

Research Article

Gender and Handedness Prediction from Offline Handwriting Using Convolutional Neural Networks

Ángel Morera, Ángel Sánchez , José Francisco Vélez, and Ana Belén Moreno

Technical School of Computer Science, Rey Juan Carlos University, Móstoles, 28933 Madrid, Spain

Correspondence should be addressed to Ángel Sánchez; angel.sanchez@urjc.es

Received 21 July 2017; Revised 5 October 2017; Accepted 12 November 2017; Published 14 January 2018

Academic Editor: Jing Na

Copyright © 2018 Ángel Morera et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Demographic handwriting-based classification problems, such as gender and handedness categorizations, present interesting applications in disciplines like Forensic Biometrics. This work describes an experimental study on the suitability of deep neural networks to three automatic demographic problems: gender, handedness, and combined gender-and-handedness classifications, respectively. Our research was carried out on two public handwriting databases: the IAM dataset containing English texts and the KHATT one with Arabic texts. The considered problems present a high intrinsic difficulty when extracting specific relevant features for discriminating the involved subclasses. Our solution is based on convolutional neural networks since these models had proven better capabilities to extract good features when compared to hand-crafted ones. Our work also describes the first approach to the combined gender-and-handedness prediction, which has not been addressed before by other researchers. Moreover, the proposed solutions have been designed using a unique network configuration for the three considered demographic problems, which has the advantage of simplifying the design complexity and debugging of these deep architectures when handling related handwriting problems. Finally, the comparison of achieved results to those presented in related works revealed the best average accuracy in the gender classification problem for the considered datasets.

1. Introduction

In spite of current technological advances, there are not still algorithms allowing a computer to transcribe the content of any “difficult” handwritten document (e.g., a historical document). The general handwriting recognition problem presents many difficulties produced by interpersonal and intrapersonal variations when writing, the cursive nature of handwriting, the use of different pen types, or the presence of paper with noisy background [1]. Srihari et al. [2] have studied and determined with scientific rigor the individuality of handwriting. Regarding the handwriting recognition problem, there are two variants: *offline* and *online* recognition [1]. The offline problem consists in recognizing handwritten text that has previously been written on paper and then digitized. The online handwriting problem aims to recognize the text that was written using some kind of electronic digitizer device. The sensors of this device also record a set of dynamic measures about how the act of writing is produced (e.g., writing pressure, pen altitude, and azimuth). In recent years, there

has been more progress on the online modality but the offline one is still far to be solved in an unrestricted manner [3].

There exist additional complex recognition problems associated with handwriting. Automatic classification of individuals into different demographic categories [4–6] using handwriting presents interesting applications in areas such as Forensic Biometrics, Psychology, Human-Computer Interaction, or Biometric Security [7, 8]. For example, when an anonymous piece of handwritten text is found at a crime scene and it is possible to automatically recognize that the writer is a “left-handed woman,” this can reduce the group of suspects to be investigated. Psychology can also get benefits from research on handwriting style since it could be possible to identify correlations between the handwriting and some personality attributes of the writer. In the field of Human-Computer Interaction, if gender and/or handedness of a user can be automatically predicted, the computer applications could offer him/her a more personalized interaction (e.g., gender-oriented advertising). Biometric Security can also benefit from handwriting prediction since this fact can



FIGURE 1: Some samples of handwritten text lines in English from IAM database (a, c, e, and g) and in Arabic from KHATT database (b, d, f, and h). These texts were written by (a)-(b) right-handed male, (c)-(d) left-handed male, (e)-(f) right-handed female, and (g)-(h) left-handed female, respectively.

be combined with other biometric modalities in order to improve security when accessing computer systems.

These handwriting-based demographic prediction problems include gender, handedness, age ranges, or even nationality of a person [9]. This group of supervised learning problems can be considered as binary or multiclass ones. The most common binary problems are gender prediction (where handwriting texts can be classified as written by men or by women) and handedness prediction (where handwriting texts can be classified as produced by right-handed or by left-handed writers). Among the multiclass problems, one can discriminate among texts written by people included in different age intervals, in specific human races, or even in groups of nationalities. A property of all these problems is that they can be either balanced (i.e., where approximately half of the population belong to each class) as in the case of gender classification or unbalanced as in the case of the handedness classification (where the “left-handed” class only includes approximately 10% of the individuals). In general, these demographic classification problems are very complex, even for humans, since it is quite difficult to find which handwriting features properly characterize each involved class. An example of this occurs in the classification of gender. Although it is accepted that feminine writing is rounder and neater than masculine one, there are some cases where masculine writing may have a “feminine” appearance and vice versa. Figure 1 illustrates different handwriting text lines written by a “right-handed male,” a “left-handed male,” a “right-handed female,” and a “left-handed female” using two different alphabets (Latin and Arabic, resp.). In this paper, we additionally aim to analyze the relationships between the gender and handedness handwriting features.

1.1. Related Work. There are relatively few works in the literature on these problems (mostly, on the binary ones) which have been started to be investigated recently in an automatic form [9–11]. One important difficulty is that there are few handwriting databases with annotated demographic information of the writers. Other aspects that hinder this problem are similar to those presented by the general handwriting recognition problem (e.g., cursive features).

Neural networks (NN) have been applied for many years in the analysis of high-dimensional, nonlinear, and complex classification problems [12], as is the case of automatic handwriting recognition [1]. The handwriting problem has been investigated since many years using different types of NN [13, 14] for both online and offline cases [1] and even also for alphabets different from Latin (e.g., Arabic in [15]).

Two main situations can be distinguished in the automatic offline handwriting recognition of text: first, the recognition of isolated characters, which is actually solved with error rates lower than 1% [16]; second, the recognition of groups of connected characters (e.g., words or text patches), where the success rates are still far from this value. Traditionally, continuous handwriting recognition [17] from digitized documents followed a sequence of stages including preprocessing, segmentation, feature extraction, and classification [18]. Handwritten character segmentation is a particularly complex problem because it is sometimes impossible to determine where one letter ends and where the next one begins. To overcome this difficulty, holistic methods have been recently proposed, which handle each word as a whole. These solutions were usually based on hidden Markov models (HMM) [19] or neural networks (NN) [3]. In recent years, this has changed with the emergence of algorithms that

allow training deep networks presenting multiple hidden layers which are able to extract more complex and relevant features. Since each hidden layer computes a nonlinear transformation of the previous layer, a deep network can have significantly greater representational capacity (i.e., it can learn more complex functions) than a shallow network. In a 2015 survey, Patel and Thakkar [18] pointed out that a 100% success rate is still far behind in the problem of continuous handwriting recognition. Holistic methods eliminate the need to perform complex segmentation tasks on handwriting. In 2016, Bluche [20] presented a system that uses a modification of a Long Short-Term Memory (LSTM) neural network that performs the processing and recognition of complete paragraphs. However, these methods limit the vocabulary that may appear in the text. For this reason, only good recognition results are obtained in cases of limited vocabularies [18]. To break this line of reduced vocabularies, some authors are successfully employing recurring networks such as Connectionist Temporal Classification (CTC) [20, 21].

Regarding the considered demographic classification problems using handwritten texts [22, 23], gender prediction has been the most addressed one. It was studied by Graphonomics and Psychology in a nonautomatic form since the beginning of last century [24, 25]. One of the first automatic methods to classify gender from offline handwriting was presented by Hecker in 1996 [26]. Using handwriting of 96 males and 96 females and automatic pixel intensity statistics, the author achieved an overall classification rate of 71.5%. In 2003, Koppel and collaborators [27] used automatic learning algorithms with manuscript documents extracted from the British National Corpus (BNC) [28]. Each document was represented by a feature vector of characteristics, whose dimensionality was reduced by eliminating irrelevant features. Their experiments produced an average correct classification higher than 85% for gender classification. In 2004, Tomai et al. [29] applied a k -nearest neighbor (knn) classifier to microfeatures extracted from offline characters from the CEDAR letter database [2] to diverse demographic problems and reported gender classification results of around 70%. Liwicki et al. [10] proposed two online gender classification approaches, respectively, based on SVM classifier and a mixed Gaussian model (GMM). The experiments performed for the evaluation were carried out with the IAM database and showed a correct prediction of 62% with SVM and 67% with GMM in gender classification. These same authors in 2011 [30], using again GMM, obtained global accuracy results of 67.57% for both offline and online gender recognition using the IAM database. Al Maadeed and Hassaine (2014) [9] focused their research on the problem of automatic gender prediction from offline manuscripts using two approaches. In the first one, all individuals wrote the same text, while in the second one, each individual wrote a different text. From each document, they extracted a set of shape features (e.g., curvatures, chain codes, or stroke orientations) that were classified using Random Forests (RF) and Kernel Discriminant Analysis (KDA). The evaluation of the system was performed using the QUWI database [31] through different experiments with Arabic texts, English texts, and the combination of both.

Best prediction results were achieved by combining both languages and when the handwritten texts were the same, with an accuracy of 69.8% with RF and 72.3% with KDA, respectively. Bouadjenek and collaborators (2015) [11] have addressed the gender classification problem using features from Histogram of Oriented Gradients (HOG) and an SVM classifier. Their evaluation was performed using the IAM and KHATT databases, which contain handwritten documents in English and Arabic, respectively, and achieved average precision of 75.45% for IAM and 68.89% for KHATT. Siddiqi et al. (2015) published a study on gender classification from handwriting [32] which focused on features based on slant/orientation, roundedness/curvature, neatness/legibility, and writing texture. These features were classified using ANN and SVM and evaluated on the QUWI and the MSHD databases. The best classification results for the two databases were achieved using slant and curvature features with an SVM classifier (68.75% for QUWI and 73.02% for MSHD, resp.). In 2016, two studies regarding the gender classification problem were published at the ICDAR conference. A first study, by Mirza et al. [33], used texture features that were extracted using a bank of multiscale and multiorientation Gabor filters, and these features were classified with feed forward neural networks. Best experimental results reported by these authors were achieved using only Arabic texts from the QUWI dataset. A second study, by Tan and collaborators [34], proposed the extraction of multiple geometrical (e.g., local curvature of strokes) and transformed (e.g., Fourier coefficients) features and the use of Mutual Information to select an optimal subset of features in classifying the writer's gender. This study reported an average accuracy of 67.2% using ICDAR 2013 and RDF datasets. In 2017, Akbari et al. [35] proposed an effective technique to predict gender that converts a handwritten image into a textured one that is decomposed into various subbands at various levels. These subbands are used to construct Probabilistic Finite State Automata (PFSA) that generate the feature vectors. With these vectors, they trained a neural network (NN) and an SVM. To evaluate both classifiers, text-dependent and text-independent tests have been performed with the QUWI and MSHD [36] databases. Their experiments showed correct classification results of 77.8% with SVM and 79.3% with NN in the case of QUWI dataset, whereas with the MSHD dataset these results were, respectively, 79.9% with SVM and 79% with NN. Finally, also in 2017, Bouadjenek et al. [37] compared Histogram of Oriented Gradients (HOG) with Local Binary Patterns (LBP) as feature extractors for gender classification on the IAM dataset. Using separately for the extracted HOG and LBP features an SVM classifier, the HOG produced better correct gender prediction (74% versus 70%).

The problem of handedness classification from handwriting has also been more recently studied in an automatic way [24, 38]. According to Saran et al. [39], it is possible to discriminate handedness based on direction of strokes and slope of letters (i.e., left-handed writers produce strokes in right-to-left direction and the slope of letters is backwards, whereas right-handed ones produce opposite features).

Bandi and Srihari [4] in 2005 presented an online handedness system based on pen pressure and writing movement

with a classification result of 74.4%. In 2007 Liwicki et al. [10] also proposed an online method for handedness detection using SVM and GMM for classification using the IAM database and reported results of 62% with SVM and 84.6% with GMM, respectively. Al-Maadeed and others [40] studied in 2013 the offline handedness classification problem (i.e., without using dynamic information from handwriting). They extracted shape and curvature features from strokes and used a knn classifier, reporting results of 71.5% on the QUWI database (with both English and Arabic texts). A work of 2015 by Bouadjenek et al. [11] applied to handedness prediction the same offline system that they used for gender classification (i.e., HOG for feature extraction and SVM as classifier) on the KHATT dataset (also with English and Arabic texts) reporting 83,93% of success. More recently, Al-Maadeed et al. [41] have presented a novel framework for handedness detection, using offline handwriting and fuzzy logic. These authors collected a database of handwritten texts (in Arabic and English) from 121 writers and extracted a high number of shape features from the texts. A dimensionality reduction stage, based on fuzzy conceptual reduction by applying the Lukasiewicz implication, was included. The classification stage was performed using a knn method, producing an average result of 83.43% for their dataset.

Most recent works present results for more than one demographic problem using handwriting (e.g., they separately handle both gender and handedness problems; see, e.g., [10]). Other recent papers additionally include some multiclass problems like age range prediction [11, 42] and nationality [9].

1.2. Proposed Approach. In general, there is an inherent difficulty in identifying the best features to discriminate between the subclasses (e.g., men versus women) in demographic classification problems based on handwriting [29]. Some types of deep networks like convolutional neural networks can find automatically good features and also perform the classification task. Convolutional neural networks had proven better capabilities to extract relevant handwriting features when compared to using hand-crafted ones for the automatic text transcription problem.

In this paper, we describe a detailed experimental study on the application of these deep neural networks to several automatic demographic classification problems based on handwriting. In particular, we address three types of demographic problems: gender, handedness, and the combined “gender-and-handedness” classification. In order to test our proposal, two public handwriting datasets are used: IAM with English texts and KHATT containing Arabic texts.

To the best of our knowledge, our work also presents the first approach to the combined gender-and-handedness prediction, which has not been addressed before by other researchers. Moreover, this multiclass approach for gender and handedness problems produced better average accuracy results than handling successively the two binary problems. Our solution exhibits generic behavior because it has a unique configuration of convolutional neural network for the three considered demographic problems.

1.3. Contributions and Outline of the Paper. The main contributions of this work are the following ones:

- (i) This is the first paper on the application of deep networks to demographic classification problems from handwriting. A different problem is identifying a writer from his/her handwriting using deep learning models, which has been recently studied by Xing and Qiao [43]. Moreover, although there exist other deep learning approaches to predict the gender, these are based on other types of input patterns different from handwriting. For example, Bartle and Zheng [44] used stylistic information in computer blogs, and Levi and Hassner [45] used facial images.
- (ii) In addition to the separated gender and handedness classification problems from handwriting, we introduce the combined gender-and-handedness problem, where four subclasses are defined: right-handed men, left-handed men, right-handed women, and left-handed women, respectively. This novel multiclass problem, which is not handled by previous works, is more complex than separate binary gender and handedness ones, and it is of interest to Forensic Biometrics applications [8].
- (iii) For the sake of simplicity in the proposed solutions, we have designed a unique configuration of convolutional neural network, with specific parameter values for each of the three considered demographic problems.
- (iv) Our prediction method remains relatively robust for more than one considered alphabet (i.e., Latin and Arabic), and it achieved competitive classification results in two of the most used datasets for these problems: IAM and KHATT.

This paper is organized as follows. Section 2 describes the methods and materials used in this research. Section 3 describes the experimental setup, presents the results achieved for each of the considered demographic problems, and discusses these results. Finally, Section 4 summarizes the conclusions of the work.

2. Materials and Methods

In this section, we summarize some fundamentals of deep learning and convolutional neural networks. Next, the common characteristics of the proposed convolutional model, used for all considered handwritten-based demographic problems, are described. We continue with a description of the preprocessing applied to training data. Next, the specific features of the convolutional networks applied to respective gender, handedness, and combined classification problems are explained. Finally, two databases used in our experimentation are summarized.

2.1. Deep Learning and Convolutional Neural Networks. The essence of deep learning is the application to learning problems of artificial NN that contain more than two hidden layers. Deep learning has produced extraordinary advances

in difficult computational problems that have resisted the attempts of the AI community during decades. This new paradigm has been used to discover complex structures in high-dimensional data [46]. Deep learning is currently being applied to many scientific domains and, in particular to image recognition problems where it has beaten other machine-learning techniques [46].

Convolutional neural network (CNN or ConvNet) is a well-studied deep learning architecture that was inspired by the natural visual perception mechanism. LeCun and collaborators [13] presented in 1990 the framework for the CNN, and they created a multilayer network called LeNet-5 which was able to classify handwritten digits. This type of NN included three types of layers: convolutional, pooling (or subsampling), and fully connected (or dense) layers. Convolutional layers aim to learn feature representations of inputs. Each of them is composed of several convolution kernels which are used to compute different feature maps. Each neuron of a feature map is connected to a region of neighbor neurons of the previous layer. The new feature map is calculated by first convolving the input with a learned kernel and then applying an element-wise nonlinear activation function on the convolved results [47]. Note that the kernel is shared by all spatial locations of the input. The complete feature maps are obtained by using several different kernels. Each pooling layer searches to achieve shift invariance and reduces the resolution of the feature maps. It is usually placed between two convolutional layers. Finally, after several stacks of convolutional and pooling layers, there appears one or more fully connected layers which perform the final classification task. Like other multilayer networks, CNNs were trained using types of backpropagation algorithms.

However, due to the need of large training data and the lack of computing power at that time, these original LeNet-5 networks could not perform well on complex problems. In 2012, Krizhevsky et al. [48] proposed a new CNN model with a deeper structure, called ImageNet, which showed significant improvements upon other image classification methods. It included data augmentation to enlarge the training dataset, “dropout” (i.e., dropping out a percentage of neuron units, both hidden and visible) for reducing overfitting, ReLU activation function for reducing the effect of gradient vanishing during backpropagation, and the use of GPUs for accelerating the overall training process. Moreover, the application of proposed good practices [48] when designing and training convolutional networks is also important for achieving effective results.

The inputs of the CNN for our considered problems are order-3 tensors (i.e., a monochannel image with H rows and W columns). These inputs are processed sequentially through all network layers and produce as output a c -dimensional vector for the classification problem with c classes. Using some mathematical notation, the value at position (i, j) in the k -th feature map of the l -th network layer, represented as: $z_{i,j,k}^l$, can be calculated as follows:

$$z_{i,j,k}^l = w_k^l x_{i,j}^l + b_k^l \quad (1)$$

where w_k^l and b_k^l are the respective weight and bias vectors of the k -th filter in the l -th layer and $x_{i,j}^l$ is the local input region for this position and layer. Network weight masks (which define convolution kernels) w_k^l are shared, thus reducing the training time. Like other types of NN, in order to recognize nonlinear features, the value computed by (1) is passed through the ReLU activation function:

$$a_{i,j,k}^l = \text{ReLU}(z_{i,j,k}^l). \quad (2)$$

These results, produced after the inputs pass through a convolutional layer, are then processed by a pooling layer (i.e., it can be a *max-pooling* layer, placed between two convolutional layers) in order to achieve invariance and reduce the size of feature maps. New intermediate values are computed as follows:

$$y_{i,j,k}^l = \text{max_pooling}(a_{i,m,n}^l), \quad (3)$$

where $(m, n) \in \mathcal{R}_{i,j}$ represents a local neighborhood around position (i, j) . Note that kernels of lower layers can detect low-level features while kernels in higher layers detect high-level features. Finally, after several convolutional and pooling layers, there exists one or more fully connected layers, and the last one is the output layer which classifies the input test pattern into one of the predefined categories (i.e., supervised classification).

2.2. Proposed Deep Learning Architecture Framework for Demographic Problems. This subsection describes the common characteristics in our solution for the considered demographic problems using handwritten text. Next, in the successive subsections, we point out the specific aspects of each particular problem, namely, gender classification, handedness classification, and combined gender-and-handedness classification. For predicting the subclasses in the three problems, we used the same CNN architecture shown in Figure 2. The general proposed neural model has 6 trainable layers, grouped in 2 stacks of convolutional and subsampling (or max-pooling) layers, and 2 final dense layers. The network receives input images with a spatial resolution of 30×100 . After some experimentation, we used kernels of size 5×5 for the convolutional layers and of size 2×2 for the subsampling layers. These experiments showed us that smaller kernels produced worst results and bigger kernels did not improve significantly the results. Parameters a , b , and c in this figure, respectively, correspond to the number of feature maps for the first convolutional layer, the number of feature maps for the second convolutional layer, and the number of output neurons in the last layer (i.e., the problem subclasses) for each of the three demographic problems. The corresponding values of these parameters for each considered problem are detailed in Section 2.4.

In all the convolutional layers, we used zero padding to preserve the spatial size, all hidden layers include the nonlinear rectification units (ReLU), and the output layer used the *SoftMax* activation function. Dropout regularization with value of 0.25 was applied to each of the convolutional layers and with value of 0.5 to the first dense layer. The binary

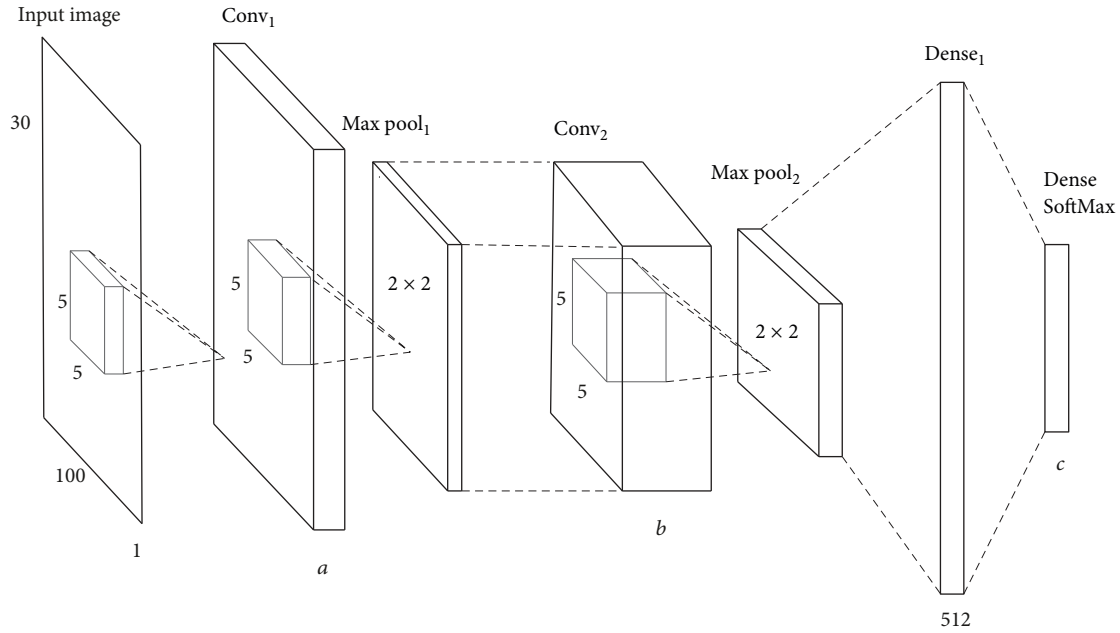


FIGURE 2: Proposed CNN architecture for demographic problems using handwriting.

models were trained using *Stochastic Gradient Descent* (SGD) and the multiclass one was trained using *Adam* optimization algorithms, respectively, both with a learning rate value of 0.001 and net *weight decay value* of $1E-7$. All these parameter values were determined through experimentation.

Figure 3 sketches the prediction method followed to address all demographic classification problems. Each dataset, composed by a collection of handwritten separated lines (each one with its associated demographic information), is partitioned into subsets of text images: training, validation, and test ones, respectively. There exists also separation between the “training” and “test” individuals in order to prevent the CNN model from “learning” the specific handwriting of each individual. Given one handwritten line, it is automatically splitted into their component “words” (i.e., text patches) that, after being preprocessed, will be the inputs to the network. The extraction of “words” given in a text line is computed by first applying a morphological dilation to the line, then extracting the contours from the resulting dilated binary image, and finally computing the bounding rectangles from the connected contours.

The CNN model can predict for a given unknown word its subclass in each considered problem. Finally, the predicted results of test words, contained into a text line, are combined by a majority-voting scheme to determine the final prediction result for the considered test line. The advantage of this approach is making available to the network a higher number of training samples (i.e., thus allowing it to achieve internal representations of smaller pieces of text when analyzing the involved graphisms). Moreover, we use a *Learn-on-Demand* method [49] when training the CNN models, thus avoid generating in advance all the possible training samples for the network.

2.3. Preprocessing of Training Data. When using deep learning neural networks in classification problems, it is necessary to have a large amount of training data (in some cases, millions) so that the network is able to discriminate correctly among the different classes. Data augmentation is an elegant solution to the problem and it consists in transforming the available data into new data without altering their nature. Some common data augmentation methods [47] are geometric transformations (such as normalization, rotation, shifting, or rescaling), morphological operations, and various photometric transformations. Of course, these transformations can be successively applied to the same input image [50].

Pseudocode 1 summarizes our data augmentation approach, which is applied to any training word image w .

Using Pseudocode 1, we produce synthetic word images as shown in Figure 4. These generated images are rescaled to be the training inputs of the CNN classifier.

2.4. Specific Model Features for Gender, Handedness and Combined Classification. Regarding our solution to the binary gender problem with a convolutional network, the used architecture configuration is presented in Figure 2 with respective parameter values of $a = 128$ (i.e., number of feature maps for the first convolutional layer), $b = 256$ (i.e., number of feature maps for the second convolutional layer), and $c = 2$ (number of output neurons or subclasses in the last layer). The number of training epochs for this problem was 200. In each epoch an amount of 100,000 synthetic training and 20,000 validation words (obtained from the original ones using the algorithm of Pseudocode 1) were presented to the network. One-half of the synthetic training and validation sets of words correspond to masculine writers and the other half to feminine ones.

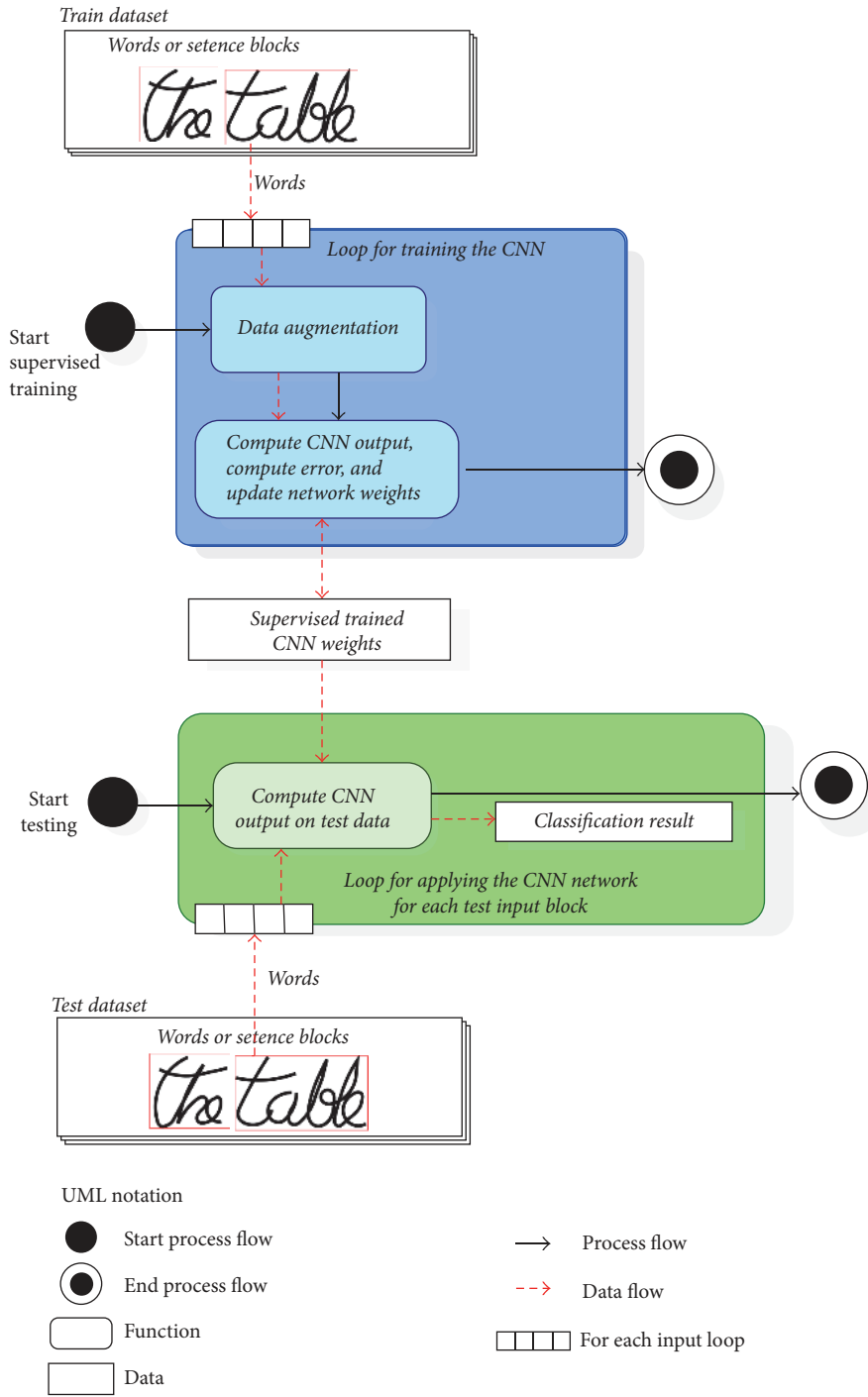


FIGURE 3: Overview of proposed method for demographic classification problems.

Handedness prediction is also a binary problem (i.e., “right-handed” and “left-handed” subclasses), where the number of original patterns in both subclasses is unbalanced for most of available datasets. In general, the databases have around 90% of samples for right-handed writers and 10% for left-handed ones, which is approximately the proportion of both subclasses in the world. The CNN architecture configuration used is the same shown in Figure 2 with respective parameter values of $a = 64$, $b = 128$, and $c = 2$. The number

of training epochs for this problem was 200. In each epoch, a total of 100,000 synthetic training and 25,000 validation words (obtained from the original ones using the algorithm of Pseudocode 1) were presented to the network. One-half of the synthetic training and validation words corresponded to right-handed writers and the other half to left-handed ones.

The combined multiclass problem categorizes the subclasses of combining gender with handedness. In particular, it needs previous partitioning of the datasets into individuals

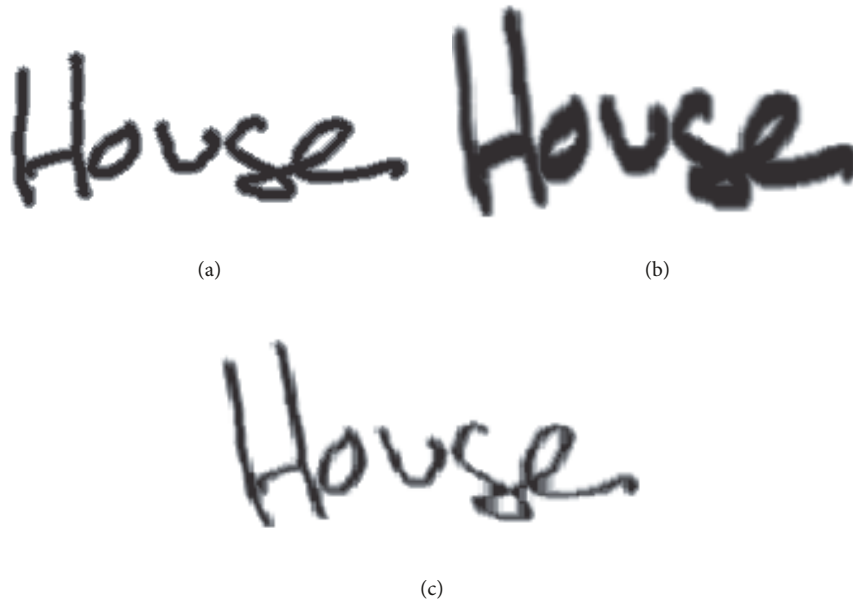


FIGURE 4: Some examples of combined transformations applied to a sample training word: (a) original image; (b) synthetic image obtained from (a) by successively applying a -5° inclination, a vertical scaling of 20%, and a morphological dilation; (c) synthetic image obtained from (a) by successively applying a -10° inclination, both vertical and horizontal scaling of 10% and a morphological erosion.

```

algorithm generate_modify_image ( $w$ ):
  # affine inclination
   $i$  = randomInclination in  $\{-10^\circ, -5^\circ, 0^\circ, 5^\circ, 10^\circ\}$ 
   $w_r$  = affine_rotation ( $w, i$ )
  # positive scaling
  vs = randomVerticalScaling in  $\{0\%, 10\%, 20\%\}$ 
  hs = randomHorizontalScaling in  $\{0\%, 10\%, 20\%\}$ 
   $w_{r,s}$  = increasing_scaling ( $w_r, vs, hs$ )
  # binary morphological filter with a  $3 \times 3$  structuring element
   $m$  = randomMorphologyOperation in  $\{\text{erosion, dilation, none}\}$ 
   $w_{r,s,m}$  = morphology_filter ( $w_{r,s}, m$ )
  # produce  $30 \times 100$  rescaled image  $w_o$  using bilinear interpolation
   $w_o$  = normalize_size ( $w_{r,s,m}$ , "bilinear") # input training image for the CNN
  return  $w_o$ 

```

PSEUDOCODE 1: Pseudocode of successive random transformations applied on training images.

who correspond to “right-handed men,” “left-handed men,” “right-handed women,” and “left-handed-women,” respectively. Regarding our convolutional network solution, we also used the CNN architecture configuration presented in Figure 2 with parameter values of $a = 32$, $b = 64$, and $c = 4$, respectively. The number of training epochs for this problem was 250. In each epoch, a total of 130,000 synthetic training and 20,000 validation words (also obtained from the original ones using the algorithm of Pseudocode 1) were presented to the network. One-quarter of synthetic training and validation words corresponded to right-handed masculine, left-handed masculine, right-handed feminine, and left-handed feminine writers, respectively.

All of our algorithms were coded in Python using the OpenCV Computer Vision library and the Keras high-level

API for neural networks. Our models were trained and tested using a NVIDIA GeForce GTX TITAN Black GPU with 6 GB of frame buffer memory.

2.5. IAM and KHATT Databases. The IAM database [51–53] was created by the Computer Vision and Artificial Intelligence Research Group in the University of Bern (Switzerland). This dataset includes both an online version and an offline one. The database is specially designed to train and test text recognizers, as well as performing identification and verification experiments for writers.

The complete version of IAM Handwriting Database 3.0 is structured as follows. A number of 657 writers contributed samples of their handwriting. There are 1,539 pages of scanned text, 5,685 isolated and labeled sentences, 13,353

TABLE 1: Number of training and text lines, per subclass and experiment, from Offline IAM database.

Gender							
Male (M)				Female (F)			
Train		Test		Train		Test	
5,228		1,659		3,234		805	
Handedness							
Right-handed (R)				Left-handed (L)			
Train		Test		Train		Test	
6,946		3,011		728		250	
Combined							
RM		LM		RF		LF	
Train		Test		Train		Test	
4,851		1,473		429		134	
2,497		1,113		297		132	

isolated and labeled text lines, and 115,320 isolated and labeled words. This dataset contains forms of unconstrained handwritten text, which were scanned at a resolution of 300 DPI and saved as PNG images with 256 gray levels. From each writer, the following information was stored in the database: the gender, native language, and other features relevant for the analysis such as if he/she is right-handed or left-handed writer.

In our experiments, we have only used a subset of the offline sentences of this dataset (which are here named as ‘‘Offline IAM’’). Table 1 shows the number of training and test lines used for each class and considered problem for the Offline IAM dataset.

The KHATT database [54, 55] was created by a research group of the King Fahd University (Saudi Arabia). It contains offline handwritten Arabic texts of approximately 1,000 writers from different countries, genders, handwriting, and educational levels. This database can be used in problems of identification of writers, techniques of binarization and elimination of noise, handwriting recognition, and techniques of line segmentation. Each of the 1,000 writers, 677 men and 323 women, wrote four paragraphs which contained a common part for all writers and a free part where each one wrote a different text. A total of 4,000 paragraphs were segmented into text lines with about 200,000 different words. In addition, 928 of the writers were right-handed and 72 were left-handed. The database also contains information related to writers such as name, age, gender, or handedness. So, it can be very useful when using the data for a particular demographic problem. Table 2 shows the number of training and test lines used for each class and problem considered for the KHATT dataset.

3. Results and Discussion

This section describes the experiments and corresponding results on the two used databases: Offline IAM and KHATT, respectively. Next, these results are compared to those presented by related works. Finally, an analysis and discussion on the achieved results are also included.

TABLE 2: Number of training and test lines used, per subclass and experiment, from KHATT database.

Gender							
Male (M)				Female (F)			
Train		Test		Train		Test	
3,931		609		1,846		358	
Handedness							
Right-handed (R)				Left-handed (L)			
Train		Test		Train		Test	
5,344		877		433		90	
Combined							
RM		LM		RF		LF	
Train		Test		Train		Test	
3,601		547		330		62	
1,743		330		103		28	

In order to evaluate our approach, we use some standard performance metrics for binary and multiclass categorization. These measures, which are calculated for each subclass of a given demographic problem, are precision, recall, and F -measure. They are defined for a binary problem and given subclass c as given by

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c} \quad (4)$$

$$\text{Recall}_c = \frac{TP_c}{TP_c + FN_c} \quad (5)$$

$$F\text{-measure}_c = 2 \frac{\text{Precision}_c \times \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c}, \quad (6)$$

where TP_c , FP_c , and FN_c are, respectively, the number of true positives, false positives, and false negatives in the class c .

The overall accuracy of the binary model can be directly computed from any of the two classes c , since it has the same value for the two classes due to the exchange of positives and negatives property [56]. This accuracy value is computed as follows:

$$\text{Overall}_{\text{Accuracy}} = \frac{TP_c + TN_c}{TP_c + TN_c + FP_c + FN_c}. \quad (7)$$

The previous formulae can be extended to multiclass categorization problems [56]. Definitions of Precision_c and Recall_c are now adapted for our 4-class combined demographic problem. Given the confusion matrix $M = [(M_{i,j}) \mid 1 \leq i, j \leq 4]$ corresponding to our multiclass problem, these metrics are now computed as follows:

$$\text{Precision}_c = \frac{M_{c,c}}{\sum_j M_{j,c}} \quad (8)$$

$$\text{Recall}_c = \frac{M_{c,c}}{\sum_j M_{c,j}}$$

The expression of F -measure for each class in the multiclass problem is computed using (6) but with accuracy and precision values, respectively, computed by (8). Finally,

TABLE 3: Performance results for gender classification on IAM database.

Male (M)	Precision	86.81
	Recall	84.14
	<i>F</i> -measure	85.46
Female (F)	Precision	69.26
	Recall	73.66
	<i>F</i> -measure	71.39
Overall accuracy		80.72

TABLE 4: Performance results for handedness classification on IAM database.

Right-handed (R)	Precision	96.75
	Recall	93.02
	<i>F</i> -measure	94.85
Left-handed (L)	Precision	42.60
	Recall	63.40
	<i>F</i> -measure	50.64
Overall accuracy		90.70

the average accuracy for the multiclass problem [56] can be computed as follows:

$$\text{Average Accuracy} = \frac{\sum_{c=1}^l ((TP_c + TN_c) / (TP_c + TN_c + FP_c + FN_c))}{l}, \quad (9)$$

where the respective accuracy values of all l classes (with $l = 4$) are averaged.

In our context, the precision of a subclass c is the quotient between the number of correctly classified handwritten text lines into the subclass c and the total number of text lines classified into this subclass c . The recall of a subclass c is the quotient between the number of correctly classified handwritten text lines into the subclass c and the number of text lines that truly belong to the class c . The *F*-measure combines precision and recall and reflects the relative importance of recall with respect to the precision. Finally, the average accuracy represents a global measure of the classifier’s performance for each considered problem. As recommended by [56] for binary and multiclass classification, the previous evaluation measures have been applied to determine the performance of our proposals.

3.1. Experiments Using the Offline IAM and KHATT Datasets.

The previous evaluation measures have been applied to determine the performance of our models in the considered demographic prediction problems using English and Arabic texts. Tables 3, 4, and 5, respectively, present the calculated scores (in %) for the gender, handedness, and combined problem using the Offline IAM dataset, according to the measures given by (4)–(9).

Note that if both binary gender and handedness problems were independently handled, the joint average accuracy produced by the corresponding classification models should

TABLE 5: Performance results for combined gender-and-handedness classification on IAM database.

Right-handed male (RM)	Precision	69.98
	Recall	76.10
	<i>F</i> -measure	72.91
Left-handed male (LM)	Precision	53.38
	Recall	52.99
	<i>F</i> -measure	53.18
Right-handed female (RF)	Precision	62.51
	Recall	58.58
	<i>F</i> -measure	60.48
Left-handed female (LF)	Precision	66.22
	Recall	37.12
	<i>F</i> -measure	47.57
Average accuracy		83.19

TABLE 6: Performance results for gender classification on KHATT database.

Male (M)	Precision	71.21
	Recall	85.14
	<i>F</i> -measure	77.55
Female (F)	Precision	62.11
	Recall	41.43
	<i>F</i> -measure	49.70
Overall accuracy		68.90

be the product of their individual accuracies. This would produce, using overall accuracy values given in Tables 3 and 4, an average accuracy of 73.21%. This result is worse than 83.19% (see Table 5) obtained when we train a unique 4-class combined classification system. This fact, together with the economy in training times, shows that the proposed combined multiclass approach for the two considered problems is more effective than independently solving one binary problem and on the first classification apply the second one (i.e., in a hierarchical fashion).

Due to a substantially lower number of original training images in the KHATT database, we have applied the *transfer learning* technique (also known as *inductive training* or *pretraining*) in order to improve the classifications results for such dataset. This pretraining was only applied to the handedness and combined gender-and-handedness problems. Instead of randomly initializing the weights of the CNN connections, we have used the pretrained models built for the Offline IAM database and, after that, trained with them the respective networks with the corresponding training patterns of the KHATT dataset. This way, the knowledge gained for the CNN, while learning to recognize handwritten words with IAM, is transferred to the KHATT network. This practice is common for these networks (e.g., ImageNet) because many datasets do not have the sufficient size for enabling convolutional networks to extract relevant features for producing good classification results.

Tables 6, 7, and 8, respectively, present the calculated scores (in %) for the gender, handedness, and combined

TABLE 7: Performance results for handedness classification on KHATT database.

Right-handed (R)	Precision	92.18
	Recall	74.24
	<i>F</i> -measure	82.24
Left-handed (L)	Precision	13.34
	Recall	38.64
	<i>F</i> -measure	19.83
Overall accuracy		70.91

TABLE 8: Performance results for combined gender-and-handedness classification on KHATT database.

Right-handed male (RM)	Precision	69.23
	Recall	44.07
	<i>F</i> -measure	53.86
Left-handed male (LM)	Precision	11.84
	Recall	14.75
	<i>F</i> -measure	13.14
Right-handed female (RF)	Precision	41.81
	Recall	44.27
	<i>F</i> -measure	43.01
Left-handed female (LF)	Precision	3.76
	Recall	25.93
	<i>F</i> -measure	6.57
Average accuracy		70.84

problem using the KHATT dataset (i.e., Arabic script), according to the measures given by (4)–(9).

Note that the same fact with respect to proposed combined classification happens for the KHATT database. If both binary gender and handedness problems were independently handled, the average accuracy (obtained from values of Tables 6 and 7) is 48.86%. This result is much worse than 70.84% obtained when we train a unique 4-class combined classification system. By averaging the improvement accuracy rates of both Offline IAM and KHATT databases, our multiclass approach improved the accuracy by 29.26% when compared to separately and successively handling both binary problems. Perhaps, when training convolutional networks independently for the two demographic problems, the networks are not able to discover handwriting features which capture the interconnections between both individual problems. Moreover, when training a network for the combined multiclass problem, these related handwritten characteristics are better discovered.

With respect to the training times required for the convolutional network models using the IAM dataset, the gender problem used 100,000 training sample images and other 25,000 ones for validation (of the training). These steps were performed during 200 epochs (i.e., about 61 hours). Similar training times were required for the other two considered problems using the IAM dataset. When training our models for the KHATT dataset, the training times were significantly increased since, as explained, we first applied a pretraining

of the convolutional networks with the images of the IAM dataset.

3.2. Comparison with Related Works. Comparing our research results to those ones published on the same gender and handedness problems using the same datasets is difficult because of the differences in experimental aspects and the way the classification results are reported. Differences in experimental aspects are as follows: the different number and distribution of original images between the categories for training, validating, and testing the classification systems; the different alphabets used; the usage of the same texts written by all the writers or of different texts for each writer; and/or whether there is any preprocessing on the original datasets images. With respect to published results, there are several works [9, 35] that only report an overall accuracy result for each classification method used in the problems. In the common case of unbalanced classes, as is the case of “left-handed” in the handedness problem, this overall accuracy is not appropriate and specific measures per each class are more convenient.

Taking into account the previous remarks, we have compared our results with those reported in [10, 11, 30] which use the same databases and the same performance measures per class. The analyzed results are presented in Table 9, and they show that our approach produces the best scores in the gender problem for both IAM and KHATT databases, while the results presented in [11] are the best ones for the handedness problems for the considered datasets.

3.3. Analysis and Discussion. The analysis of our experimental results for the three considered handwritten-based demographic problems on the IAM and KHATT datasets have raised the following aspects:

- (i) The proposed combined multiclass approach for gender and handedness problems produced better average accuracy results than handling successively the two binary problems.
- (ii) Our common convolutional architecture framework for the three demographic problems has produced acceptable prediction results, even for the combined gender-and-handedness prediction problem, where there are fewer training text lines in the involved subclasses.
- (iii) Classification results on the KHATT database are worse than the corresponding ones on the Offline IAM database. This can be caused by the more reduced number of original training examples in the Arabic dataset. In spite of applying data augmentation and transfer learning as optimization techniques for improving the classification results of the convolutional networks, we noticed that, when there is a more reduced number of original training samples (i.e., those ones provided by the dataset without data augmentation), the prediction results are worse.
- (iv) We have to remark the importance of transfer learning (or pretraining) when training convolutional networks in problems with a reduced number of original

TABLE 9: Comparison of overall accuracy results for gender and handedness problems on Offline IAM and KHATT databases presented in related papers.

	Liwicki et al. [30]		Liwicki et al. [10]		Bouadjenek et al. [11]		Our approach	
	Gender	Handedness	Gender	Handedness	Gender	Handedness	Gender	Handedness
IAM	67.57	84.66	75.45	100.0	80.72	90.70		
KHATT			68.89	83.93	68.90	70.91		

samples per class. This is the case of “left-handed” in the KHATT database.

- (v) Some papers addressing demographic problems from handwriting only report global accuracy results in classification. However, these results are not valuable at all when the number of patterns per subclass is highly unbalanced (e.g., “left-handed”). It is more important to report the correct prediction results per class (i.e., using precision and recall measures).

Next, discussion of results is completed in several directions: complexity of the proposed model respect classical approaches, necessity of data augmentation, and computing times, respectively. Regarding the complexity of the proposed model with respect to classical approaches (i.e., feature-based ones), from a developer’s viewpoint using convolutional neural networks (CNN) is simpler than determining which features are the best ones for discriminating each class. Differently from other analyzed feature-based proposals (see, e.g., [11, 35, 40]), when using CNN one does not have to discover which features are relevant to solve the problem (i.e., this approach is a drop-in replacement to hand-made feature descriptors). Since these good internal representations are now found by the network, the model is much simpler and powerful at the same time. Regarding data augmentation, it is true that these networks require a very high number of training examples to learn well the involved classes. These examples were obtained synthetically, by creating word images through applying combinations of multiple transformations with different parameters on the original training images. In our approach, the considered transformations were random left/right sloping, vertical/horizontal scaling, and morphological erosion/dilation. With respect to training times, despite advances in CNNs these models are still highly time consuming. However, and as is a common practice, we have drastically reduced the training times of our neural networks using a cluster of GPUs.

4. Conclusion

This paper presented a detailed experimental study on the application of deep neural networks to several automatic demographic classification problems based on handwriting. In particular, we have addressed three problems: gender, handedness, and the combined “gender-and-handedness” classification. We tested our proposal on two public handwriting datasets (IAM with English texts and KHATT containing Arabic texts). Convolutional neural networks had proven better capabilities to extract relevant handwriting

features when compared to using hand-crafted ones for the automatic text transcription problem. Our work also tackled the combined gender-and-handedness prediction, which has not been addressed before by other researchers. Moreover, this combined multiclass approach for gender and handedness problems produced better average accuracy results than handling successively the two binary problems. Our solution exhibits generic behavior because it has a unique configuration of convolutional neural network for the three considered demographic problems. Finally, the comparison of these results to other connected works reveals that our solution produced the best accuracy results for the gender classification problem on both tested handwriting databases.

In summary, the advantages and novel aspects of our proposal are the following ones:

- (1) To the best of our knowledge, this is the first paper on the application of deep networks to demographic classification problems from handwriting.
- (2) We introduce and effectively address the combined multiclass “gender-and-handedness” problem.
- (3) Our approach only used a unique configuration of convolutional neural network, with specific parameter values for the three considered demographic problems.
- (4) Finally, the proposed gender/handedness prediction method remains relatively robust for more than one alphabet, and it achieves competitive classification results for two of the most used datasets in this problem: IAM and KHATT.

Future work will include the extension of this research to additional handwriting datasets containing texts written in other alphabets. We are also interested in studying new additional multiclass handwritten-based problems, especially the age prediction one. Another planned future research is the adaptation of our proposed framework in order to predict some types of demographic information from writers, which is present in historical handwritten documents.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research has been supported by the Spanish Ministerio de Economía y Competitividad (MINECO), under the Projects TIN2014-57458-R and TIN2017-85221-R.

References

- [1] R. Plamondon and S. N. Srihari, "On-line and off-line handwriting recognition: A comprehensive survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 63–84, 2000.
- [2] S. N. Srihari, S.-H. Cha, H. Arora, and S. Lee, "Individuality of Handwriting," *Journal of Forensic Science*, vol. 47, no. 4, pp. 856–872, 2002.
- [3] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," in *Proceedings of the 21st Intl. Conf. Neural Information Processing Systems (NIPS'08)*, pp. 545–552, 2008.
- [4] K. R. Bandi and S. N. Srihari, "Writer demographic classification using bagging and boosting," in *Proceedings of the Intl. Graphonomics Society Conference (IGS'05)*, pp. 133–137, 2005.
- [5] S.-H. Cha and S. N. Srihari, "A priori algorithm for sub-category classification analysis of handwriting," in *Proceedings of the 6th International Conference on Document Analysis and Recognition, ICDAR 2001*, pp. 1022–1025, IEEE, Seattle, WA, USA, September 2001.
- [6] S. Hamid and K. M. Loewenthal, "Inferring gender from handwriting in Urdu and English," *The Journal of Social Psychology*, vol. 136, no. 6, pp. 778–782, 1996.
- [7] J. Chapran, "Biometric writer identification: Feature analysis and classification," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 20, no. 4, pp. 483–503, 2006.
- [8] L. Ballard, D. Lopresti, and F. Monrose, "Evaluating the security of handwriting biometrics," in *Proceedings of the Proc.*, pp. 461–466, 2006.
- [9] S. Al Maadeed and A. Hassaine, "Automatic prediction of age, gender, and nationality in offline handwriting," *Eurasip Journal on Image and Video Processing*, vol. 2014, article no. 10, 2014.
- [10] M. Liwicki, A. Schlappach, P. Loretan, and H. Bunke, "Automatic detection of gender and handedness from on-line Handwriting," in *Proceedings of the 13th Conf. Intl Graphonomics Society*, 2007.
- [11] N. Bouadjeneq, H. Nemmour, and Y. Chibani, "Histogram of Oriented Gradients for writer's gender, handedness and age prediction," in *Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications, INISTA 2015*, IEEE, Madrid, Spain, August 2015.
- [12] G. P. Zhang, "Neural networks for classification: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 30, no. 4, pp. 451–462, 2000.
- [13] Y. LeCun et al., "Handwritten digit recognition with a back-propagation network in," in *Advances in neural information processing systems 2*, pp. 396–404, Morgan Kaufmann Publishers, 1990.
- [14] B. Gosselin, "Multilayer perceptions combination applied to handwritten character recognition," *Neural Processing Letters*, vol. 3, no. 1, pp. 3–10, 1996.
- [15] M. T. Parvez and S. A. Mahmoud, "Offline arabic handwritten text recognition: A survey," *ACM Computing Surveys*, vol. 45, no. 2, article no. 23, 2013.
- [16] U. Meier, D. C. Cireşan, L. M. Gambardella, and J. Schmidhuber, "Better digit recognition with a committee of simple neural nets," in *Proceedings of the 11th International Conference on Document Analysis and Recognition, ICDAR*, pp. 1250–1254, September 2011.
- [17] A. L. Koerich, R. Sabourin, and C. Y. Suen, "Large vocabulary off-line handwriting recognition: a survey," *Pattern Analysis & Applications*, vol. 6, no. 2, pp. 97–121, 2003.
- [18] M. Patel and S. P. Thakkar, "Handwritten character recognition in english: a survey," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 2, pp. 345–350, 2015.
- [19] T. Plötz and G. A. Fink, *Markov Models for Handwriting Recognition*, Springer Briefs in Computer Science, Springer, London, UK, 2011.
- [20] T. Bluche, "Joint Line Segmentation and Transcription for End-to-End Handwritten Paragraph Recognition," in *Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS'16)*, Allen Institute for Artificial Intelligence, Barcelona, Spain, 2016.
- [21] O. Nina, *Connectionist Temporal Classification for Offline Handwritten Text Recognition*, BYU Conference Center, 2016.
- [22] R. J. Klimoski and A. Rafaeli, "Inferring personal qualities through handwriting analysis," *Journal of Occupational Psychology*, vol. 56, no. 3, pp. 191–202, 1983.
- [23] T. Dziejczak, "Right hand writing vs. Left hand writing of one person. A comparative study," *Z Zagadnien Nauk Sadowych*, vol. 94, pp. 564–577, 2013.
- [24] R. A. Huber and A. M. Hendrick, *Handwriting Identification: Facts and Fundamentals*, CRC Press, Boca Raton, FL, USA, 1999.
- [25] E. S. Nogueras, M. F. Zanuy, and J. R. Alcobé, "Gender classification by means of online uppercase handwriting: a text-dependent allographic approach," *Cognitive Computation*, vol. 8, no. 1, pp. 15–29, 2016.
- [26] R. Hecker, "The scientific examination of sex differences," *Proc. Fifty-Fourth Annual Meeting of the American Society of Questioned Document Examiners*, 1996.
- [27] M. Koppel, S. Argamon, and A. R. Shimoni, "Automatically categorizing written texts by author gender," *Literary and Linguistic Computing*, vol. 17, no. 4, pp. 401–412, 2003.
- [28] A. G. Burnage and G. Baguley, *The British National Corpus, Library and Information Briefings*, 1996, <http://www.natcorp.ox.ac.uk/archive/papers/gblibs.html>.
- [29] C. I. Tomai, D. Kshirsagar, and S. N. Srihari, "Group Discriminatory power of handwritten characters," *Document Recognition and Retrieval XI, Proc. of SPIE-IS & T Electronic Imaging*, vol. 5296, pp. 116–123, 2004.
- [30] M. Liwicki, A. Schlappach, and H. Bunke, "Automatic gender detection using on-line and off-line information," *PAA. Pattern Analysis and Applications*, vol. 14, no. 1, pp. 87–92, 2011.
- [31] S. Al Maadeed, W. Ayouby, A. Hassaine, and J. M. Aljaam, "QUWI: An Arabic and English handwriting dataset for offline writer identification," in *Proceedings of the 13th International Conference on Frontiers in Handwriting Recognition, ICFHR 2012*, pp. 746–751, Italy, September 2012.
- [32] I. Siddiqi, C. Djeddi, A. Raza, and L. Souici-meslati, "Automatic analysis of handwriting for gender classification," *Pattern Analysis and Applications*, vol. 18, no. 4, pp. 887–899, 2015.
- [33] A. Mirza, M. Moetesum, I. Siddiqi, and C. Djeddi, "Gender classification from offline handwriting images using textural features," in *Proceedings of the 15th International Conference on Frontiers in Handwriting Recognition, ICFHR 2016*, pp. 395–398, IEEE, Shenzhen, China, October 2016.
- [34] J. Tan, N. Bi, C. Y. Suen, and N. Nobile, "Multi-feature selection of handwriting for gender identification using mutual information," in *Proceedings of the 15th International Conference on*

- Frontiers in Handwriting Recognition, ICFHR 2016*, pp. 578–583, IEEE, Shenzhen, China, October 2016.
- [35] Y. Akbari, K. Nouri, J. Sadri, C. Djeddi, and I. Siddiqi, “Wavelet-based gender detection on off-line handwritten documents using probabilistic finite state automata,” *Image and Vision Computing*, vol. 59, pp. 17–30, 2017.
- [36] C. Djeddi, A. Gattal, L. Souici-Meslati, I. Siddiqi, Y. Chibani, and H. El Abed, “LAMIS-MSHD: a multi-script offline handwriting database,” in *Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition, ICFHR 2014*, pp. 93–97, IEEE, Heraklion, Greece, September 2014.
- [37] N. Bouadjenek, H. Nemmour, and Y. Chibani, “Writer’s gender classification using HOG and LBP features,” *Lecture Notes in Electrical Engineering*, vol. 411, pp. 317–325, 2017.
- [38] R. N. Morris, *Forensic Handwriting Identification: Fundamental Concepts and Principles*, Academic Press, Cambridge, Massachusetts, USA, 2000.
- [39] V. Saran, S. Kumar, A. K. Gupta, and S. Ahmad, “Differentiation of handedness of writer based on their strokes and characteristic features,” *Journal of Forensic Research*, vol. 4, no. 5, Article ID 1000204, pp. 1–3, 2013.
- [40] S. Al-Maadeed, F. Ferjani, S. Elloumi, A. Hassaine, and A. Jaoua, “Automatic handedness detection from off-line handwriting,” in *Proceedings of the 2013 7th IEEE GCC Conference and Exhibition, GCC 2013*, pp. 119–124, IEEE, Doha, Qatar, November 2013.
- [41] S. Al-Maadeed, F. Ferjani, S. Elloumi, and A. Jaoua, “A novel approach for handedness detection from off-line handwriting using fuzzy conceptual reduction,” *Eurasip Journal on Image and Video Processing*, vol. 2016, no. 1, article no. 1, pp. 1–14, 2016.
- [42] N. Bouadjenek, H. Nemmour, and Y. Chibani, “Age, gender and handedness prediction from handwriting using gradient features,” in *Proceedings of the 13th International Conference on Document Analysis and Recognition, ICDAR 2015*, pp. 1116–1120, IEEE, Tunis, Tunisia, August 2015.
- [43] L. Xing and Y. Qiao, “DeepWriter: A multi-stream deep CNN for text-independent writer identification,” in *Proceedings of the 15th International Conference on Frontiers in Handwriting Recognition, ICFHR 2016*, pp. 584–589, China, October 2016.
- [44] A. Bartle and J. Zheng, “Gender classification with deep learning,” Stanford cs224d Course Project Report, 2015.
- [45] G. Levi and T. Hassner, “Age and gender classification using convolutional neural networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2015*, pp. 34–42, IEEE, Boston, MA, USA, June 2015.
- [46] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [47] S. Albelwi and A. Mahmood, “A framework for designing the architectures of deep convolutional neural networks,” *Entropy*, vol. 19, pp. 1–20, 2017.
- [48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Proc. Advances in Neural Information Processing Systems (NIPS’12)*, vol. 25, pp. 1097–1105, 2012.
- [49] F. Khosrowshahi, “Innovation in artificial neural network learning: Learn-On-Demand methodology,” *Automation in Construction*, vol. 20, no. 8, pp. 1204–1210, 2011.
- [50] S. Zhang, C. Zhang, and Q. Yang, “Data preparation for data mining,” *Applied Artificial Intelligence*, vol. 17, no. 5-6, pp. 375–381, 2003.
- [51] U.-V. Marti and H. Bunke, “The IAM-database: an english sentence database for offline handwriting recognition,” *International Journal on Document Analysis and Recognition*, vol. 5, no. 1, pp. 39–46, 2003.
- [52] M. Liwicki and H. Bunke, “IAM-OnDB - An on-line English sentence database acquired from handwritten text on a whiteboard,” in *Proceedings of the 8th International Conference on Document Analysis and Recognition*, pp. 956–961, IEEE, Seoul, South Korea, September 2005.
- [53] IAM Homepage, <http://www.fki.inf.unibe.ch/databases/iam-on-line-handwriting-database>.
- [54] S. A. Mahmoud, I. Ahmad, M. Alshayeb et al., “KHATT: Arabic offline Handwritten Text Database,” in *Proceedings of the 13th International Conference on Frontiers in Handwriting Recognition, ICFHR 2012*, pp. 449–454, IEEE, Bari, Italy, September 2012.
- [55] KHATT Homepage, <http://khatt.ideas2serve.net>.
- [56] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.

