Learning and Nonlinear Models - Journal of the Brazilian Society on Computational Intelligence (SBIC), Vol. 18, Iss. 2, pp. 56-65, 2020

A STUDY OF THE INFLUENCE OF DATA COMPLEXITY AND SIMILARITY ON SOFT BIOMETRICS CLASSIFICATION PERFORMANCE IN A TRANSFER LEARNING SCENARIO

Marcelo Romero*, Matheus Gutoski*, Leandro Takeshi Hattori*, Manassés Ribeiro, Heitor Silvério Lopes***

*Federal University of Technology Paraná, Curitiba, Brazil

**Catarinense Federal Institute of Education, Science and Technology, Videira, Brazil

nmarceloromero@gmail.com, matheusgutoski@gmail.com, leandrotakeshihattori@gmail.com

manasses.ribeiro@ifc.edu.br, hslopes@utfpr.edu.br

Abstract – Transfer learning is a paradigm that consists in training and testing classifiers with datasets drawn from distinct distributions. This technique allows to solve a particular problem using a model that was trained for another purpose. In the recent years, this practice has become very popular due to the increase of public available pre-trained models that can be fine-tuned to be applied in different scenarios. However, the relationship between the datasets used for training the model and the test data is usually not addressed, specially where the fine-tuning process is done only for the fully connected layers of a Convolutional Neural Network with pre-trained weights. This work presents a study regarding the relationship between the datasets used in a transfer learning process in terms of the performance achieved by models complexities and similarities. For this purpose, we fine-tune the final layer of Convolutional Neural Networks with pre-trained weights using diverse soft biometrics datasets. An evaluation of the performances of the models when tested with datasets that are different from the one used for training the model is presented. Complexity and similarity metrics are also used to perform the evaluation.

Keywords – Neural Network, Convolutional Neural Network, Transfer Learning, Soft Biometrics, Data Complexity, Data Similarity.

1 Introduction

Transfer learning is a learning paradigm based on employing knowledge that was previously acquired to solve new problems. Opposite to traditional machine learning algorithms, which assume that the data used to initially obtain the knowledge lies within the same feature space or share the same distribution of the test data [1], a transfer learning approach considers that the domains of the training and testing data may be different [2]. In this sense, the domains, tasks and distributions used in training and testing datasets to be distinct [3]. Although the domains of the datasets may differ, transfer learning applications are able to sufficiently solve problems that are similar. For instance, a model capable of classifying different types of images of guitars should be also able to classify basses. This allows to overcome a very common problem in real-world applications: the lack of available labelled data.

Among various application fields, image classification is a widely tackled problem through transfer learning. Several works address the use of pre-trained Convolutional Neural Networks (CNNs) [4] to perform this task. CNNs are neural networks that allow to deal with raw high-dimensional input data, such as images or videos; they learn hierarchies of features that allow to represent the original data in a more abstract manner. CNNs have two main parts: the feature extractor, composed of Convolutional and Pooling layers; and the classifier, composed of Fully Connected layers. This structure permits the network to learn both, feature extractor and classifier, during the training. Hence, they are end-to-end classifiers. This allows to develop feature extractors able to extract features that are specific for treating the data that was used to train the classifier avoiding the need for hand-crafted feature extractors.

Considering the transfer learning context, once the feature extractor of the CNN has been trained using a dataset containing a large number of instances, it is common to fine-tune the classifier segment of the network using a much smaller dataset, usually for a specific domain in which labelled data is more difficult to obtain. Recent work addressing this subject are based on the use the Inception-v3 model [5] pre-trained with the ImageNet dataset [6]. For instance, in [7], non-popular dog breeds were classified using a fine-tuned Inception-v3 network. The approach achieved 96% of accuracy on a custom dataset composed of images obtained on the internet. The work presented by [8] also applied a fine-tuning process of the Inception-v3 to classify images of Arabidopsis and Tobacco plants. The method achieved overall accuracy of 98%. The same strategy was used in [9] to classify species of flowers, achieving performances of 95% and 94%, in terms of accuracy, for the datasets used in the work. Not all works are focused on the Inception-v3 model, such as [10], in which a CNN was trained to classify videos of the Sports-1M dataset [10] and tested with the UCF-101 [11] Activity Recognition dataset. Results of the work showed that retraining the top few layers of the network achieved satisfactory results, whilst training the entire CNN from scratch led the network to a massive overfitting. On the other hand, in [12] the parameters of a CNN (with the architecture of [13] and trained with the

ImageNet dataset) were transferred into another model which had its final fully connected layers trained for a specific dataset. The method yielded state-of-the-art results on challenging benchmark datasets.

In spite of the large use of transfer learning to solve problems within different application domains, such as the works described before. Most works do not study the factors that may have influence the performance of models that are trained with data from different distributions, besides the proper architecture of the models. Moreover, existing studies are based on fine-tuning a pre-trained model to perform a specific task instead of considering the use of different datasets for training and testing. To the best of our knowledge, no work has addressed this aspect with the objective to analyse factors such as complexity or similarity of the datasets used for transfer learning. To overcome this gap, this work aims to evaluate the results obtained by a model that was trained and tested with datasets with different levels of complexity and similarity. We aim at analysing the possible impact of these factors on the classification performance of neural networks. To perform this task, we use small soft biometrics datasets [14], one of which is introduced in this work: the UTFPR-SBD2 dataset. Soft biometrics are physiological or behavioural characteristics from humans that provide information useful to differentiate one individual from another. Although soft biometrics are usually not unique for each individual, they can provide some prior information about the subjects that can lead to identification of a subject. Indeed, using just one soft biometric may not be a suitable option to identify a particular individual. However, used in combination they can lead to achieve satisfactory results to improve a recognition task under highly variable conditions. Soft biometrics can also be used to complement other primary biometric identifiers such as fingerprints and faces.

The main contributions of this work are: (1) the introduction of a novel soft biometrics dataset; (2) the application of a transfer learning approach to perform soft biometrics classification; (3) an evaluation of the relationship between the complexity of the datasets and the performance obtained by the models; (4) a study of the influence of the similarities between datasets when performing transfer learning.

This paper is organised as follows: Section 2 presents the methodology of this work and describes details of the architecture of the neural network used to perform the experiments. Section 3 present the metrics used for the classification task, and the complexity and similarity metrics used to analyse the datasets. Section 4 reports the experiments and the results obtained. Finally, Section 5 presents the conclusions and future work.

2 Methodology

This work aims at analysing the performance obtained by Neural Network models when trained and tested with different datasets that have diverse complexities and similarities. For this purpose, we use image datasets containing pedestrians and the objective is to classify soft biometrics attributes. Figure 1 presents an overview of our method for training and evaluating a model k. We use a Business Process Model and Notation (BPMN) diagram [15] to describe the process.

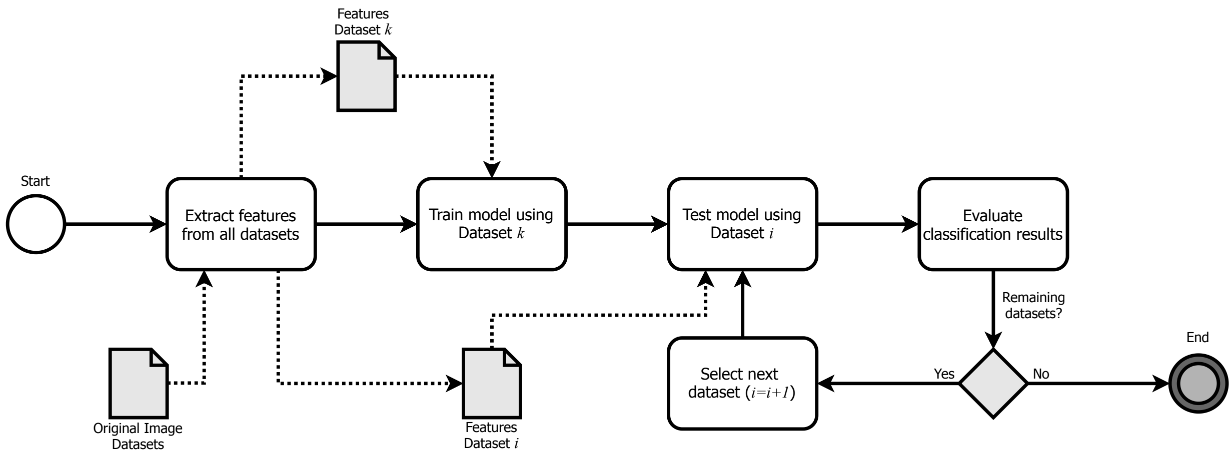


Figure 1: Methodology proposed in this work. The performance of a model trained with a dataset k is evaluated using N other datasets. The features are extracted from all datasets using the Inception-v3 model, with pre-trained weights using the ImageNet dataset.

The process starts with the extraction of the features from all the available datasets. For this purpose, we use the Inception-v3 architecture [5] with pre-trained weights using instances of the ImageNet dataset [6], which contains 100000 images of 1000 classes. After, a dataset k to be used for training is selected. A neural network composed of fully connected layers is then trained using the dataset k and tested using the remaining datasets. The classification results are evaluated after each test step using a dataset i, given $i \in N$, for a set of N datasets, for such purpose, we consider not only classification metrics to assess results but also aspects such as dataset complexity and the similarity between the datasets. Details regarding the Inception-v3 architecture and the training parameters used in this work are presented in Section 2.1. On the other hand, the complexities and similarities of the datasets, as well as the classification results, are measured through the metrics presented in Section 3.

2.1 Classification using Inception-v3

The Inception-v3 architecture is a CNN with Convolutional and Poling layers at the first segment of its structure. Local Response Normalization (LRN) [13] operations are also used in this first part of the model. The next part contains sequentially stacked Inception modules introduced in [16]. The concrete structure of the Inception modules used in Inception-v3 are described in [5]. The output of the last Inception module is fed to a Fully Connected layer that acts as the classifier and has its output filtered by a Softmax layer to produce the final classification.

In this work, we use a pre-trained version of the Inception-v3 and retrain only the Fully Connected layer using a soft biometrics dataset. The trained model is then tested using other datasets, as presented in Figure 1. The final layer of the network was trained with the Stochastic Gradient Descent (SGD) [17] optimisation method and Cross-Entropy as cost function, presented in Equation 1, in which L_i is the loss of the i-th pattern fed to the network considering n possible classes, $t_{i,j}$ is the ground-truth probability of the class j for the pattern i, and $p_{i,j}$ is the probability that the pattern i belongs to the class j predicted by the model.

$$L_i = -\sum_{j}^{n} t_{i,j} \log(p_{i,j}) \tag{1}$$

3 Evaluation Metrics

This Section presents the metrics used to evaluate the classification performance of the models, the complexity of each dataset and the similarity between the datasets.

3.1 Classification Metrics

The classification performances of the models are evaluated through Receiver Operating Characteristics (ROC) charts [18], which allow to visually analyse the performance of binary classifiers. ROC graphs represent the trade-off between True Positive Rates (TPR) and True Negative Rates (TNR) achieved by a classifier. If the classifier is discrete (it outputs only a class label), the chart contains a pair corresponding to a single point in ROC space. If the classifier yields a score or a probability value, a threshold can be used to produce a discrete classifier: the output of the classifier correspond to one class if it is above the threshold and another class if it is below. Each threshold value produces a different point in ROC space. Therefore, a ROC curve is produced for this type of classifier. We mainly consider the Area Under the ROC Curve (AUC) in this work, which is the area covered by a curve in the ROC space. The AUC value ranges between 0 and 1. If it is equal to 1, the classifier achieved the best possible performance.

3.2 Complexity Metrics

Besides the performances of the models, we also evaluate the complexity of the datasets. The first metric is the Kolmogorov complexity $C(x)$, which is a compression-based measure [19] defined by the length of the shortest computer program p generating a binary string x using a given description language L on an universal Turing machine U [20]. It is presented in Equation 2, in which $|p|$ is the length of the program p.

$$C(x) = \min_{p}\{|p| : U(p) = x\} \tag{2}$$

The complexity of a set of images can be calculated through a Normalised Compression Rate (NCR) based on the Kolmogorov complexity. Although $C(x)$ is non-computable [20], it is possible to obtain an approximation using a real-world compressor [21], such as the *bzip2* algorithm [22], which has been used in this work. Hence, we define NCR for a set of images X with Equation 3, in which C is the compressor, $s(\cdot)$ defines the size of its input, $C_0(X)$ and $C_{\max}(X)$ represent, the no-compression and maximum compression of X, respectively.

$$NCR(X) = 1 - \frac{s(C_0(X)) \ - \ s(C_{\max}(X))}{s(C_0(X))} \tag{3}$$

We also consider two other complexity metrics: Entropy and Spatial Information (SI). Since both metrics measure the complexity of a single image, we calculate their means to obtain the complexity of a dataset.

The entropy of a grey-scale image is defined by Equation 4, in which K is the number of grey levels and p_k is the probability related to the grey level k.

$$Entropy = -\sum_{k=1}^{K} p_k \log_2 p_k \tag{4}$$

The SI represents edge energy. It is calculated by first applying horizontal and vertical Sobel kernels on a grey-scale image, obtaining Sh and Sv filtered images, respectively. The mean SI then calculated to explain the complexity of an entire image through Equation 5, in which N and M are the number of rows and columns of the image, respectively.

$$SI = \frac{1}{NM} \sum_{j=1}^{N} \sum_{i=1}^{M} \sqrt{Sh_{i,j}^2 + Sv_{i,j}^2} \tag{5}$$

3.3 Similarity Metric

This work also addresses the influence of the similarity between datasets when applying transfer learning. For this purpose we used the Structural Similarity Index (SSIM) [23], which takes the texture of two images into account to compare them. Let x and y denote two image patches extracted from the same spatial location from two images being compared; and let μ_x, σ_x^2 and σ_{xy} be the mean of x, the variance of x, and the covariance of x and y, respectively. SSIM is calculated through Equation 6, in which L denotes the dynamic range of the pixel-values, and k_1 and k_2 are two scalar constants.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + (k_1L)^2)(2\sigma_{xy} + (k_2L)^2)}{(\mu_x^2\mu_y^2 + (k_1L)^2)(\sigma_x^2\sigma_y^2 + (k_2L)^2)} \tag{6}$$

To measure the SSIM between two datasets, we compare all images from both datasets and calculate the mean. Since this evaluation process is computationally expensive, it is recommended only for small datasets, or for statistical significant sets of samples taken from bigger datasets.

4 Experiments and Results

This section presents the experiments that were performed and the results obtained. First, we describe the configuration of the system used to train and test the neural networks. After, we briefly describe the datasets used in this work. Then, we introduce an evaluation of the datasets in terms of their complexities and similarities. This initial analysis will allow to establish relationships between classification performance and the complexity and similarity of the datasets used for training and testing the models. Finally, we present the classification results obtained by the models considering three soft biometrics attributes: Gender, Lower Clothes, and Hat. Likewise, we also define a baseline that consists in training and testing the models using the same dataset. Finally, we provide a discussion addressing the results obtained in this work.

4.1 System Setup

In this work, a Nvidia Titan-XP GPU was used for training and testing the models. The GPU was built within a server with an Intel(R) Core(TM) i7-6700 @ 3.40GHz processor and 32Gb RAM memory, running Ubuntu 14.04.3 LTS. The Python library TensorFlow [24], version 1.2.1, was used to train and test the neural network. The following Python libraries were also used to evaluate the performance of the models and for image processing: OpenCV2, version 2.4.9, SciPy, version 0.19.1, Scikit Image, version 0.12.3, and Scikit Learn, version 0.18.2.

4.2 Datasets

Three datasets were used for testing our methodology: VHH, UTFPR-SBD1 and UTFPR-SBD2. The first one was introduced in [25]. It is composed of a mixture of samples from the datasets: ViPer [26], H3D [27], and HATdb [28]. We refer to this dataset as VHH in this work. It contains 300 images that were manually labelled. The dataset was created by combining other datasets in order to help improve the generalisation capability of the trained models [25]. Due to that it contains a combination of subsets of diverse distributions, this is the most complex dataset, as we formally demonstrate in Section 4.3. The UTFPR-SBD1 dataset is composed of 264 manually labelled images. Its samples are frames of videos containing individuals walking in front of a simple white background. We consider this dataset as the simplest. On the other hand, this work introduces the UTFPR-SBD2 dataset, which is a more complex version of UTFPR-SBD1 in terms of the number of individuals and samples. It contains 360 manually labelled videos containing individuals walking in four different angles: left-to-right, right-to-left, back-to-front and front-to-back. Figure 2 presents sample frames from this dataset considering these four walking angles. In this work, we use one frame of each video of UTFPR-SBD2 as our dataset.

The datasets were selected pursuing variety when considering their complexity. Although they provide several labels, the ones that were used for this work are: Gender, Lower Clothes and Hat. The UTFPR-SBD1 and UTFPR-SBD2 datasets are publicly available for download in a public repository[1].

[1] `http://labic.utfpr.edu.br/datasets/UTFPR-SBD.html`

a) Left-to-Right b) Right-to-Left c) Back-to-Front d) Front-to-Back

Figure 2: Sample frames from the UTFPR-SBD2 dataset. Each frame corresponds to one of the four walking angle of the individuals.

4.3 Datasets Evaluation

We evaluate the datasets in terms of their complexities using the metrics presented in Section 3.2. Table 1 presents the complexities of the datasets used in this work.

Table 1: Complexity of each dataset considering different metrics

	VHH	**UTFPR-SBD2**	**UTFPR-SBD1**
NCR	0.000122	0.000062	0.000066
Entropy	5.950083	5.472752	4.913418
SI	0.103205	0.070842	0.045732

Results show that the most complex dataset is VHH. UTFPR-SBD2 is the second most complex considering Entropy and SI. Although UTFPR-SBD1 achieved a higher NCR than UTFPR-SBD2, the gap between them is very small. Thus, we consider the other metrics as more relevant, for which UTFPR-SBD2 proves to be more complex than UTFPR-SBD1.

As for the similarities between the datasets, which are presented in Figure 3, UTFPR-SBD1 and UTFPR-SBD2 are the most similar ones with the highest SSIM value. On the other hand, there is little similarity between VHH and the other datasets. Thus, we consider VHH as the most complex and dissimilar dataset.

4.4 Experiments

Three attributes labelled for all datasets were used to analyse the influence of the complexities and similarities of the datasets on the results obtained by a model using transfer learning: Gender, Lower Clothes and Hat. All attributes are binary and mostly unbalanced. We also train and test the models using the same dataset in order to serve as baseline.

4.4.1 Baseline: Inception-v3 Only

In this section we present the results obtained by Inception-v3 fine-tuned with the same dataset that was used for testing. We used 10-fold cross-validation to perform the experiments. The objective of the experiments is to compare the performance of models trained and tested with data from the same dataset with the performance achieved by models that were trained and evaluated with data from different datasets. Table 2 presents the results for the cross-validation considering the mean AUC.

Learning and Nonlinear Models - Journal of the Brazilian Society on Computational Intelligence (SBIC), Vol. 18, Iss. 2, pp. 56-65, 2020

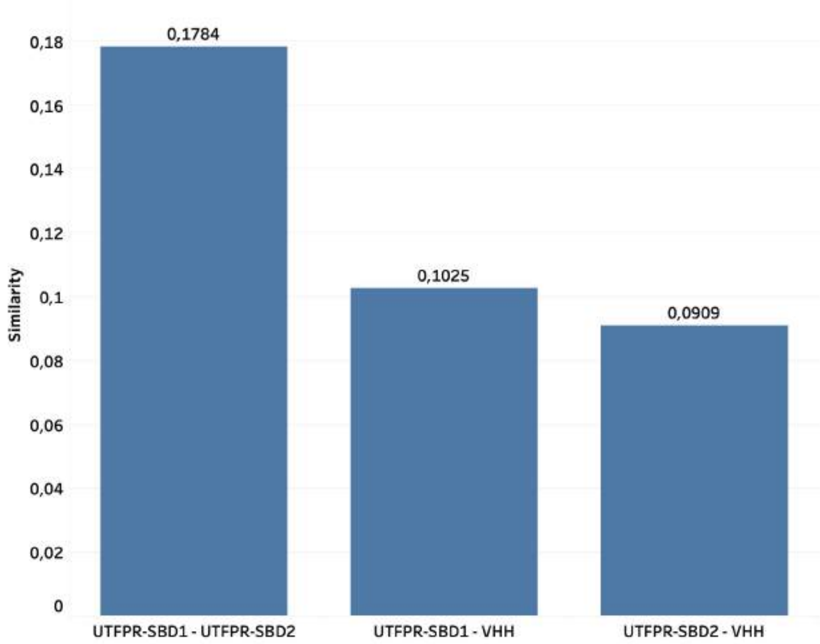


Figure 3: Structural Similarity Index (SSIM) between the datasets.

Table 2: Mean AUC and Standard Deviation obtained by the models for each dataset without transfer learning.

	UTFPR-SBD1	**UTFPR-SBD2**	**VHH**
Gender	0.99 ± 0.01	0.89 ± 0.07	0.72 ± 0.08
Lower Clothes	0.97 ± 0.04	0.99 ± 0.01	0.77 ± 0.07
Hat	0.89 ± 0.13	0.76 ± 0.11	0.59 ± 0.11

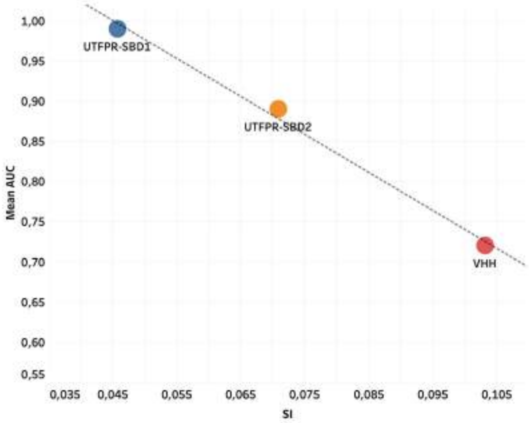
Results show a high mean AUC for the *Lower Clothes* attribute when using the datasets UTFPR-SBD1 and UTFPR-SBD2. The mean AUC for *Gender* is also high for UTFPR-SBD1, whilst it is around 0.90 for UTFPR-SBD2. For VHH, the mean AUC is lower, specially for the *Hat* attribute, for which the lowest mean AUC values were also obtained using the other two datasets. Figure 4 graphically shows this tendency. The plot presents a linear regression trend represented by the blue line. In almost all cases, except for the attribute *Lower Clothes* considering the UTFPR-SBD1 and UTFPR-SBD2 datasets, there is an inverse relationship between the complexity (SI) of the dataset and the performance obtained by the model (mean AUC) for that dataset. The results confirms the initial hypothesis that the more complex the dataset the harder it is to classify soft biometrics attributes.

4.4.2 Experiment #1 - Gender

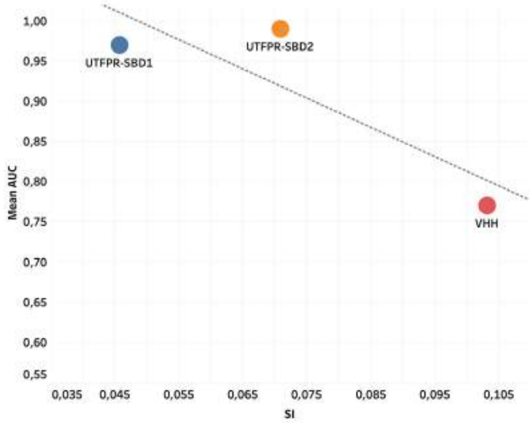
The performances obtained by the models for the attribute *Gender* when using transfer learning are presented in Table 3. Results show that the best performances were obtained by the models that were tested with the UTFPR-SBD1 dataset, the most simple one. Training the model with UTFPR-SBD2 dataset led to obtain the best result for UTFPR-SBD1 (AUC equal to 0.90). Therefore, the best result was achieved training the model with a more complex train set than the test set with both being the most similar ones. On the other hand, the worst results were obtained when the model was trained with the most simple dataset (UTFPR-SBD1) and when the similarity between the datasets was not very high (training with UTFPR-SBD1 and testing with VHH, training with VHH and testing with UTFPR-SBD2).

Table 3: AUC obtained using transfer learning with different datasets for the attribute *Gender*.

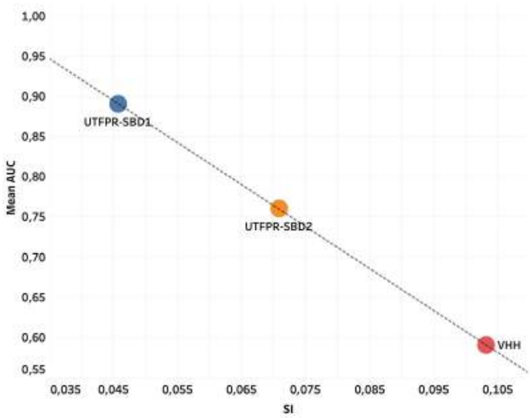
		Test		
		UTFPR-SBD1	**UTFPR-SBD2**	**VHH**
Train	**UTFPR-SBD1**	X	0.67	0.65
	UTFPR-SBD2	0.90	X	0.68
	VHH	0.85	0.65	X



a) Gender



b) Lower Clothes



c) Hat

Figure 4: Relationship between the performance (mean AUC) of the models and the complexities (SI) of the datasets.

No transfer learning approach surpassed the AUC baseline (obtained by models trained and tested with the same dataset). However, for the simplest dataset (UTFPR-SBD1), the approach based on training with UTFPR-SBD2 achieved 0.90 of AUC, 0.09 less than the mean value obtained when training and testing the model with the same dataset in cross-validation. Considering the most complex dataset (VHH), the result of the transfer learning when the model was trained with UTFPR-SBD2 was equal to 0.68, which is 0.04 less than the mean value obtained by the model trained and tested with VHH in cross-validation (0.72).

4.4.3 Experiment #2 - Lower Clothes

Considering the attribute *Lower Clothes*, results are similar to those obtained for the attribute Gender: the best models were those trained with more complex datasets and tested with the most simple one (UTFPR-SBD1). The best result was obtained for a model trained with VHH and tested with UTFPR-SBD2 (AUC= 0.92). This result shows that similarity may not have much relevance depending on the attribute that is being considered. For this attribute, a higher complexity may have a concrete influence since VHH is more complex than UTFPR-SBD2. As for the worst result, it was obtained when the model was trained with the simplest dataset and tested with the most complex one, similar to what was observed in experiment #1. The results for this attribute are presented in Table 4.

Table 4: AUC obtained using transfer learning with different datasets for the attribute *Lower Clothes*.

		Test		
		UTFPR-SBD1	**UTFPR-SBD2**	**VHH**
Train	**UTFPR-SBD1**	X	0.86	0.78
	UTFPR-SBD2	0.91	X	0.86
	VHH	0.89	0.92	X

For this attribute, the transfer learning approaches did not surpass the AUC achieved when training and testing the models with the same dataset. Notwithstanding, the results achieved by the model trained with VHH and tested with UTFPR-SBD2 are is worth mentioning, with an AUC equal to 0.92, whilst the mean AUC achieved by the model trained and tested with UTFPR-SBD2 was equal to 0.99 with a standard deviation equal to 0.01. We consider that this result was achieved due to the higher complexity of VHH with respect to UTFPR-SBD2 and the apparent simplicity to classify instances of datasets when considering the *Lower Clothes* attribute.

4.4.4 Experiment #3 - Hat

The best results obtained for the attribute *Hat*, presented in Table 5, are similar to those in Experiments #1 and #2: testing the model with UTFPR-SBD1 led to obtain the highest AUC values. The third best result was obtained when training the model with the least complex dataset (UTFPR-SBD1) and tested with the most complex (VHH). The other ROC curves show difficulty to classify images for this particular attribute, since the rest of the AUC values are under 0.60 and very close to a random guessing threshold equal to 0.50.

Table 5: AUC obtained using transfer learning with different datasets for the attribute *Lower Clothes*.

		Test		
		UTFPR-SBD1	**UTFPR-SBD2**	**VHH**
Train	**UTFPR-SBD1**	X	0.58	0.63
	UTFPR-SBD2	0.69	X	0.58
	VHH	0.84	0.55	X

The AUC achieved for the UTFPR-SBD1 dataset when the models were trained with other datasets did not surpassed the mean value achieved when using the same dataset for training and testing, although the value obtained when using VHH for training was equal to 0.84. On the other hand, for the most complex dataset (VHH), training with UTFPR-SBD1 led to obtain outstanding results compared to the mean AUC when training and testing with VHH in cross-validation: 0.63 versus 0.59 ± 0.11. This result implies that even simple datasets such as UTFPR-SBD1 may lead to achieve satisfactory results for complex ones when applying transfer learning. The causes of this phenomenon will be addressed in further studies, since it can be considered an exception taking into account the results from previous experiments.

4.4.5 Discussion

The results presented in this work suggest that a model tends to achieve satisfactory results when tested with a dataset less complex than that used for training. However, this may vary and would ultimately depend on the attribute to classify. In two of the three experiments performed in this work, the best result was achieved by a model trained with a more complex train set than the test set (experiments #1 and #2), and also with the highest similarity level between them. Therefore, similarity could be considered an useful metric, when combined with complexity, to guide the choice of a train dataset for a model when applying transfer learning. The inverse situation was also empirically demonstrated: training a model with a simpler train dataset led to achieve the worst results (as shown in Tables 3, 4 and 5). However, for some cases the simplest train dataset led to obtain better results than those achieved when using the other more complex train datasets, which was possibly due to that its similarity with the test dataset was higher (experiment #3). This behaviour reinforces the hypothesis that similarity could be a relevant factor to consider when transferring learning, even though it is a computationally expensive metric and it may not suitable for all real-world scenarios.

As for the evaluation of the models when using transfer learning when compared to an approach based on using the same dataset for both training and testing (baseline), results showed that there is no improvement with respect to the baseline, although some transfer learning approaches allowed to achieve close results (experiment #1). However, one experiment led to obtain a dramatic improvement (experiment #3). It was achieved when training the model with the simplest dataset (UTFPR-SBD1) and tested with the most complex one (VHH) for classifying the most difficult soft biometric attribute (Hat). Although this behaviour is punctual, further studies should be performed so as to study the factors that led to accomplish it in order to understand how to replicate it for other datasets or attributes.

5 Conclusion

This work presented a study regarding the influence of the complexity of a train dataset on the performance of an image classifier for a set of images drawn from a different distribution with distinct complexity. It also addressed the similarity between the datasets to evaluate the performance of the model. Our approach was tested using three datasets with different levels of complexity and diverse similarities between them. In this sense, we also introduced a novel dataset named UTFPR-SBD2, which is publicly available to foster future research addressing soft biometrics classification.

Future work will aim at evaluating different types of datasets with distinct labels. The quantity of datasets used may also be augmented to obtain more general results. For soft biometric datasets, distinct attributes could be considered aside from the ones used in this work. Data augmentation can also be used to modify the complexity of a train dataset or its similarity to the test dataset in order to evaluate the influence of both factors when they are dynamically adjusted. Further studies may also pursuit to study with more depth those the factors that influence the transferring of knowledge acquired with a dataset to another one. Although this work has presented a discussion of certain aspects such as similarity and complexity, there may be another ones that may have more or less influence in the process.

Acknowledgements

M. Romero thanks the Organisation of the American States, the Coimbra Group of Brazilian Universities and the Pan American Health Organization; M. Gutoski would like to thank CAPES for the scholarship; L. T. Hattori thanks CNPQ for the scholarship; H.S.Lopes would like to thank to CNPq for the research grant number 440977/2015-0. All authors would like to thank both CAPES and CNPq for the scholarships, as well as NVIDIA and Fundação Araucária. All authors would also like to thank Bruna De Paula and Lucas Albini for their work during the labelling process of the UTFPR-SBD2 dataset.

References

[1] L. Shao, F. Zhu and X. Li. "Transfer learning for visual categorization: A survey". *IEEE transactions on neural networks and learning systems*, vol. 26, no. 5, pp. 1019–1034, 2015.

[2] G. Fung, J. Yu, H. Lu and P. Yu. "Text classification without negative examples revisit". *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 6–20, 2006. cited By 112.

[3] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue and G. Zhang. "Transfer learning using computational intelligence: A survey". *Knowledge-Based Systems*, vol. 80, no. Supplement C, pp. 14 – 23, 2015. 25th anniversary of Knowledge-Based Systems.

[4] Y. LeCun and Y. Bengio. "The Handbook of Brain Theory and Neural Networks". chapter Convolutional Networks for Images, Speech, and Time Series, pp. 255–258. MIT Press, Cambridge, MA, USA, 1998.

[5] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna. "Rethinking the Inception Architecture for Computer Vision". In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei. "ImageNet Large Scale Visual Recognition Challenge". *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Dec 2015.

[7] P. Devikar. "Transfer Learning for Image Classification of various dog breeds". *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 5, no. 12, pp. 2707–2715, 2016.

[8] A. Tapas. "Transfer Learning for Image Classification and Plant Phenotyping". *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 5, no. 11, pp. 2664–2669, 2016.

[9] X. Xia, C. Xu and B. Nan. "Inception-v3 for flower classification". In *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, pp. 783–787, June 2017.

[10] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar and L. Fei-Fei. "Large-scale video classification with convolutional neural networks". In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 1725–1732, 2014.

[11] K. Soomro, A. R. Zamir and M. Shah. "UCF101: A dataset of 101 human actions classes from videos in the wild". *arXiv preprint arXiv:1212.0402*, 2012.

[12] M. Oquab, L. Bottou, I. Laptev and J. Sivic. "Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks". In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[13] A. Krizhevsky, I. Sutskever and G. E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, pp. 1097–1105. Curran Associates, Inc., 2012.

[14] P. Tome, J. Fierrez, R. Vera-Rodriguez and M. S. Nixon. "Soft Biometrics and Their Application in Person Recognition at a Distance". *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 3, pp. 464–475, March 2014.

[15] OMG. "Business Process Model and Notation (BPMN), Version 2.0", January 2011.

[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich. "Going deeper with convolutions". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[17] L. Bottou. *Large-Scale Machine Learning with Stochastic Gradient Descent*, pp. 177–186. Physica-Verlag HD, Heidelberg, 2010.

[18] T. Fawcett. "An introduction to ROC analysis". *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[19] H. Yu and S. Winkler. "Image complexity and spatial information". In *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, pp. 12–17, July 2013.

[20] M. Ribeiro, A. E. Lazzaretti and H. S. Lopes. "A study of deep convolutional auto-encoders for anomaly detection in videos". *Pattern Recognition Letters*, 2017.

[21] R. Cilibrasi and P. M. B. Vitanyi. "Clustering by compression". *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1523–1545, April 2005.

[22] J. Seward. "bzip2 and libbzip2, version 1.0. 5: A program and library for data compression". *URL http://www. bzip. org*, 2007.

[23] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli. "Image quality assessment: from error visibility to structural similarity". *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[24] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro and et al. "TensorFlow: large-scale machine learning on heterogeneous systems". *arXiv:1603.04467*, pp. 1–19, 2016.

[25] H. A. Perlin and H. S. Lopes. "Extracting human attributes using a convolutional neural network approach". *Pattern Recognition Letters*, vol. 68, pp. 250 – 259, 2015. Special Issue on Soft Biometrics.

[26] D. Gray and H. Tao. *Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features*, pp. 262–275. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[27] L. Bourdev, S. Maji and J. Malik. "Describing people: A poselet-based approach to attribute classification". In *2011 International Conference on Computer Vision*, pp. 1543–1550, Nov 2011.

[28] G. Sharma and F. Jurie. "Learning discriminative spatial representation for image classification". In *BMVC 2011 - British Machine Vision Conference*, edited by J. Hoey, S. J. McKenna and E. Trucco, pp. 1–11, Dundee, United Kingdom, August 2011. BMVA Press.