Data Driven Feature Extraction for Gender Classification using Multi-script Handwritten Texts

Momina Moetesum*, Imran Siddiqi*, Chawki Djeddi †, Yaacoub Hannad ‡, Somaya Al-Maadeed §
* Bahria University, Islamabad, Pakistan
Email: imran.siddiqi@bahria.edu.pk
†Larbi Tebessi University, Tebessa, Algeria.
‡Ibn Tofail University, Morocco.
§Qatar University, Qatar

***Abstract*—This paper presents a study on assessing the effectiveness of machine learned features to predict gender of writers from images of handwriting. Pre-trained Convolutional Neural Networks have been employed as feature extractors to discriminate male and female handwriting while classification is carried out using a number of classifiers, Linear Discriminant Analysis (LDA) being the most effective. Feature extraction is carried out by changing the scale of observation using word, patch and page images. Experiments are carried out on English and Arabic handwriting samples of the QUWI database and the realized results demonstrate the effectiveness of machine learned features in predicting gender from handwriting.**

Keywords*—*Gender Classification; Convolutional Neural Networks; Handwriting, Multi-scrip Text.

I. Introduction

Handwriting and hand-drawn shapes have long been studied by forensic experts, document examiners, neurologists, psychologists and paleographers [5], [14], [27], [41] thanks primarily to the wide variety of applications they offer. Handwriting is a complex fine motor skill [39], [18], [13] that involves a combination of cognitive and psychomotor processes [49]. A number of studies [20], [28], [35], [47] have established the fact that handwriting is indicative of rich information about the writer. This not only permits handwriting to be employed as an effective behavioral biometric modality [39], [43] but also facilitates research targeting a number of interesting applications. The most significant of these is the study of correlation between handwriting and different neurological disorders including autism [19], [30] Parkinson [48] and Alzheimer [40] etc. The changing patterns in handwriting as a function of aging [50], [38] and under the effect of medications [13] have also been investigated. Furthermore, attempts have been made to predict different personal attributes of writers from handwriting images [27], [41]. The validity of these graphological studies, however, has remained debatable [20], [28], [35], [47]. Consequently, the experts in neuro or forensic sciences generally distant themselves from such studies. The widely accepted and established correlation is between handwriting and writer demographics especially the gender of writer [6], [12], [23], [24], [25], [26], [10], [36], [46]. The differences in motor control [15], [51], [17] as well as the varied learning rate of motor skills [11] in the two gender categories lead to different writing styles in male and female writers.

The early research on prediction of gender from handwriting [23], [25], [26] primarily focused on exploration of discriminative attributes through (human) examination of writing samples. A notable development was the work of Hamid et al. [24] who presented quantitative results on classification of gender from handwriting by human experts. Experiments on 30 samples in two different scripts reported 68% recognition rate. Computerized systems on this problem mainly rely on the visual differences [12], [25] in the writing styles of male and female writers (Figure 1). Such systems algorithmically compute global or local features from given set of handwriting samples to characterize the gender.

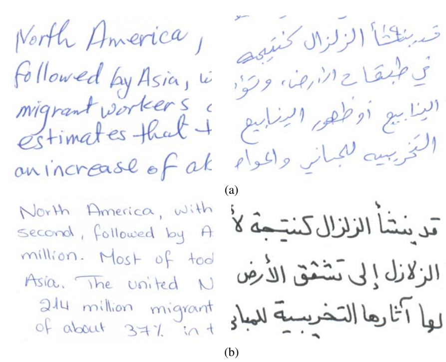

Fig. 1. Sample images of (a): Male and (b): Female writings in the QUWI database

Cha et al. [14] presented one of the earliest computerized systems for classification of demographic attributes of individuals from handwriting. A combination of different micro and macro features reported 70% classification rate. The work was later extended [5] to enhance the classification rates using boosting on multiple neural networks. Liwicki et al. [32] exploited a combination of multiple features to classify gender and handedness from online writing samples. Siddiqi et al. [44] capture the orientation, curvature and textural information in handwriting through a combination of features and report results on a comprehensive series of experiments using QUWI and MSHD databases. Likewise, combination of

978-1-5386-5875-8/18/$31.00 ©2018 IEEE
DOI 10.1109/ICFHR-2018.2018.00104

multiple descriptors is evaluated in [45] on a custom developed database of male and female writing samples.

Among other notable contributions, geometric features are employed in [4] to classify gender, age group and handedness of writers. Experiments on writing samples of the QUWI database realized classification rates of up to 74%. Bouadjenek et al. [7] capture the textural information in handwriting using Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). Experiments on writing samples of IAM-onDB database with SVM classifier reported 74% classification rate. The system was later extended [9] to include classification of handedness as well as age in addition to gender. Likewise, in addition to HOG and LBP, authors also employed Gradient Local Binary Patterns (GLBP) [8] to classify gender from images of handwriting. In other similar studies, textural information is extracted using a bank of Gabor filters by Mirza et al. [33], Youssef et al. [52] consider gradient and wavelet domain LBP while oriented basic image features (oBIFs) are employed in [21]. Similarly, Akabari et al. [2] consider wavelet sub-bands to characterize gender from handwriting using SVM classifier.

In a recent study, Morera et al. [34] consider gender, handedness, and combined gender-and-handedness classification using convolutional neural networks. A CNN with two convolutional layers is trained on word images of English and Arabic writing samples producing a 512 dimensional feature vector. Authors report classification rates of 80.72% and 68.90% on IAM and KHATT databases respectively. While most of the reported studies target the enhancement of feature extraction step (and employ traditional classifiers), Ahmed et al. [1] focus on the classification part. Traditional textural features including LBP, HOG and GLCM are considered while for classification a number of classifiers are investigated. Classifiers are combined using bagging, boosting and stacking. Experiments on the database of ICDAR 2015 competition on gender classification from handwriting [16] reported classification rates varying from 79% to 85%.

This paper presents a comprehensive study to evaluate the effectiveness of machine learned features in characterizing gender from handwriting. More specifically, we target a multi-script environment using Arabic and English handwriting samples. A pre-trained convolutional neural network is employed as feature extractor. Features are extracted at word, patch and page level. A patch represents a small part of handwriting image comprising 4-5 lines and the same number of words per line. These different images correspond to different scales of observation. For classification, we have employed a number of classifiers including Naive Bayes (NB), Decision Trees (DTA), Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM). Experiments are carried out in text-dependent as well as text-independent mode. Classification performance is also studied in a script independent mode where training and test samples come from different scripts. The realized classification rates demonstrate the effectiveness of the proposed technique to characterize gender from handwriting. The key research questions considered in the present study include the following.

- How the current state-of-the-art convolutional neural networks based feature extraction can be adapted for classification of gender from handwriting and what is the performance of such systems?
- What is the impact of scale of observation in characterizing gender from handwriting?
- Are the classification rates sensitive to the content of writing samples (text-independent vs. text-dependent evaluations)?
- Do individuals from a given gender share common characteristics across multiple scripts (script-independent evaluations)?

In the next section, we present in detail the feature extraction and classification techniques. Section III presents the details of dataset, experimental protocols and realized results along with accompanying discussion and a comparison with existing techniques. Finally, Section IV summarizes the key findings of this study with a discussion on potential directions of further research on the subject.

II. Proposed Technique

This section details the feature extraction and classification techniques employed in our study. Features are extracted using a pre-trained CNN while for classification we investigated a number of standard classifiers. Classification is carried out at word , patch and complete page levels that correspond to different observation scales. Details on feature extraction and classification are presented in the following.

A. Feature Extraction using CNNs

In the context of recognition systems, the last few years have witnessed a shift in paradigm from conventional, domain-specific hand-engineered features to automatically extracted machine-learned features. The superiority of such machine extracted features (typically through CNNs) over traditional representations has been validated for a variety of classification tasks.

Convolutional Neural Networks, came to scene, for the first time in early 1990s [31]; the current fame, however, is primarily attributed to Krizhevsky et al. [29] where the CNN based system significantly reduced the error rate in the ImageNet Large Scale Visual Recognition competition. Since then, CNNs have been widely employed to a large number of recognition problems outperforming the conventional solutions. A typical CNN mainly comprises convolutional and pooling layers which are followed by fully connected layer(s). The convolutional layers serve as feature extractors and perform convolution of input image (volume) with different filters while the pooling layers down sample the filter outputs to reduce the dimensionality and avoid over-fitting problems. The fully connected layers serve as classifier and the learning process involves finding the optimal set of filters for each layer. While CNNs can be trained from scratch for a given problem, it is also common to employ pre-trained CNNs either as feature extractors [22] or fine tune them to a specific problem [37], [42] by continuing back propagation.

TABLE I. SUMMARY OF WELL-KNOWN GENDER CLASSIFICATION METHODS

Study	Year	Features	Classifier	Database	Classification Rate
Cha & Srihari [14]	2001	Combination of macro & micro features	ANN	CEDAR	70.20%
Liwicki et al. [32]	2011	Online & offline features	GMM	IAM-OnDB	67.57%
Sokic et al. [45]	2012	Shape Descriptors		BHDH	
Siddiqi et al. [44]	2012	Textural and Structural features	SVM	QUWI & MSHD	68.75% / 73.02%
Al-Maadeed & Hassaine [4]	2014	Geometrical Features	Random Forests	QUWI	73%
Bouadjenek et al. [7]	2014	HOG & LBP	SVM	IAM	74%
Youssef et al. [52]	2013	Gradient & LBP	SVM	QUWI	74.30%
Mirza et al. [33]	2016	Gabor filters	ANN	QUWI	70%
Akbari et al. [2]	2017	Wavelet sub-bands	SVM/ANN	QUWI & MSHD	80%
Ahmed et al. [1]	2017	LBP, HOG, GLCMs	Ensemble of classifiers	QUWI	79 - 85%
Morera et al. [34]	2018	CNNs		IAM/KHATT	80.72% / 68.90%
Gattal et al. [21]	2018	oBIFs	SVM	QUWI	66% - 81%

While a number of very deep pre-trained CNNs (ZF Net, VGG Net, GoogLeNet, ResNet etc.) have been made publicly available, it is important to mention that we target a two-class gender prediction problem. Morera et al. [34] have also demonstrated that with only two convolutional layers, the system reports high classification rates. Consequently, we have chosen to employ AlexNet as pre-trained network that comprises 5 convolutional and 3 fully connected layers. The network is employed as feature extractor only and the features are extracted at $fc7$ layer.

In an attempt to study the impact of scale of observation on the classification task, we extract features at word, patch and page levels. Words are extracted from binarized images using run-length-smoothing algorithm followed by connected component labeling. While the word segmentation results are acceptable in case of English writing samples, the Arabic samples often result in over or under segmented words due to varied inter and intra word distances. However, it is important to point out that perfect word segmentation is not an objective. The idea is to extract a small part of writing the corresponds to one-two words representing a closer scale of observations. In addition to words, we also extract small patches from the handwriting images. Each patch is 20% of text width/height and typically contains portions of 4-5 text lines and corresponds to a relatively distant observation scale. Finally, the complete image of handwriting is also considered as input to the CNN and corresponds to the most distant scale of observation. Each word, patch or page image is resized to 227×227 and is fed to the CNN which outputs a 4096 dimensional feature vector. Activation maps corresponding to word and patch images for a random layer in the CNN are presented in Figure 2.

B. Classification

Features extracted from word (patch or page) images of male and female writing samples are fed to multiple classifiers. The classifiers investigated in our study include Naive Bayes (NB), Linear Discriminant Analysis (LDA), Decision Trees (DT) and Support Vector Machine (SVM). Classifiers are trained separately for each scale of observation. The performance of these classifiers in different experiments is presented in the following section.

III. EXPERIMENTS AND RESULTS

This section presents the details of the experiments carried out to assess the performance of the features learned as well as classifiers under different experimental scenarios. We first present the details of the dataset employed in our study

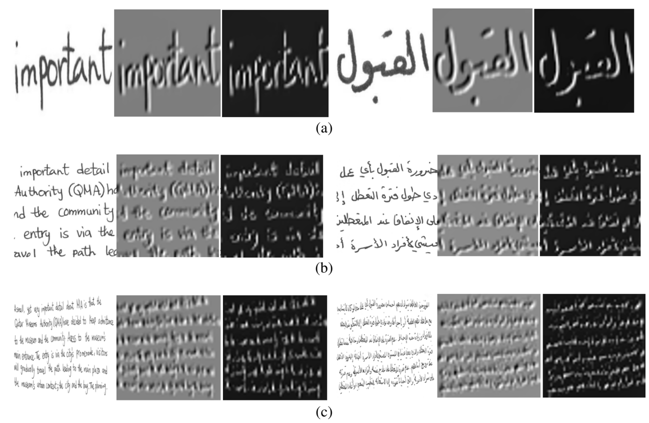

Fig. 2. Features maps of a CNN layer and ReLU outputs on Arabic and English (a): word, (b): patch and (c): page images

followed by a discussion on the experiments conducted and the realized results.

A. Dataset

The experimental study of our system is carried out on the QUWI database [3] that contains Arabic and English writing samples collected from 1017 writers. Each writer contributed four samples in Arabic (2 samples) & English (2 samples). The writing samples contain same as well as arbitrary text for each writer both in English and Arabic. This allows employing the database in text-dependent as well as text-independent modes. In all of our evaluations, we employ writing samples of 1000 writers, 700 in the training set and 300 in the test. It is also important to point out that although we target classification of gender, it is ensured that the training and test sets do not have any writer in common so that the evaluations truly correspond to characterizing gender and not writer.

B. Experiments, Results & Discussion

First, we present the results of experiments on the complete QUWI database with 700 writers ($700 \times 4 = 2800$ samples) in the training set and 300 writers ($300 \times 4 = 1200$ samples) in the test set. The classification rates against different classifiers are summarized in Table II where it can be seen that although performance of different classifiers is comparable, LDA outperforms other classifiers. Consequently, for subsequent experiments, we report the results with LDA only. It is also interesting to note that for all classifiers, patch level classification rates are better than those of word or complete page images. This observation seems to be very much natural as single word is too close a scale of observation and the amount of text is fairly limited. Likewise, complete page represents a very distant scale of observation. A small patch of handwriting with few lines of text and few words per line represents a good scale of observation that effectively characterizes gender from handwriting. The highest classification rate achieved is 70.08% when using patch images with LDA classifier.

TABLE II. CLASSIFICATION RATES ON COMPLETE QUWI DATABASE

	Classifier			
Image Scale	LDA	NB	SVM	DT
Word	67.75%	65.41%	63.66%	64.33%
Patch	70.08%	67.16%	65.03%	65.58%
Page	68.50%	66.75%	64.66%	64.33%

1) Text-Dependent vs. Text-Independent Evaluations: These experiments aim to study the performance variation as a function of textual content in the training and test sets. For text-dependent experiments, Page 2 (Page 4) of 700 writers is used in the training set and same pages of 300 writers in the test set. For text-independent experiments, we employ Page 1 (Page 3) of each writer so that training and test sets contain different textual content. The classification rates of these experiments are summarized in Table III. It can be seen from the table that similar to the performance on the complete database, the patch level images realize the highest classification rates. It can also be observed that the performance of text-independent and text-dependent modes is very much comparable with only marginal differences. In general, the classification rates on English writing samples are better than those on the Arabic text and the difference is more pronounced at word level.

TABLE III. CLASSIFICATION RATES: TEXT-DEPENDENT VS. TEXT-INDEPENDENT EVALUATIONS

	Text-Dependent			Text-Independent		
Dataset	Word	Patch	Page	Word	Patch	Page
QUWI-English	68.33%	73.33%	70.33%	67.33%	72.00%	69.33%
QUWI-Arabic	64.66%	71.66%	69.66%	65.33%	70.66%	66.33%

2) Scrip-Dependent vs. Script-Independent Evaluations: In these evaluations, we analyze the effect of script of text on the classification performance. In script-dependent experiments we employ the first two pages (last two pages for English) of 700 writers in the training and 300 writers in the test set. In script-independent experiments, the training and test sets contain writing samples in different scripts (one set is Arabic other is English). The results of these experiments are presented in Table IV where it can be seen that script-dependent experiments report higher classification rates as opposed to scrip-independent evaluations. This is very much natural as having different samples in different scripts in the training and test sets is a challenging experimental scenario. Classification rates of 65.16% and 64.83% (patch level) for these experiments are indeed very promising.

3) Comparison & Discussion: We also present a performance comparison of the proposed technique with existing systems evaluated on the same database. It is however important to point out the most of the recent work on this problem evaluated using QUWI database considers the experimental protocol of ICDAR 2015 gender classification competition [16]. The competition comprised four tasks. Tasks A and Tasks B comprised writing samples in Arabic and English respectively. Tasks C & D correspond to script-independent evaluations with Arabic (English) samples in the training and English (Arabic) samples in the test set. The training set in each of the tasks comprises 300 writing samples while the validation and test sets comprise 100 samples each. In addition to the experimental settings presented earlier, we also evaluated the CNN based feature extraction followed by LDA classification using the competition protocol. The classification rates of the participating systems, other studies evaluated using the same protocol as well as those realized by the proposed technique are summarized in Table V.

TABLE IV. CLASSIFICATION RATES: SCRIPT-DEPENDENT VS. SCRIPT-INDEPENDENT EVALUATIONS

Training Set	Test Set	Word	Patch	Page
QUWI-English	QUWI-English	68.50%	71.50%	70.83%
QUWI-Arabic	QUWI-Arabic	67.66%	69.83%	68.83%
QUWI-English	QUWI-Arabic	63.50%	65.16%	64.50%
QUWI-Arabic	QUWI-English	60.83%	64.83%	63.83%

TABLE V. COMPARISON USING EVALUATION PROTOCOL OF ICDAR 2015 GENDER CLASSIFICATION COMPETITION ([16])

Technique	Result (%) (Rank)			
	Task A	Task B	Task C	Task D
Participants of ICDAR 2015 Competition				
LISIC	60(3)	42(8)	49(5)	55(2)
ACIRS	60(3)	54(3)	53(3)	49(6)
Nuremberg	62(2)	**60(1)**	55(2)	53(3)
MCS-NUST	47(7)	51(5)	48(6)	45(8)
CVC	**65(1)**	57(2)	**63(1)**	**58(1)**
QU	44(8)	52(4)	53(3)	47(7)
UBMA	51(5)	50(6)	44(7)	50(5)
ESI-STIC	48(6)	46(7)	42(8)	53(3)
Other studies evaluated using same protocol				
Mirza et al. [33]	70	67	69	63
Ahmed et al. [1]	79	85	80	81
Gattal et al. [21]	78	81	76	76
Proposed Technique	76	77	71	73

Table V shows that the best results on the ICDAR 2015 experimental protocol are reported by Ahmed et al. [1] where different ensemble classifiers have been employed. It is also important to mention that the experimental protocol of the competition only considered 100 writers in the test set. With the increase in number of writers in the test set, the performance naturally drops. Scalability of the classification technique to larger datasets is indeed an important aspect. We, therefore, also present a comparison of the techniques evaluated on larger datasets. Al-Maadeed & Hassaine [4] considered the complete set of 1017 writers in the QUWI database while Morera et al. [34] employed 1000 writers of the KHATT database. The performance comparison of these techniques with the proposed system is presented in Table VI where the highest classification rate is reported to be 73% on 1017 writers of the QUWI database. It is also important the mention that Morera et al. [34] also employs a convolutional neural networks using word images. We, on the other hand, report results at patch level which proved to be the optimal scale of observation in our experiments.

IV. CONCLUSION

This paper presented an experimental study to evaluate the performance of state-of-the-art deep convolutional neural network based feature extraction to classify gender from offline

TABLE VI. PERFORMANCE COMPARISON OF TECHNIQUES EVALUATED ON LARGE DATASETS

Study	Database	Number of writers	Classification Rate
Al-Maadeed & Hassaine [4]	QUWI	1017	73%
Morera et al. [34]	KHATT	1000	68.90%
Proposed Technique	QUWI	1000	70.08%

handwriting. Features were extracted from writing samples by changing the scale of observation, i.e. word, patch and complete page. A pre-trained ConvNet (AlexNet) was employed as feature extractor. For classification, we investigated a number of standard classifiers; among these linear discriminant analysis reproted the highest classification rates. We carried out a comprehensive series of experiments including text-dependent, text-independent, script-dependent and script-independent experiments. For comparison purposes, we also evaluated the system on the database and experimental protocol of the ICDAR 2015 competition. The realized classification rates demonstrated the effectiveness of machine learned features in characterizing gender from writing samples.

The present work employs only a single pre-trained CNN model as feature extractor. It would be interesting to study how the performance varies by changing the model. A comprehensive series of experiments can be carried out by using other pre-traiend models. Furthermore, in addition to feature extraction, pre-trained models can also be fine tuned by continuing back propagation on the handwriting images. Another direction could be to train a network from scratch at word, patch or image levels. This will allow deeper insights into what writing features contribute to characterize gender. Using pre-trained networks constraint the size of input images. Text lines, for instance, cannot be directly fed to most of the pre-trained models as resizing them to match the input size disturbs the aspect ratio. Training a customized network will also alleviate such problems. In addition, prediction of other demographic attributes like age or handedness etc. can also be explored.

REFERENCES

[1] Mahreen Ahmed, Asma Ghulam Rasool, Hammad Afzal, and Imran Siddiqi. Improving handwriting based gender classification using ensemble classifiers. *Expert Systems with Applications*, 85:158–168, 2017.

[2] Younes Akbari, Kazem Nouri, Javad Sadri, Chawki Djeddi, and Imran Siddiqi. Wavelet-based gender detection on off-line handwritten documents using probabilistic finite state automata. *Image and Vision Computing*, 59:17–30, 2017.

[3] S. Al Ma'adeed, W. Ayouby, A. Hassaine, and J.M. Aljaam. Quwi: An arabic and english handwriting dataset for offline writer identification. In *In Proc. of 13th Intl. Conf. on Frontiers in Handwriting Recognition (ICFHR)*, pages 746–751, 2012.

[4] Somaya Al Maadeed and Abdelaali Hassaine. Automatic prediction of age, gender, and nationality in offline handwriting. *EURASIP Journal on Image and Video Processing*, (1):1–10, 2014.

[5] Karthik R Bandi and Sargur N Srihari. Writer demographic classification using bagging and boosting. In *In Proc. 12th Int. Graphonomics Society Conference*, pages 133–137, 2005.

[6] J Beech and I Mackintosh. Do differences in sex hormones affect handwriting style? evidence from digit ratio and sex role identity as determinants of the sex of handwriting. *Personality and Individual Differences*, 39(2):459–468, 2005.

[7] Nesrine Bouadjenek, Hassiba Nemmour, and Youcef Chibani. Local descriptors to improve off-line handwriting-based gender prediction. In *Proc. of the 6th International Conference on Soft Computing and Pattern Recognition (SoCPaR)*, pages 43–47, 2014.

[8] Nesrine Bouadjenek, Hassiba Nemmour, and Youcef Chibani. Age, gender and handedness prediction from handwriting using gradient features. In *Proc. of the 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1116–1120, 2015.

[9] Nesrine Bouadjenek, Hassiba Nemmour, and Youcef Chibani. Histogram of oriented gradients for writer's gender, handedness and age prediction. In *Proc. of International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)*, 2015.

[10] M. E. Broom, B. Thompson, and M. T. Bouton. Sex differences in handwriting. *Journal of Applied Psychology*, 13(2):159–166, 1929.

[11] Manuel M Buitrago, Thomas Ringer, Jörg B Schulz, Johannes Dichgans, and Andreas R Luft. Characterization of motor skill and instrumental learning time scales in a skilled reaching task in rat. *Behavioural brain research*, 155(2):249–256, 2004.

[12] V. Burr. Judging gender from samples of adult handwriting: accuracy and use of cues. *Journal of Social Psychology*, 142(6):691–700, 2002.

[13] Michael P Caligiuri and Linton A Mohammed. *The neuroscience of handwriting: Applications for forensic document examination*. CRC Press, 2012.

[14] Sung-Hyuk Cha and S.N. Srihari. A priori algorithm for sub-category classification analysis of handwriting. In *In Proc. of the 6th Intl. Conf. on Document Analysis and Recognition*, pages 1022–1025, 2001.

[15] Marsha R Cohen. Individual and sex differences in speed of handwriting among high school students. *Perceptual and motor skills*, 84:1428–1430, 1997.

[16] Chawki Djeddi, Somaya Al-Maadeed, Abdeljalil Gattal, Imran Siddiqi, Labiba Souici-Meslati, and Haikal El Abed. Icdar2015 competition on multi-script writer identification and gender classification using quwidatabase. In *Proc. of the 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1191–1195, 2015.

[17] Shoshi Dorfberger, Esther Adi-Japha, and Avi Karni. Sex differences in motor performance and motor learning in children and adolescents: an increasing male advantage in motor learning and consolidation phase gains. *Behavioural brain research*, 198(1):165–171, 2009.

[18] Katya P Feder and Annette Majnemer. Handwriting development, competency, and intervention. *Developmental Medicine & Child Neurology*, 49(4):312–317, 2007.

[19] Christina T Fuentes, Stewart H Mostofsky, and Amy J Bastian. Children with autism show specific handwriting impairments. *Neurology*, 73(19):1532–1537, 2009.

[20] Adrian Furnham and Barrie Gunter. Graphology and personality: Another failure to validate graphological analysis. *Personality and Individual Differences*, 8(3):433–435, 1987.

[21] Abdeljalil Gattal, Chawki Djeddi, Imran Siddiqi, and Youcef Chibani. Gender classification from offline multi-script handwriting images using oriented basic image features (obifs). *Expert Systems with Applications*, 99:155–167, 2018.

[22] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[23] F.L. Goodenough. Sex differences in judging the sex of handwriting. *Journal of Social Psychology*, 22:61–68, 1945.

[24] S. Hamid and K.M. Loewenthal. Inferring gender from handwriting in urdu and english. *Journal of Social Psychology*, 136(6):778–782, 1996.

[25] James Hartley. Sex differences in handwriting: A comment on spear. *British Educational Research Journal*, 17(2):141–145, 1991.

[26] William N. Hayes. Identifying sex from handwriting. *Perceptual and Motor Skills*, 83:91–800, 1996.

[27] Roy N. King and Derek J. Koehler. Illusory correlations in graphological inference. *Journal of Experimental Psychology:Applied*, 6(4):336–348, 2000.

[28] Richard J. Klimoski and Anat Rafaeli. Inferring personal qualities through handwriting analysis. *Journal of Occupational Psychology*, 56(3):191–202, 1983.

[29] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[30] Azadeh Kushki, Tom Chau, and Evdokia Anagnostou. Handwriting difficulties in children with autism spectrum disorders: A scoping review. *Journal of autism and developmental disorders*, 41(12):1706–1716, 2011.

[31] Yann LeCun, Bernhard E Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne E Hubbard, and Lawrence D Jackel. Handwritten digit recognition with a back-propagation network. In *Advances in neural information processing systems*, pages 396–404, 1990.

[32] Marcus Liwicki, Andreas Schlapbach, and Horst Bunke. Automatic gender detection using on-line and off-line information. *Pattern Analysis and Applications*, 14(1):87–92, 2011.

[33] Ali Mirza, Momina Moetesum, Imran Siddiqi, and Chawki Djeddi. Gender classification from offline handwriting images using textural features. In *Frontiers in Handwriting Recognition (ICFHR), 2016 15th International Conference on*, pages 395–398, 2016.

[34] Ángel Morera, Ángel Sánchez, José Francisco Vélez, and Ana Belén Moreno. Gender and handedness prediction from offline handwriting using convolutional neural networks. *Complexity*, 2018, 2018.

[35] Efrat Neter and Gershon Ben-Shakhar. The predictive validity of graphological inferences: A meta-analytic approach. *Personality and Individual Differences*, 10(7):737–745, 1989.

[36] S. M. Newhall. Sex differences in handwriting. *Journal of Applied Psychology*, 10(2):151–161, 1926.

[37] Hong-Wei Ng, Viet Dung Nguyen, Vassilios Vonikakis, and Stefan Winkler. Deep learning for emotion recognition on small datasets using transfer learning. In *Proceedings of the 2015 ACM on international conference on multimodal interaction*, pages 443–449. ACM, 2015.

[38] Sara Rosenblum, Batya Engel-Yeger, and Yael Fogel. Age-related changes in executive control and their relationships with activity performance in handwriting. *Human movement science*, 32(2):363–376, 2013.

[39] Lambert Schomaker. Advances in writer identification and verification. In *Proc. of the 9th International Conference on Document Analysis and Recognition*, volume 2, pages 1268–1273, 2007.

[40] A Schröter, R Mergl, K Bürger, H Hampel, H-J Möller, and Ulrich Hegerl. Kinematic analysis of handwriting movements in patients with alzheimers disease, mild cognitive impairment, depression and healthy subjects. *Dementia and Geriatric Cognitive Disorders*, 15(3):132–142, 2003.

[41] Viv Shackleton and Sue Newel. European management selection methods: A comparison of five countries. *International Journal of Selection and Assessment*, 2(2):91–102, 1994.

[42] Arjun Sharma et al. Adapting off-the-shelf cnns for word spotting and recognition. In *Proc. of 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 986–990, 2015.

[43] I. Siddiqi and N. Vincent. Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. *Pattern Recognition*, 43(11):3853 – 3865, 2010.

[44] Imran Siddiqi, Chawki Djeddi, Ahsen Raza, and Labiba Souici-meslati. Automatic analysis of handwriting for gender classification. *Pattern Analysis and Applications*, pages 1–13, 2014.

[45] E. Sokic, A. Salihbegovic, and M. Ahic-Djokic. Analysis of off-line handwritten text samples of different gender using shape descriptors. In *In Proc. of IX International Symposium on Telecommunications (BIHTEL)*, pages 1–6, 2012.

[46] H. Tenwolde. More on sex differences in handwriting. *Journal of Applied Psychology*, 18:705–710, 1934.

[47] Robert P. Tett and Cynthia A. Palmer. The validity of handwriting elements in relation to self-report personality trait measures. *Personality and Individual Differences*, 22(1):11–18, 1997.

[48] Hans-Leo Teulings and George E Stelmach. Control of stroke size, peak acceleration, and stroke duration in parkinsonian handwriting. *Human Movement Science*, 10(2):315–334, 1991.

[49] Gerard P Van Galen. Handwriting: Issues for a psychomotor theory. *Human movement science*, 10(2):165–191, 1991.

[50] Judie Walton. Handwriting changes due to aging and parkinson's syndrome. *Forensic science international*, 88(3):197–214, 1997.

[51] Naomi Weintraub, Anat Drory-Asayag, Rivki Dekel, Hanna Jokobovits, and Shula Parush. Developmental trends in handwriting performance among middle school children. *OTJR: Occupation, Participation and Health*, 27(3):104–112, 2007.

[52] A. E. Youssef, A. S. Ibrahim, and A. L. Abbott. Automated gender identification for arabic and english handwriting. In *5th International Conference on Imaging for Crime Detection and Prevention (ICDP 2013)*, 2013.