: *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*. [S.l.: s.n.], 2017. p. 1–8. ISSN 2154-512X.

HAMADENE, A.; CHIBANI, Y. One-class writer-independent offline signature verification using feature dissimilarity thresholding. *IEEE Transactions on Information Forensics and Security*, IEEE, v. 11, n. 6, p. 1226–1238, 2016.

HART, P. The condensed nearest neighbor rule (corresp.). *IEEE transactions on information theory*, Citeseer, v. 14, n. 3, p. 515–516, 1968.

HASSAN, R.; COHANIM, B.; WECK, O. D.; VENTER, G. A comparison of particle swarm optimization and the genetic algorithm. In: *Proceedings of the 1st AIAA multidisciplinary design optimization specialist conference*. [S.l.: s.n.], 2005. p. 1–13.

HOUMANI, N.; GARCIA-SALICETTI, S.; DORIZZI, B.; MONTALVÃO, J.; CANUTO, J. C.; ANDRADE, M. V.; QIAO, Y.; WANG, X.; SCHEIDAT, T.; MAKRUSHIN, A. et al. Biosecure signature evaluation campaign (esra'2011): evaluating systems on quality-based categories of skilled forgeries. In: *2011 International Joint Conference on Biometrics (IJCB)*. [S.l.: s.n.], 2011. p. 1–10.

HU, J.; CHEN, Y. Offline signature verification using real adaboost classifier combination of pseudo-dynamic features. In: IEEE. *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*. [S.l.], 2013. p. 1345–1349.

KALERA, M. K.; SRIHARI, S.; XU, A. Offline signature verification and identification using distance statistics. *International Journal of Pattern Recognition and Artificial Intelligence*, World Scientific, v. 18, n. 07, p. 1339–1360, 2004.

KENNEDY, J.; EBERHART, R. C. Particle swarm optimization. In: *Proceedings of IEEE International Conference on Neural Networks*. [S.l.: s.n.], 1995. p. 1942–1948.

KUMAR, M. M.; PUHAN, N. B. Off-line signature verification: upper and lower envelope shape analysis using chord moments. *IET Biometrics*, IET, v. 3, n. 4, p. 347–354, 2014.

KUMAR, R.; KUNDU, L.; CHANDA, B.; SHARMA, J. D. A writer-independent off-line signature verification system based on signature morphology. In: *Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia*. [S.l.]: ACM, 2010. (IITM '10), p. 261–265. ISBN 978-1-4503-0408-5.

KUMAR, R.; SHARMA, J.; CHANDA, B. Writer-independent off-line signature verification using surroundedness feature. *Pattern recognition letters*, Elsevier, v. 33, n. 3, p. 301–308, 2012.

LORENA, A. C.; GARCIA, L. P.; LEHMANN, J.; SOUTO, M. C.; HO, T. K. How complex is your classification problem? a survey on measuring classification complexity. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 52, n. 5, p. 1–34, 2019.

MARTINS, J. G.; OLIVEIRA, L. S.; BRITTO, A. S.; SABOURIN, R. Forest species recognition based on dynamic classifier selection and dissimilarity feature vector representation. *Machine Vision and Applications*, v. 26, n. 2, p. 279–293, Apr 2015. ISSN 1432-1769.

MASOUDNIA, S.; MERSA, O.; ARAABI, B. N.; VAHABIE, A.-H.; SADEGHI, M. A.; AHMADABADI, M. N. Multi-representational learning for offline signature verification using multi-loss snapshot ensemble of cnns. *Expert Systems with Applications*, v. 133, p. 317 – 330, 2019. ISSN 0957-4174.

MIRJALILI, S.; LEWIS, A. S-shaped versus v-shaped transfer functions for binary particle swarm optimization. *Swarm and Evolutionary Computation*, v. 9, p. 1 – 14, 2013. ISSN 2210-6502.

NEW off-line Handwritten Signature Verification method based on Artificial Immune Recognition System. *Expert Systems with Applications*, v. 51, p. 186 – 194, 2016. ISSN 0957-4174.

OKAWA, M. Offline signature verification based on bag-of-visual words model using kaze features and weighting schemes. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. [S.l.: s.n.], 2016. p. 184–190.

OKAWA, M. From bovw to vlad with kaze features: Offline signature verification considering cognitive processes of forensic experts. *Pattern Recognition Letters*, v. 113, p. 75 – 82, 2018. ISSN 0167-8655. Integrating Biometrics and Forensics.

OOI, S. Y.; TEOH, A. B. J.; PANG, Y. H.; HIEW, B. Y. Image-based handwritten signature verification using hybrid methods of discrete radon transform, principal component analysis and probabilistic neural network. *Applied Soft Computing*, Elsevier, v. 40, p. 274–282, 2016.

ORTEGA-GARCIA, J.; FIERREZ-AGUILAR, J.; SIMON, D.; GONZALEZ, J.; FAUNDEZ-ZANUY, M.; ESPINOSA, V.; SATUE, A.; HERNAEZ, I.; IGARZA, J.-J.; VIVARACHO, C. et al. Mcyt baseline corpus: a bimodal biometric database. *IEE Proceedings-Vision, Image and Signal Processing*, IET, v. 150, n. 6, p. 395–401, 2003.

PAN, S. J.; YANG, Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, v. 22, n. 10, p. 1345–1359, Oct 2010. ISSN 1041-4347.

PEKALSKA, E.; DUIN, R. P.; PACLIK, P. Prototype selection for dissimilarity-based classifiers. *Pattern Recognition*, v. 39, n. 2, p. 189 – 208, 2006. ISSN 0031-3203. Part Special Issue: Complexity Reduction.

RADTKE, P. V. W.; WONG, T.; SABOURIN, R. An evaluation of over-fit control strategies for multi-objective evolutionary optimization. In: *The 2006 IEEE International Joint Conference on Neural Network Proceedings*. [S.l.: s.n.], 2006. p. 3327–3334. ISSN 2161-4407.

RANTZSCH, H.; YANG, H.; MEINEL, C. Signature embedding: Writer independent offline signature verification with deep metric learning. In: SPRINGER. *International Symposium on Visual Computing*. [S.l.], 2016. p. 616–625.

RIVARD, D.; GRANGER, E.; SABOURIN, R. Multi-feature extraction and selection in writer-independent off-line signature verification. *International Journal on Document Analysis and Recognition (IJDAR)*, v. 16, n. 1, p. 83–103, Mar 2013. ISSN 1433-2825.

SANTOS, E. M. D.; SABOURIN, R.; MAUPIN, P. Overfitting cautious selection of classifier ensembles with genetic algorithms. *Information Fusion*, v. 10, n. 2, p. 150 – 162, 2009. ISSN 1566-2535.

SERDOUK, Y.; NEMMOUR, H.; CHIBANI, Y. Handwritten signature verification using the quad-tree histogram of templates and a support vector-based artificial immune classification. *Image and Vision Computing*, v. 66, p. 26 – 35, 2017. ISSN 0262-8856.

SHAO, L.; ZHU, F.; LI, X. Transfer learning for visual categorization: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, v. 26, n. 5, p. 1019–1034, May 2015. ISSN 2162-237X.

SHEKAR, B.; BHARATHI, R.; PILAR, B. Local morphological pattern spectrum based approach for off-line signature verification. In: SPRINGER. *International Conference on Pattern Recognition and Machine Intelligence*. [S.l.], 2013. p. 335–342.

SMITH, M. R.; MARTINEZ, T.; GIRAUD-CARRIER, C. An instance level analysis of data complexity. *Machine Learning*, v. 95, n. 2, p. 225–256, May 2014. ISSN 1573-0565.

SOLEIMANI, A.; ARAABI, B. N.; FOULADI, K. Deep multitask metric learning for offline signature verification. *Pattern Recognition Letters*, v. 80, p. 84 – 90, 2016. ISSN 0167-8655.

SOUZA, V. L. F.; OLIVEIRA, A. L. I.; CRUZ, R. M. O.; SABOURIN, R. Characterization of handwritten signature images in dissimilarity representation space. In: *2019 International Conference on Computational Science (ICCS)*. [S.l.]: Springer International Publishing, 2019. p. 192–206.

SOUZA, V. L. F.; OLIVEIRA, A. L. I.; CRUZ, R. M. O.; SABOURIN, R. On dissimilarity representation and transfer learning for offline handwritten signature verification. In: IEEE. *2019 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2019. p. 1–9.

SOUZA, V. L. F.; OLIVEIRA, A. L. I.; CRUZ, R. M. O.; SABOURIN, R. A white-box analysis on the writer-independent dichotomy transformation applied to offline handwritten signature verification. *Expert Systems with Applications*, v. 154, p. 113397, 2020. ISSN 0957-4174.

SOUZA, V. L. F.; OLIVEIRA, A. L. I.; SABOURIN, R. A writer-independent approach for offline signature verification using deep convolutional neural networks features. In: IEEE. *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)*. [S.l.], 2018. p. 212–217.

TRAN, B.; XUE, B.; ZHANG, M. Variable-length particle swarm optimization for feature selection on high-dimensional classification. *IEEE Transactions on Evolutionary Computation*, IEEE, v. 23, n. 3, p. 473–487, 2018.

TRAN, B.; ZHANG, M.; XUE, B. A pso based hybrid feature selection algorithm for high-dimensional classification. In: IEEE. *2016 IEEE congress on evolutionary computation (CEC)*. [S.l.], 2016. p. 3801–3808.

Triguero, I.; Derrac, J.; Garcia, S.; Herrera, F. A taxonomy and experimental study on prototype generation for nearest neighbor classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, v. 42, n. 1, p. 86–100, 2012.

TSOUROUNIS, D.; THEODORAKOPOULOS, I.; ZOIS, E. N.; ECONOMOU, G.; FOTOPOULOS, S. Handwritten signature verification via deep sparse coding architecture. In: *2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. [S.l.: s.n.], 2018. p. 1–5.

UNLER, A.; MURAT, A. A discrete particle swarm optimization method for feature selection in binary classification problems. *European Journal of Operational Research*, Elsevier, v. 206, n. 3, p. 528–539, 2010.

VARGAS-BONILLA, J.; FERRER, M.; TRAVIESO, C.; ALONSO, J. Off-line handwritten signature gpds-960 corpus. In: . [S.l.: s.n.], 2007. v. 2, p. 764–768.

VARGAS, J.; FERRER, M.; TRAVIESO, C.; ALONSO, J. B. Off-line signature verification based on grey level information using texture features. *Pattern Recognition*, Elsevier, v. 44, n. 2, p. 375–385, 2011.

WALMSLEY, F. N.; CAVALCANTI, G. D.; OLIVEIRA, D. V.; CRUZ, R. M.; SABOURIN, R. An ensemble generation method based on instance hardness. In: *2018 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2018. p. 1–8. ISSN 2161-4407.

WEN, J.; FANG, B.; TANG, Y. Y.; ZHANG, T. Model-based signature verification with rotation invariant features. *Pattern Recognition*, Elsevier, v. 42, n. 7, p. 1458–1466, 2009.

XUE, B.; ZHANG, M.; BROWNE, W. N.; YAO, X. A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, IEEE, v. 20, n. 4, p. 606–626, 2015.

YILMAZ, M. B.; OZTURK, K. Hybrid user-independent and user-dependent offline signature verification with a two-channel cnn. In: IEEE. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. [S.l.], 2018. p. 639–6398.

YILMAZ, M. B.; YANIKOĞLU, B. Score level fusion of classifiers in off-line signature verification. *Information Fusion*, Elsevier, v. 32, p. 109–119, 2016. ISSN 1566-2535.

ZHANG, Y.; XIONG, X.; ZHANG, Q. An improved self-adaptive pso algorithm with detection function for multimodal function optimization problems. *Mathematical Problems in Engineering*, vol. 2013, p. 8 pages, 2013.

ZHANG, Z.; LIU, X.; CUI, Y. Multi-phase offline signature verification system using deep convolutional generative adversarial networks. In: IEEE. *Computational Intelligence and Design (ISCID), 2016 9th International Symposium on*. [S.l.], 2016. v. 2, p. 103–107.

ZOIS, E. N.; ALEWIJNSE, L.; ECONOMOU, G. Offline signature verification and quality characterization using poset-oriented grid features. *Pattern Recognition*, Elsevier, v. 54, p. 162–177, 2016.

ZOIS, E. N.; ALEXANDRIDIS, A.; ECONOMOU, G. Writer independent offline signature verification based on asymmetric pixel relations and unrelated training-testing datasets. *Expert Systems with Applications*, v. 125, p. 14 – 32, 2019. ISSN 0957-4174.

ZOIS, E. N.; PAPAGIANNOPOULOU, M.; TSOUROUNIS, D.; ECONOMOU, G. Hierarchical dictionary learning and sparse coding for static signature verification. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. [S.l.: s.n.], 2018.

ZOIS, E. N.; THEODORAKOPOULOS, I.; ECONOMOU, G. Offline handwritten signature modeling and verification based on archetypal analysis. In: *The IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2017.

ZOIS, E. N.; TSOUROUNIS, D.; THEODORAKOPOULOS, I.; KESIDIS, A. L.; ECONOMOU, G. A comprehensive study of sparse representation techniques for offline signature verification. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, v. 1, n. 1, p. 68–81, Jan 2019. ISSN 2637-6407.

ZOTTESSO, R. H.; COSTA, Y. M.; BERTOLINI, D.; OLIVEIRA, L. E. Bird species identification using spectrogram and dissimilarity approach. *Ecological Informatics*, v. 48, p. 187 – 197, 2018. ISSN 1574-9541.

APPENDIX A – STUDY ON "GOOD" AND "BAD" QUALITY SKILLED FORGERIES

In this section we present a complementary study on the Figure 20 of this thesis, including the instance hardness analysis. In addition, an image-level analysis is also carried out for an easier and better understanding of the scenarios. Figure 23 depicts the same behaviour as in Figure 20 , highlighting key instances, which are:

- Positive sample: Genuine signature
- Negative sample: Random forgery
- Negative sample: "Bad quality" skilled forgery
- Negative sample (correctly classified): "Good quality" skilled forgery
- Negative sample (wrongly classified): "Good quality" skilled forgery

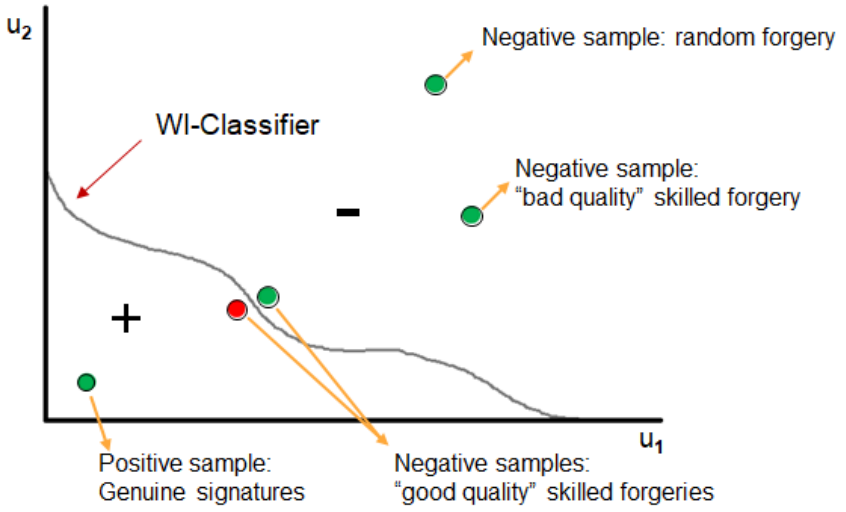

Fig. 23 – Synthetic decision frontiers. The same as in Figure 20 of this thesis.

As presented in Figure 23, while the negative region of space is located on the right of the decision boundary, the positive region is located on the left side. So, all correctly classified instances are colored in green, the wrongly classified one is presented in red.

Figures 24, 25, 26, 27 and 28 present, respectively, the behavior of genuine signature, random forgery, "bad quality" skilled forgery, "good quality" skilled forgery (correctly classified) and "good quality" skilled forgery (wrongly classified) at the image level for the GPDS dataset (VARGAS-BONILLA et al., 2007).

In all these figures, while left side presents the tested sample, the right side (in purple) contains the neighborhood of the tested sample in the training set (same methodology as

in the original article). Recall, to obtain a dissimilarity vector we need a reference signature and a query. That is the reason why in each sample two signatures are presented. Each sample, also contains the index of the writer and the index of the signature.

On the top of each figure, the instance hardness value related to the tested sample is presented. Also, in the lower left corner of each figure, the location of the respective instance in Figure 23 is depicted.

Fig. 24 – A positive tested sample on the left and its neighborhood on the right.

As depicted in Figure 24, the tested positive sample is formed by two genuine signatures from the same writer (index 349). As can be seen, all instances of the neighborhood belong to the positive class, since both signatures used to obtain the dissimilarity vector are from the same writer. So, the $IH = 0.0$. Finally, as both references and queries are formed by similar signatures all these dissimilarity vectors are located close to the origin (as highlighted in lower left corner of the figure).

As depicted in Figure 25, the tested negative sample is formed by two signatures from different writers (index 349 and 481), which clearly have different formats. The same behavior can also be seen in all instances that belong to the neighborhood. As all neighbors belong to the negative class, then $IH = 0.0$. Finally, as both references and queries are formed by signatures from different writers and have different formats all these dissimilarity vectors are located far from the origin (as highlighted in lower left corner of the figure).

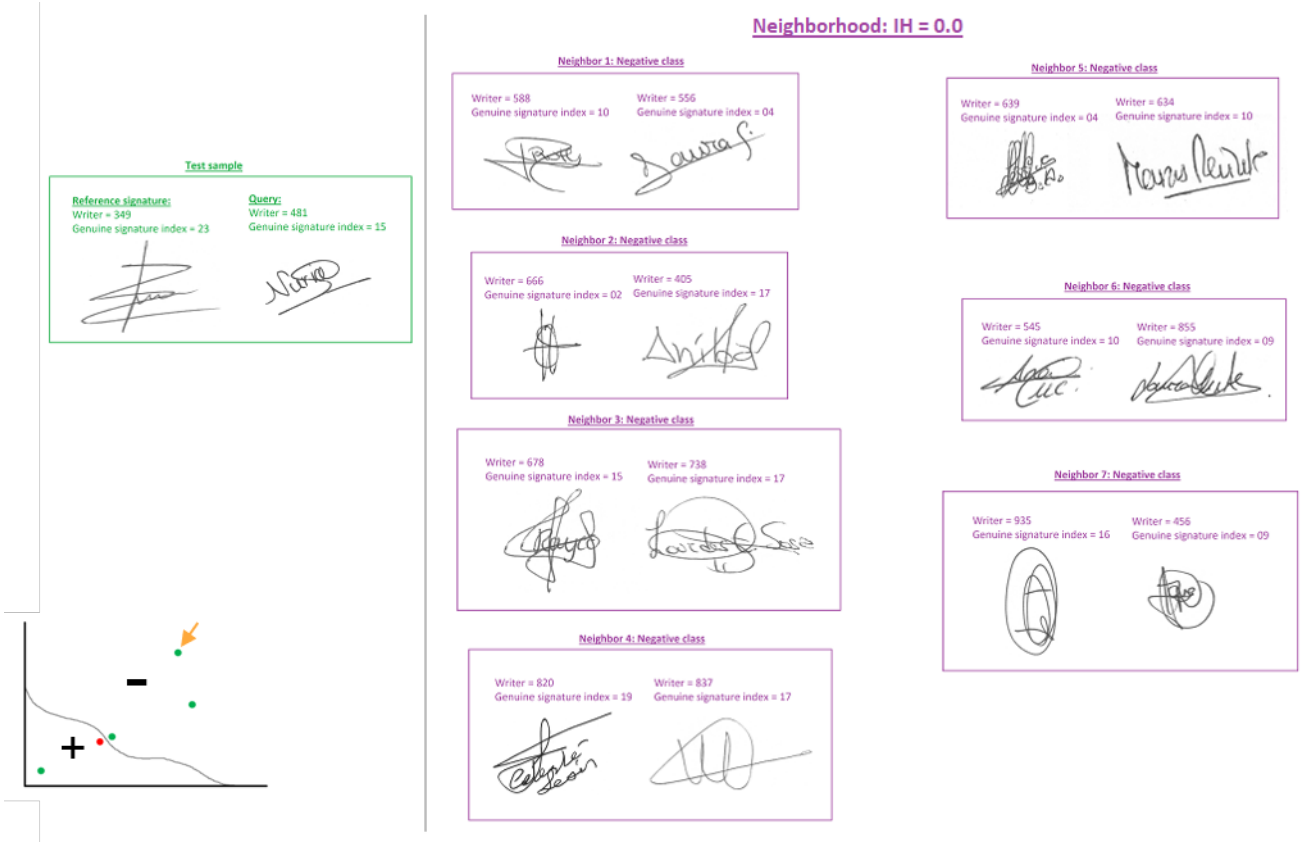

Fig. 25 – A negative tested sample (random forgery) on the left and its neighborhood on the right.

As depicted in Figure 26, the tested negative sample is formed by genuine signature as reference and a "bad quality" skilled forgery. It can be cleary seen that the forgery is not good. Thus, the "bad quality" skilled forgery behaves similarly to a random forgery in the dissimilarity space. So that, as can be seen, all neighbors belong to the negative class and are formed by dissimilarity vectors formed by signatures of different writers and different formats ($IH = 0.0$). As highlighted in lower left corner of the figure, it is located far from the origin.

As presented in Figures 27 and 28, a "good quality" skilled forgery actually looks like the genuine signature. Thus, the "good quality" skilled forgery behaves similarly to a genuine signature in the dissimilarity space. So that, as can be seen, all neighbors belong to the positive class and are formed by dissimilarity vectors formed by signatures from the same writers ($IH = 1.0$).

The fact of being located close to the WI decision boundary results in hard to classify instances. Consequently, correct (Figure 27) and wrong (Figure 28) classification may occur for this type of test samples.

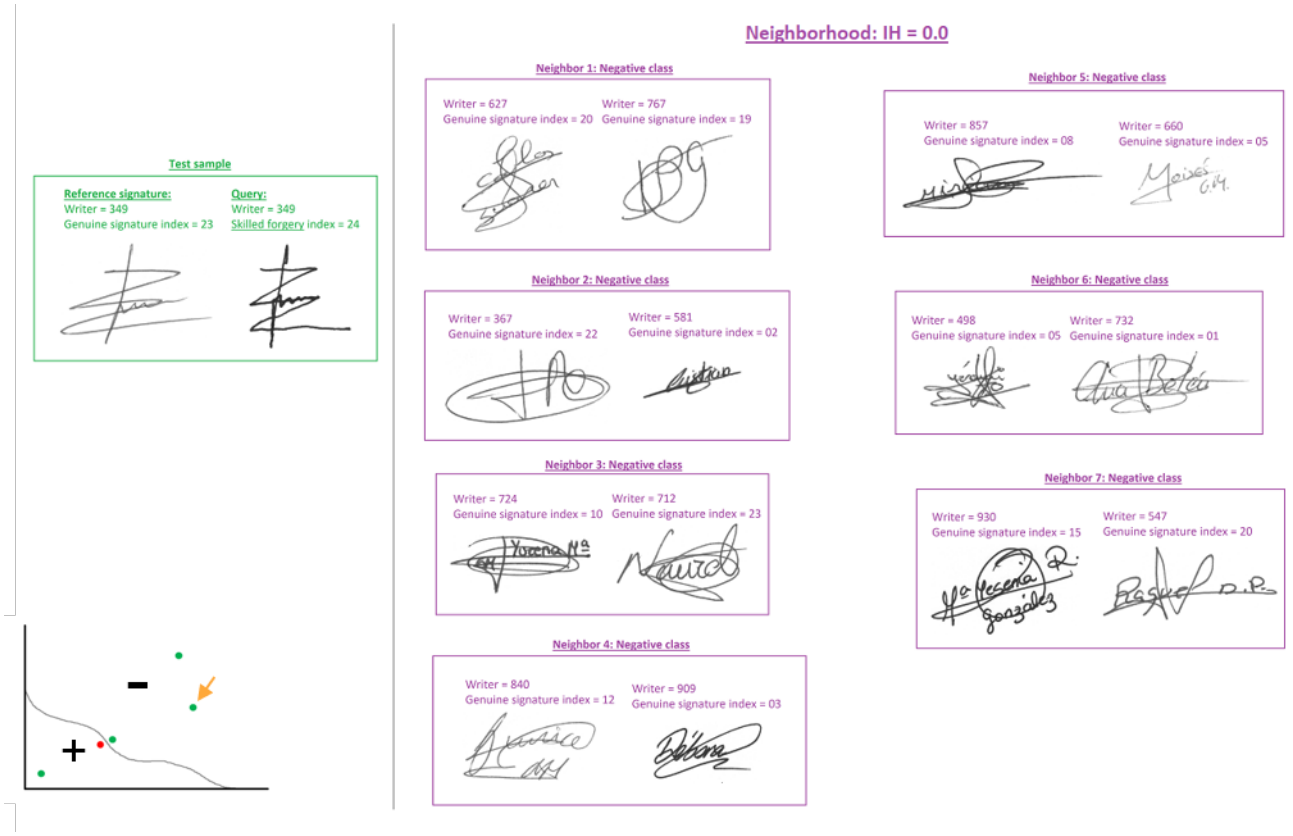

Fig. 26 – A negative tested sample ("Bad quality" skilled forgery) on the left and its neighborhood on the right.

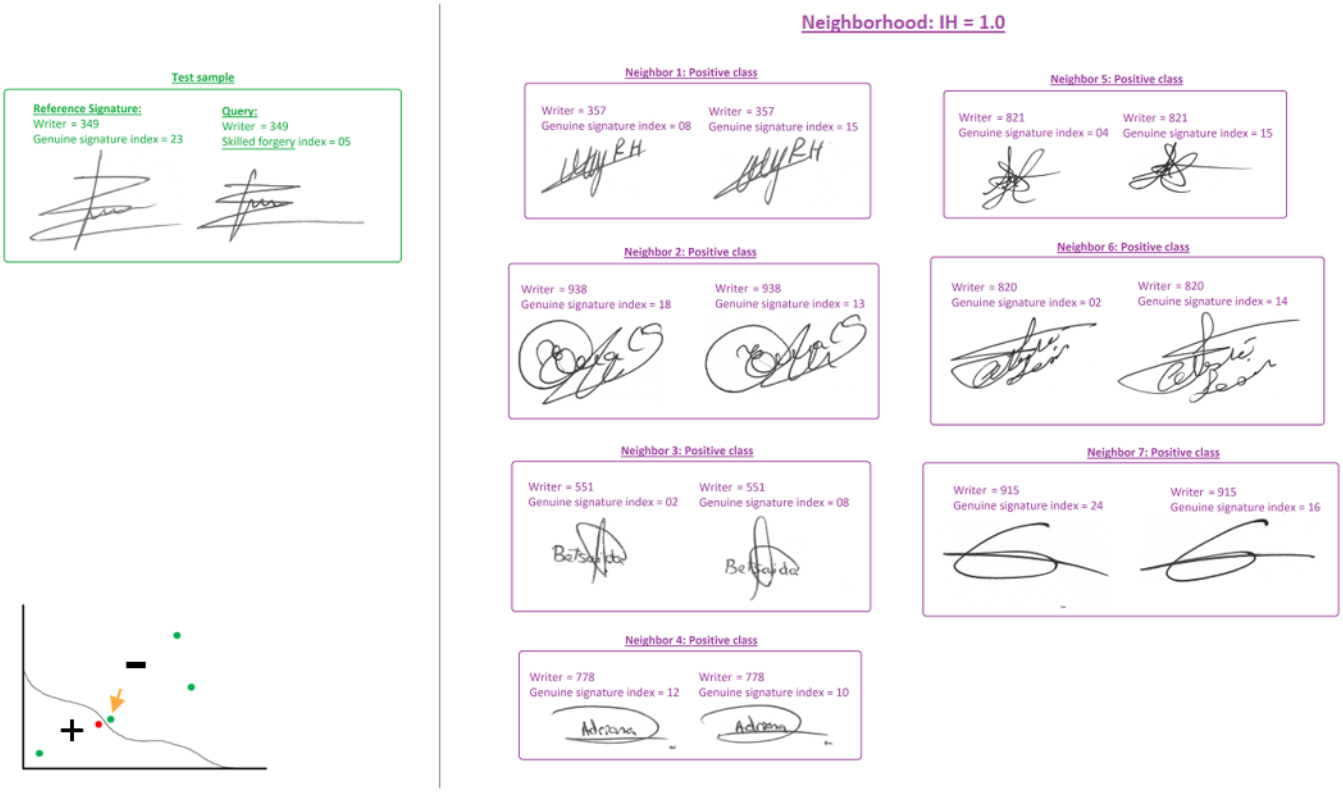

Fig. 27 – A negative tested sample correctly classified ("Good quality" skilled forgery) on the left and its neighborhood on the right.

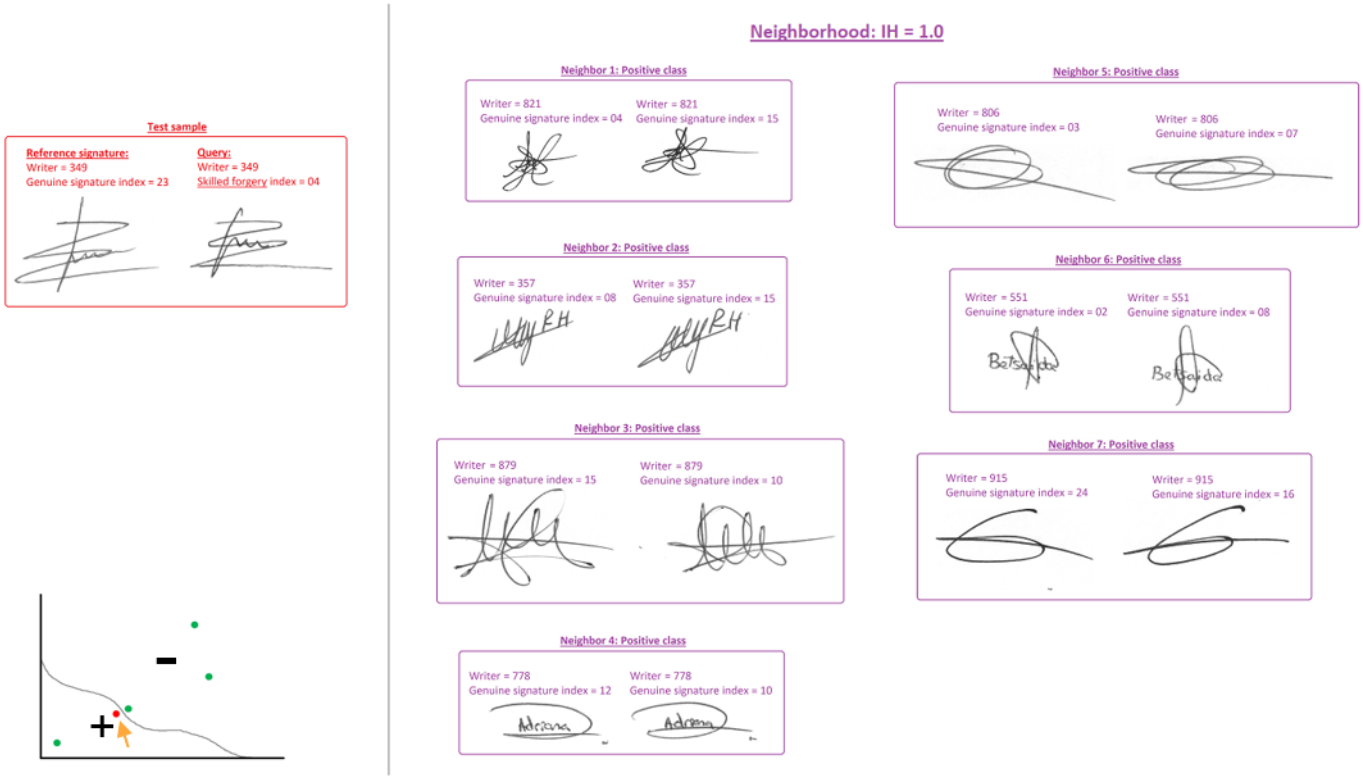

Fig. 28 – A negative tested sample wrongly classified ("Good quality" skilled forgery) on the left and its neighborhood on the right.

APPENDIX B – BRAZILIAN RESULTS

Tables 48, 49, 50 and 51 present the relationship of IH and the accuracy (%) of the model when the user threshold of EER is used as decision threshold, respectively for the positive samples, negative samples from the random forgeries, negative samples from the simple forgeries and negative samples from the skilled forgeriess, for the BRAZILIAN dataset. In the tables, the first column represents the possible IH values ($K = 7$), in the second column the number of samples for the respective IH value. The other columns represent the accuracy (%) when considering the CNN-SVM trained using the training set of GPDS-300 dataset and using respectively one (R1), five ($R5_{max}$), fifteen ($R15_{max}$) and thirty references ($R30_{max}$).

As in the other datasets, positive samples form a dense cluster (almost all positive samples have $IH \leq 0.14$), and the negative samples are scattered throughout space. The negative (random) samples may be disjoint to the positive set. The negative samples formed by the "good quality skilled forgeries" overlap the positive region of the DS, resulting in the need for a classifier with complex decision boundary.

Table 48 – Relationship between IH and accuracy (%) for the **positive samples**, for the BRAZILIAN dataset

IH	#Samples	R1	$R5_{max}$	$R15_{max}$	$R30_{max}$
0.00	**591**	**96.27**	**97.63**	**99.49**	**99.66**
0.14	7	57.14	71.42	71.42	85.71
0.28	1	100.00	100.00	100.00	100.00
0.42	0	-	-	-	-
0.57	1	0.00	0.00	100.00	100.00
0.71	0	-	-	-	-
0.85	0	-	-	-	-
1.00	0	-	-	-	-

Table 49 – Relationship between IH and accuracy (%) for the **negative samples from the random forgeries**, for the BRAZILIAN dataset

IH	#Samples	R1	$R5_{max}$	$R15_{max}$	$R30_{max}$
0.00	79	100.00	100.00	100.00	100.00
0.14	122	100.00	100.00	100.00	100.00
0.28	96	100.00	100.00	100.00	100.00
0.42	72	100.00	100.00	100.00	100.00
0.57	55	100.00	100.00	100.00	100.00
0.71	47	100.00	100.00	100.00	100.00
0.85	52	100.00	100.00	100.00	100.00
1.00	77	100.00	100.00	100.00	100.00

Table 50 – Relationship between IH and accuracy (%) for the **negative samples from the simple forgeries**, for the BRAZILIAN dataset

IH	#Samples	R1	$R5_{max}$	$R15_{max}$	$R30_{max}$
0.00	69	100.00	100.00	100.00	100.00
0.14	98	100.00	100.00	100.00	100.00
0.28	69	100.00	100.00	100.00	100.00
0.42	45	100.00	100.00	100.00	100.00
0.57	62	100.00	100.00	100.00	100.00
0.71	69	100.00	100.00	100.00	100.00
0.85	76	100.00	100.00	100.00	100.00
1.00	112	98.21	100.00	100.00	100.00

Table 51 – Relationship between IH and accuracy (%) for the **negative samples from the skilled forgeries**, for the BRAZILIAN dataset

IH	#Samples	R1	$R5_{max}$	$R15_{max}$	$R30_{max}$
0.00	5	100.00	100.00	100.00	100.00
0.14	9	100.00	100.00	100.00	100.00
0.28	29	100.00	100.00	100.00	100.00
0.42	23	100.00	100.00	100.00	100.00
0.57	39	100.00	100.00	100.00	100.00
0.71	63	100.00	100.00	100.00	100.00
0.85	115	100.00	100.00	100.00	100.00
1.00	**317**	**93.05**	**95.89**	**98.42**	**98.73**

APPENDIX C – CEDAR RESULTS

Tables 52, 53 and 54 present the relationship of IH and the accuracy (%) of the model when the user threshold of EER is used as decision threshold, respectively for the positive samples, negative samples from the random forgeries and negative samples from the skilled forgeries, for the CEDAR dataset. In the tables, the first column represents the possible IH values ($K = 7$), in the second column the number of samples for the respective IH value. The other columns represent the accuracy (%) when considering the CNN-SVM trained using the training set of GPDS-300 dataset and using respectively one (R1), five ($R5_{max}$) and twelve ($R12_{max}$) references.

As in the other datasets, positive samples form a dense cluster (almost all positive samples have $IH \leq 0.14$), and the negative samples are scattered throughout space. The negative (random) samples may be disjoint to the positive set. The negative samples formed by the "good quality skilled forgeries" overlap the positive region of the DS, resulting in the need for a classifier with complex decision boundary.

Table 52 – Relationship between IH and accuracy (%) for the **positive samples**, for the CEDAR dataset

IH	#Samples	R1	$R5_{max}$	$R12_{max}$
0.00	**482**	**92.53**	**94.60**	**95.85**
0.14	60	80.00	93.33	95.00
0.28	1	0.00	0.00	100.00
0.42	0	-	-	-
0.57	0	-	-	-
0.71	2	50.00	100.00	100.00
0.85	1	0.00	100.00	100.00
1.00	4	0.00	100.00	100.00

Table 53 – Relationship between IH and accuracy (%) for the **negative samples from the random forgeries**, for the CEDAR dataset

IH	#Samples	R1	$R5_{max}$	$R12_{max}$
0.00	11	100.00	100.00	100.00
0.14	68	95.58	100.00	100.00
0.28	89	95.50	100.00	100.00
0.42	58	98.27	100.00	100.00
0.57	69	100.00	100.00	100.00
0.71	79	100.00	100.00	100.00
0.85	110	100.00	100.00	100.00
1.00	66	100.00	100.00	100.00

Table 54 – Relationship between IH and accuracy (%) for the **negative samples from the skilled forgeries**, for the CEDAR dataset

IH	#Samples	R1	$R5_{max}$	$R12_{max}$
0.00	1	100.00	100.00	100.00
0.14	15	80.00	100.00	100.00
0.28	16	100.00	100.00	100.00
0.42	18	100.00	100.00	100.00
0.57	29	100.00	100.00	100.00
0.71	35	97.14	97.14	100.00
0.85	147	96.59	95.91	97.95
1.00	**289**	**87.54**	**93.42**	**94.46**

APPENDIX D – CEDAR RESULTS OPTIMIZED SPACE

Tables 55, 56 and 57 present the relationship of the IH and the accuracy (%) of the model when the user threshold of EER is used as decision threshold, respectively for the positive samples, negative samples from the random forgeries and negative samples from the skilled forgeries (for the CEDAR dataset).

The CNN-SVM trained in the training set of GPDS-300 dataset the model responsible for carrying out the verification. In the tables, the first column lists the IH values ($K = 7$), the second and third columns contain respectively the number of samples and the accuracy (%) when considering the model using twelve reference signatures, for each IH value in the original space. The fourth and fifth columns, the same information in the optimized space (using the BPSO method as in Section 4.5).

As can be seen from the presented tables, even with different number of features in each space, in general, for the CEDAR dataset, there is no significant difference between the frequencies of the number of samples for each IH value, for both spaces. This analysis can be observed for positive, negative (random) or for negative (skilled) samples. This behavior is similar to that presented in other databases

Table 55 – Relationship between IH and accuracy (%) for the **positive samples**, for the CEDAR dataset

	Original Space		**Optimized Space**	
IH	#Samples	$R12_{max}$	#Samples	$R12_{max}$
0.00	482	95.85	499	96.79
0.14	60	95.00	41	85.37
0.28	1	100.00	2	50.00
0.42	0	-	0	-
0.57	0	-	1	100.00
0.71	2	100.00	0	-
0.85	1	100.00	5	100.00
1.00	4	100.00	2	100.00

Table 56 – Relationship between IH and accuracy (%) for the **negative samples from the random forgeries**, for the CEDAR dataset

	Original Space		**Optimized Space**	
IH	#Samples	$R12_{max}$	#Samples	$R12_{max}$
0.00	11	100.00	55	100.00
0.14	68	100.00	71	100.00
0.28	89	100.00	67	100.00
0.42	58	100.00	67	100.00
0.57	69	100.00	73	100.00
0.71	79	100.00	93	100.00
0.85	110	100.00	51	100.00
1.00	66	100.00	43	100.00

Table 57 – Relationship between IH and accuracy (%) for the **negative samples from the skilled forgeries**, for the CEDAR dataset

	Original Space		**Optimized Space**	
IH	#Samples	$R12_{max}$	#Samples	$R12_{max}$
0.00	1	100.00	9	100.00
0.14	15	100.00	13	100.00
0.28	16	100.00	20	100.00
0.42	18	100.00	21	100.00
0.57	29	100.00	25	100.00
0.71	35	100.00	61	100.00
0.85	147	97.95	130	98.46
1.00	289	94.46	271	95.44