UDC 004.9 **doi: 10.26906/SUNZ.2025.1.138-142**

M. Shupyliuk, V. Martovytskyi

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

ANALYSIS OF PERSONALITY DETECTION AND WRITER IDENTIFICATION METHODS

Abstract. Handwritten text as multi-sensory activity can show one's personality and at the same time can serve as one's biometric identifier. Handwriting analysis is used in various fields including history, forensic, education, security, personnel matters etc. In this article handwriting analysis methodologies were considered and categorized in four groups highlighting advantages and disadvantages of each group. Also, this article depicts various problems associated with developing handwriting analysis systems such as improper feature extraction, overfitting, underfitting, unreliable training data, picking model for assessing personality types, etc. Both methods for robust offline writer identification and methods for prediction of human personality that are used in state-of-the-art handwriting analysis systems are presented. In addition, current studies and common approaches for performance measurement and database selection in both writer identification and personality detection fields were analyzed. Also, perspective development directions of modern handwriting analysis systems are presented.

Keywords: handwriting, handwriting analysis, personality, handwriting features, graphology, personality detection, writer identification.

Introduction

Handwriting analysis systems aim to automatically extract important information from handwriting, therefore research in the field of graphology has a wide range of applications, such as writer identification, personality detection etc. M. Hengl [1], Hemlata et al. [2], Wirmanto et al. [15], Samsuryadi et al. [31] presented an overview of modern methods of handwriting analysis that are used in various fields. This paper presents some challenges in handwriting analysis systems, the motivation for further research, application areas, design approaches and issues, writer identification and personality detection methods. In addition, perspective development directions are presented.

Formulation of the problem and its connection with important scientific or practical tasks

Digital transformation is a continuous activity and affects various aspects of human life. Despite that fact handwriting remains a core human activity and considered as one of the most complex human activities. Everyone's handwriting is unique and has characteristics that are typical of the given individual which creates a basis for handwriting analysis [1]. Handwriting analysis is a scientific method for recognizing, assessing and understanding a writer's personality and identity through the shapes and word patterns in the handwriting [2].

Handwriting analysis has three major branches: graphology, handwriting expert and forensic linguistics [1]. The main fields of application for handwriting analysis are criminalistic, education, history, mental hygiene, medical diagnosis, public life, personnel matters and personality profiles. In the criminalistic field handwriting analysis is associated with crimes that involve written documents such as kidnapping, defamation, false witness, false accusations, violation of rules for accounting and forgery of public documents.

Detecting an individual based on handwriting is a kind of behavior biometric identification well acknowledged by psychologists, neurologists, paleographers, forensic analysts, document analysts, and computer science researchers. Similarly, the existence of a relationship between handwriting and different demographic attributes of writers, such as gender, handedness, and age, is also confirmed by psychologists and neurologists [3]. So, handwriting is also called brain writing. Each personality trait has a neurological brain pattern in the human brain. Each neurological brain pattern design delivers one of a kind neuromuscular movement which is the same for each individual who has that specific personality trait. Each stroke or movement in handwriting reveals a particular personality trait [2].

Writer identification and personality detection have been studied for many years but they remain an open research problem for such reasons as diversity, similarity and high intra-variability among authors in handwriting. Digital era brings new challenges into writer identification and personality detection because now text can be handwritten not only on paper but also on digital devices. This leads to the necessity to differentiate between online and offline. In online writer identification and personality detection data contains temporal information about the text formation [4]. In contrast, offline writer identification and personality detection deals only with the handwritten text itself without any additional information [4]. This paper mainly focuses on the offline handwriting analysis process to detect the personality of a person and perform writer identification based on their handwritten documents. Section III describes research results, where subsection I describes categorized handwriting analysis methodologies, subsection II describes the related works in the field of handwriting analysis for writer identification, subsection III describes related works for personality detection, subsection IV provides discussion topics and future scope, section IV gives conclusion.

Research results

Handwriting analysis methodologies. Handwriting analysis is a complex problem so there have been many studies proposed. To categorize these studies and identify advantages and disadvantages different aspects can be used.

Considering the type of features as a functional context, the methods reported in the literature can be categorized into four main groups: structural-based, textural-based, grapheme-based and auto-learned based methods [17].

Structural-based methods consider the allograph shapes and then apply a grapheme-emission probability distribution. For structural-based systems cost of runtime is high due to complexity of preprocessing steps which include segmentation, binarization and edge detection. Strong reliance on preprocessing is a limitation of these methods, since preprocessing step failure means feature extraction and classification steps are prone to failure. Concentration of such methods mainly on allograph shapes means that additional information drawn in the same word between allographs is omitted. Additional drawback of such methods is vulnerability to any variations in the character or allograph characteristics (i.e. slant, aspect ratio).

Textural-based (transformation-based) methods work with digitized image of handwriting which are considered as special textures and extract features(after some transformation techniques are applied) from the whole document, Regions of Interest (or ROIs like blocks, grid cells, connected-components, words, etc.) or Writing Fragments (WFs). Textural-based methods are more efficient runtime-wise, as they do not need additional preprocessing steps such as segmentation and binarization. Another advantage is parameter-free learning. On the other hand, textural-based methods usually need more data to extract reliable and highly discriminative features, which might not be the case (i.e. forensic experts to predict the authenticity have to deal with small queried handwriting samples). Additional drawback of such methods is a need for high contrast images.

Grapheme-based (bags of features (BOF)-based) methods rely on codebook generation for all character types. In such methods, handwritten text segmented into character segments and graphemes are generated for each character, after that a codebook is generated using any clustering algorithm, for each codebook histogram distance scores are calculated and the resulting histogram used for classification. Such methods are considered as effective for the writer's identification because they do not require an excessive amount of the writer's original handwriting sample to predict the authenticity of the queried handwriting sample. However, this type of methods have some drawbacks such as the need to spend much time extracting and comparing grapheme details. Additionally, such methods require a substantial amount of memory due to the large number of grapheme features, especially for methods that apply a single clustering algorithm.

Auto learned(model-based) methods rely on the usage of deep learning techniques to extract automatic features learned by deep models. Such methods are robust and report high identification rates, especially for word-level identification. Another advantage is that models automatically learn representations from data (i.e. automatic feature extraction) that's why Convolutional Neural Networks have been widely used to extract features from the images (i.e., words, pages, patches). The main drawback of model-based methods is training because a large number of samples is needed, the training could be very time consuming and models tend to overfit. Another important drawback of model-based methods is the difficulty of selecting the best values for a large number of parameters. Additionally, this type of methods require very high computational resources and time.

Writer identification. During handwriting analysis to perform a writer identification the method of individual personal identification is used, the essence of which is the full exploration of the individual characteristics of the document being examined and the comparison of these with the known manifestations of these characteristics in other documents (comparative material) [1].

The writer identification methodologies share common/standard steps of preprocessing, feature extraction and writer identification (frequently referred as classification) as shown in Figure 1.

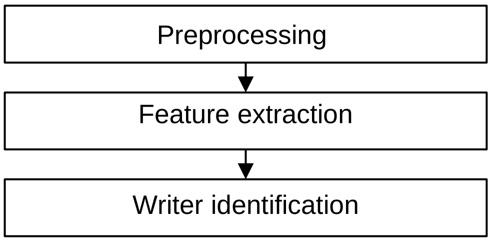

Fig. 1. Writer identification methodologies common steps

Studies in the field of writer identification also fall into four handwriting analysis methodologies categories.

In textural-based category Christlein et al. [4] proposed to use local descriptor RootSIFT to capture local properties of handwritten documents, use GMM supervectors as encoding method and use Exemplar-support vector machines (SVM) for similarity measure.

In auto learned category Christlein et al. [5] proposed use of Convolutional Neural Network (CNN) to train a powerful patch representation using cluster memberships as targets and a global image descriptor is created by means of Vector of Locally Aggregated Descriptors (VLAD) encoding. Chen et al. [6] proposed deep CNN (ResNet-50) and the weighted label smoothing regularization (WLSR) for data augmentation to allow more discriminative features to be learned to represent the properties of different writing styles. Helal et al. [8] experimented CNN DL with an additional dissimilarity approach with an SVM classifier. Sulaiman et al. [9] proposed a local binary pattern (LBP) as a hand-crafted feature descriptor and AlexNet structure for extracting deep description as a second descriptor, both descriptors then assembled into a data matrix and encoded using VLAD. A SEGmentation-free model presented by Kumar and Sharma [10] where they apply the CNN to characterize the writer of a given sample. Koepf et al. [12] presented a novel method based on vision transformer (ViT) with K-Nearest Neighbors (KNN) classifier. Semma et al. [13] presented a deep CNN trained to capture deep features in small patches which extracted from handwritten images using the FAST key points and Harris Corner (HC) detector and the deep learned features are encoded using VLAD and the Triangulation Embedding with a KNN classifier. At first He and Schomaker [7] proposed deep adaptive CNN based on multi-task adaptation of the AlexNet structure, after that He and Schomaker [11] proposed a deep neural network (FragNet, inspired by Fraglets and BagNet) based on fragments segmented on the input document image and feature pyramid maps in the CNN, in more recent work He and Schomaker [14] proposed global-context residual recurrent neural network (GR-RNN) where spatial relationship between the sequence of fragments is modeled by the recurrent neural network (RNN) to strengthen the discriminative ability of the local fragment features and the complementary information between the global-context and local fragments is leveraged. Wirmanto et al. [15] proposed to segment data into two inputs and use Xception as feature extractor in the Siamese network, the last one generates encoded representation and extracts identification label.

State-of-the-art writer identification methodologies usually measure performance based on publicly available datasets (i.e CVL, IAM ect.) and metrics used for comparison are Top-1 Accuracy and Top-5 Accuracy. Where Top-1 accuracy means expected result is exactly the expected answer (i.e. the one with highest probability) and Top-5 accuracy means expected result is among first five highest probability answers. State-of-the-art writer identification methods with high recognition rates on public CVL dataset with Top-1 accuracy are shown in Table 1.

Table 1 – **State-of-the-art writer identification methods accuracy with CVL dataset**

Author	**Method**	**Acc (%)**
Christlein et al. [4]	GMM super vector + E-SVM	99.4
Christlein et al. [5]	CNN + VLAD	99.5
Chen et al. [6]	ResNet-50	99.2
He and Schomaker [7]	Adaptive CNN	79.1
Helal et al. [8]	CNN	99.80
Sulaiman et al. [9]	CNN + LBP	97.55
Kumar and Sharma [10]	SEG WI model	99.35
He and Schomaker [11]	FragNet-64	99.1
Koepf et al. [12]	ViT	99.0
Semma et al. [13]	Resnet-34	99.5
He and Schomaker [14]	GR-RNN	99.3
Wirmanto et al. [15]	Xception+siamese	99.88

Personality detection. БDuring handwriting analysis to perform personality detection one investigates the differences of the written samples from standard writing and on the basis of this the personality of the person providing the writing sample is deduced [1].

Similar to writer identification, personality detection methodologies share the common/standard steps of preprocessing, feature extraction and personality detection (frequently referred as classification) as shown in Fig. 2.

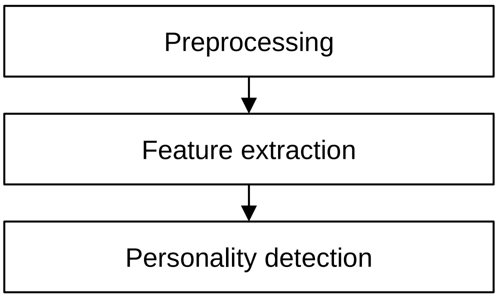

Fig. 2. Personality detection methodologies common steps

To perform personality detection one need to pick a personality psychology measurement model. There are two popular models for assessing personality types: Big Five and Myers Briggs Type Indicator (MBTI). The Big Five model formalizes personality types as scores while MBTI as classes. Article [32] performed a comparison and concluded that even though algorithms trained on MBTI could have better performance, the Big Five is much more informative and has great variability in performance depending on the algorithm.

Studies in the field of personality detection show a wide variety of approaches and they, for the most part, fall into the auto learned category.

So in auto learned category Gavrilescu [18] proposed 3-level neural network architecture with a ANN, SVM and KNN classification on 3-d level. Polap and Wozniak [19] proposed a flexible neural network where some neurons adapt to classification purposes. Topaloglu and Ekmekci [20] proposed to use decision tree and data mining techniques in conjunction with ID3 and J48 algorithms. Gavrilescu and Vizireanu [21] continued the 3-level architecture approach and this time base layer includes normalization and segmentation, intermediary layer generates Handwriting Map (HM) using binary code, top layer performs classification using feed-forward neural network(FFNN). Joshi et al. [22] developed a classification framework based on the SVM and used template-matching for letter extraction. Wijaya et al. [23] compared left, right, top and bottom margins segmentation and used SVM for classification. Fatimah et al. [24] proposed CNN based classification with SGD and AdaDelta optimization methods. Chitlangia and Malathi [25] proposed to perform feature extraction using Histogram of Oriented Gradient (HOG) technique and classification using SVM. Thomas et al. [26] collected texts written by ten subjects using a CNN model and logistic regression, and selected seven features for the classification task. Pathak et al. [27] proposed deep neural network architecture based on CNN with Long Short-Term Memory (LSTM) and Connectionist Temporal Classification (CTC). Elngar et al. [28] proposed a novel Personality Trait Level Detection Model (PTLDM) where they use Personality Analyzing Network (PAN) and PersonaNet for classification. Bernardo et al. [29] proposed as a feature extractor to use SqueezeNet lightweight architecture model optimized with hyper-parameter optimization method and SVM for classification. Rahman et al. [30] proposed extracting features using graph-based writing representation approach and performing classification using Semi-supervised Generative Adversarial Network (SGAN). Samsuryadi et al. [31] performed a complex comparison of SVM (three variations of the kernel: linear, RBF,and polynomial), KNN, and decision tree approaches.

Personality detection researchers, unlike writer identification researchers, usually instead of publicly available datasets measure performance based on private handwriting datasets and metrics used for the comparison are accuracy, precision, recall, F1 score, true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Personality detection literature methods with high recognition rates on datasets with accuracy metric are shown in Table 2.

Discussion. Handwriting analysis systems aim to automatically extract important information from handwriting. They include such disciplines as personality detection, writer identification and many others.

For further development of such systems, it is necessary to overcome many problems, for example, improper feature extraction, overfitting, underfitting, unreliable training data, picking model for assessing personality types etc. The current directions in the development of modern handwriting analysis systems are the study and use of various architectures of neural networks and deep learning. The current high accuracy rates also suggest the need for larger datasets. This would also widen the scope for techniques relying on more training data. Since writer identification and personality detection fields have so much in common it might be reasonable to use techniques from one field to improve the other one.

Table 2 – **Personality detection literature methods**.

Author	**Method**	**Acc (%)**
Gavrilescu [18]	ANN, SVM, and KNN	88.6
Polap and Wozniak [19]	Flexible neural network	93
Topaloglu and Ekmekci [20]	Decision tree	93.75
Gavrilescu and Vizireanu [21]	FFNN and template-matching	84.4
Joshi et al. [22]	SVM and template-matching	97
Wijaya et al. [23]	SVM	82.73
Fatimah et al. [24]	CNN	97.31
Chitlangia and Malathi [25]	HOG+SVM	80
Thomas et al. [26]	CNN	65
Pathak et al. [27]	Deep NN	97.7
Elngar et al. [28]	ANN + PersonaNet	65
Bernardo et al. [29]	Hybrid two-stage SqueezeNet and SVM	91.26
Rahman et al. [30]	SSL, SGAN	91.30
Samsuryadi et al. [31]	SVM, KNN, and decision tree	99

Conclusions

Thus, this article presents four groups of methods for handwriting analysis categorized by the type of features as a functional context. In addition the actual advantages and disadvantages with the use and development of such methods are presented. Both methods for identifying the personality of a person and writer identification methods which are used in the state-of-the-art handwriting analysis systems are presented.

The studies in these fields and comparative analysis are presented. Also, discussion about performance measurements, databases and typical metrics are presented.

Based on analysis, conclusions about future development and possible directions of handwriting analysis systems were presented.

References

1. Hengl M. Comparison of the Branches of Handwriting Analysis, Часопис Національного університету "Острозька академія". Серія: Право. – 2014. – № 2(10)
2. Hemlata S. Personality detection using handwriting analysis: Review / Hemlata S., Singh S. // In Seventh International Conference on Advances in Computing, Electronics and Communication. – 2018. – 18-19 August – P. 85-89.
3. Alaei F. Review of age and gender detection methods based on handwriting analysis / Alaei F., Alaei A. // Neural Computing and Applications. – 2023. – September – P. 23909-23925.
4. Christlein V. Writer Identification Using GMM Supervectors and Exemplar-SVMs / Christlein V., Bernecker D., Hönig F., Maier A., Angelopoulou E. // Pattern Recognition. – 2017. – Vol. 63 – March – P. 258-267.
5. Christlein V. Unsupervised Feature Learning for Writer Identification and Writer Retrieval / Christlein V., Gropp M., Fiel S., Maier A., // In 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). – 2017. – Vol. 13 – 9-15 November.
6. Chen S. Semi-supervised Feature Learning For Improving Writer Identification / Chen S., Wang Y., Lin C., Ding W., Cao Z. // Information Sciences. – 2019. – Vol. 482 – May – P. 156-170.
7. He S. Deep adaptive learning for writer identification based on single handwritten word images / He S., Schomaker L. // Pattern Recognition. – 2019. – Vol. 88 – April – P. 64-74.
8. Helal L. G. Representation Learning and Dissimilarity for Writer Identification / Helal L. G., [et al.]. // 2019 International Conference on Systems, Signals and Image Processing (IWSSIP). – 2019. – June – P. 63-68.

9. Sulaiman A. Length Independent Writer Identification Based on the Fusion of Deep and Hand-Crafted Descriptors / Sulaiman A., Omar K., Nasrudin M. F., Arram A. // IEEE Access. – 2019. – Vol. 7 – June – P. 91772–91784.
10. Kumar P. Segmentation-free writer identification based on convolutional neural network / Kumar P., Sharma A. // Computers & Electrical Engineering. – 2012. – Vol. 85 – June.
11. He S. FragNet: Writer Identification Using Deep Fragment Networks / He S., Schomaker L. // IEEE Transactions on Information Forensics and Security. – 2020. – Vol. 15 – March – P. 3013–3022.
12. Koepf M. Writer Identification and Writer Retrieval Using Vision Transformer for Forensic Documents / Koepf M., Kleber F., Sablatnig R., // Document Analysis Systems. – 2022. – May – P. 352–366.
13. Semma A. Writer Identification using Deep Learning with FAST Keypoints and Harris corner detector / Semma A., Hannad Y., Siddiqi I., Djeddi C., El Youssfi El Kettani M. // Expert Systems with Applications. – 2021. – Vol. 184 – Dec– P. 115473.
14. He S. GR-RNN: Global-Context Residual Recurrent Neural Networks for Writer Identification / He S., Schomaker L. // Pattern Recognition. – 2021. – Vol. 117 – Apr.
15. Wirmanto S. Offline Handwriting Writer Identification using Depth-wise Separable Convolution with Siamese Network / Wirmanto S., Agustini D.A.R., Atmanto D.A., // International Journal On Informatics Visualization. – 2024. P. 535–541.
16. Purohit N. State-of-the-Art: Offline Writer Identification Methodologies / Purohit N., Panwar S., // International Conference on Computer Communication and Informatics (ICCCI). – 2021. – Jan – P. 1–8.
17. Ahmed B. Q. Offline text-independent writer identification using a codebook with structural features / Ahmed B. Q., Hassan Y. F., Elsayed A. S., // PLOS ONE 18. – 2023. – Vol. 18(4) – April – P. 1–31.
18. Gavrilescu M. Study on determining the Myers-Briggs personality type based on individual's handwriting / Gavrilescu M., // In Proceedings of the 5th IEEE International Conference on E-Health and Bioengineering. – 2015. – Nov – P. 1–6.
19. Połap D. Flexible neural network architecture for handwritten signatures recognition / Połap D., Wozniak M., // International Journal of Electronics and Telecommunications. – 2016. – Vol. 62(2) – April – P. 197–202.
20. Topaloglu M. Gender detection and identifying one's handwriting with handwriting analysis / Topaloglu M., Ekmekci S., // ExpertSystems with Applications. – 2017. – Vol. 79 – March – P. 236–243.
21. Gavrilescu M. Predicting the big five personality traits from handwriting / Gavrilescu M., Vizireanu N., // EURASIP Journal on Image and Video Processing – 2018. – Vol. 2018(1) – July.
22. Joshi P. A machine learning approach to employability evaluation using handwriting analysis / Joshi P., Ghaskadbi P., Tendulkar S., // In Proceedings of the Communications in Computer and Information Science ICAICR 2018. P. 253–263.
23. Wijaya W. Personality analysis through handwriting detection using android based mobile device / Wijaya W., Tolle H., Utaminingrum F., // International Journal of Information Technology and Computer Science – 2018. – Vol. 2(2). P. 114–128.
24. Fatimah S. H. Personality features identification from handwriting using convolutional neural networks / Fatimah S. H., Djamal E. C., Ilyas R., Renaldi F. // In Proceedings of the 4th International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE – 2019. – Nov. – P. 119–124.
25. Chitlangia A. Handwriting analysis based on histogram of oriented gradient for predicting personality traits using SVM / Chitlangia A., Malathi G., // Procedia Computer Science – 2019. – Vol. 165 – Jan. – P. 384–390.
26. Thomas S. A framework for analyzing financial behavior using machine learning classification of personality through handwriting analysis / Thomas S., Goel M., Agrawal D., // Journal of Behavioral and Experimental Finance, 2020. Vol. 26(2).
27. Pathak A. R. Personality analysis through handwriting recognition / Pathak A. R., Raut A., Pawar S., Nangare M., Abbott H. S., Chandak P., // Journal of Discrete Mathematical Sciences and Cryptography – 2020. – Vol. 23(1) – Jan. – P. 19–33.
28. Elngar A. A. A deep learning based analysis of the big five personality traits from handwriting samples using image processing / Elngar A. A., [et al.], // Journal of Information Technology Management – 2021. – Vol. 12 – P. 3–35.
29. Bernardo L. S. A hybrid two-stage SqueezeNet and support vector machine system for Parkinson's disease detection based on handwritten spiral patterns / Bernardo L. S., Damasevicius R., De Albuquerque V. H. C., Maskeliunas R., // International Journal of Applied Mathematics and Computer Science – 2021. – Vol. 31(4) – Dec. – P. 549–561.
30. Rahman A. U. Predicting the big five personality traits from hand-written text features through semi-supervised learning / Rahman A. U., Halim Z., // Multimedia Tools and Applications – 2022. – Vol. 81(23) – Sep. – P. 1–17.
31. Samsuryadi S.. A Framework for Determining the Big Five Personality Traits Using Machine Learning Classification through Graphology / Samsuryadi S., [et al.], // Journal of Electrical and Computer Engineering. 2023. Vol. 2023(1). Jan. P. 1–15.
32. Lepri F. Is big ve better than MBTI? / Lepri F., Lepri. B., // Proceedings of the Fifth Italian Conference on Computational Linguistics CLiC-it – 2018. – Jan. – P. 93–98.

Received (Надійшла) 15.12.2024
Accepted for publication (Прийнята до друку) 29.01.2025

Аналіз методів визначення особистості та ідентифікації письменника

М. В. Шупилюк, В. О. Мартовицький

Анотація. Рукописний текст як мультисенсорна діяльність може відображати особистість і водночас може слугувати біометричним ідентифікатором. Аналіз почерку використовується в різних сферах, включаючи історію, криміналістику, освіту, безпеку, кадрові питання тощо. У цій статті методології аналізу почерку було розглянуто та розділено на чотири групи, виділивши переваги та недоліки кожної групи. Також у цій статті описуються різні проблеми, пов'язані з розробкою систем аналізу рукописного тексту, як-от неправильне виділення ознак, перенавчання, недонавчання, ненадійні навчальні дані, вибір моделі для оцінки типів особистості тощо. Представлено як методи надійної ідентифікації письменника поза мережею, так і методи прогнозування людської особистості, які використовуються в найсучасніших системах аналізу почерку. Крім того, були проаналізовані поточні дослідження та загальні підходи до вимірювання продуктивності та вибору бази даних як для ідентифікації письменника, так і для визначення особистості. Також представлено перспективні напрями розвитку сучасних систем аналізу рукописного тексту.

Ключові слова: почерк, аналіз почерку, особистість, риси почерку, графологія, визначення особистості, ідентифікація письменника.