Predictive Analysis of Demographic Information and Handedness from Handwriting using CNNs and Transformers

Mia Liang
MIT
77 Massachusetts Ave
mliang51@mit.edu

Aleksander Jovanovic-Hacon
MIT
77 Massachusetts Ave
ajovanov@mit.edu

Abstract

Handwriting analysis holds potential for predictive analysis in demographic categorization and health assessment, including gender, age, handedness, and disease onset. This study aims to investigate the predictive capabilities of deep learning techniques, specifically convolutional neural networks (CNNs) and transformers, in discerning demographic attributes such as gender and native language from handwriting patterns. Building upon prior research, we introduce the novel integration of self-attention mechanisms into CNN architectures handwriting analysis tasks using the publicly available IAM On-line Dataset [10]. Evaluation metrics including accuracy, precision, and recall will be utilized to assess the performance of our models in predicting supplementary information about the writer. By addressing the intrinsic challenges of feature extraction in handwriting analysis and exploring novel combinations of predictive variables, this research contributes to both the fields of computer vision and psychological diagnostics. The outcomes of this study bear implications for forensic analysis, and understanding handwriting evolution.

1. Introduction

Handwriting analysis, often referred to as graphology, involves the examination of an individual's handwriting to gain insights into various characteristics, such as demographics and health conditions. This multidisciplinary practice intersects fields such as forensics, healthcare, psychology, and human-computer interaction. For instance, forensic experts use handwriting analysis to verify signatures, identify anonymous correspondents, and profile suspects based on handwritten samples.

Despite its broad application spectrum, the most common challenge in handwriting analysis remains its variability driven by personal and cultural factors. Previous attempts have faced challenges in achieving high accuracy, particularly in diverse population groups. Specifically, gender and handedness have been traditional focus areas within handwriting analysis due to their relatively binary nature, yet even these have met with mixed successes.

Our study aims to address these limitations by expanding the scope of predictive variables and improving upon existing methodologies. Building upon the work done by Morera et al. (2018), who primarily focused on gender and handedness using CNNs, our study integrates advanced deep learning techniques such as self-attention-based CNNs. We consider a more diverse set of demographic and linguistic variables, such as native language, which may reveal a "written accent" in English script handwriting samples. By utilizing the robust IAM dataset and employing novel neural network architectures, we aim to achieve higher accuracies in demographic classification and contribute meaningfully to applications in forensic sciences, medical diagnostics, and perhaps even the evolution of handwriting.

1.1. Related Work

Previous research in this field has explored various aspects of graphology. While the literature on this topic is limited, studies have begun to investigate automated approaches to address these challenges.

Advancements in artificial intelligence as recent as this year, have already seen huge steps in the direction of automating handwriting feature comparison. Although still a work in progress, the AI model TextOracle has shown promising results, achieving an 85% accuracy rate for recognition of neat, regular handwriting and a 60% success rate for cursive handwriting [2]. We also see other automated processes like the Handwriting Analysis Tool (HAT), which enables the concurrent analysis of multiple handwriting styles, and the Visual-Pattern Detector (VPD), which automates the recognition of visual patterns like words, drawings, and seals. Other similar pattern analysis software include the Line Detection Tool (LDT), XRF-Data Analysis Tool (XRF-DAT) and the Artefact-Feature Analysis Tool (AFAT) [3]. Zhao et al. proposes a Spatial Variation-dependent Verification (SVV) scheme using textural features (TF) to authenticate digital or handwritten signatures with experimental results of accuracy 95.587% [7]. Intrapersonal parameter optimization methods have also been developed for offline handwritten signature augmentation, with 95.6% performance improvement results [9]. This type of automated handwriting comparison makes a crucial difference in the field of forensics for example, in identifying suspects, verifying signatures, and providing evidence in legal investigations in a prompt manner

From the comparison of handwriting, emerges the predictive analysis portion. What can we infer about the writer from the way they write? How might the slant of their letters hint at underlying personality traits? What about disease detection in the tremor of one's hand? Many studies have delved into graphology. One notable thesis, through techniques mapping visual features to personality traits, demonstrated a 90% accuracy in mapping handwriting to user traits across 100 participants [4]. Another study focused on algorithmic accuracy in training and evaluating handwriting samples from the IAM database, demonstrating the effectiveness of artificial neural networks (ANN) with an 86% accuracy rate [5]. A different machine learning-based methodology predicts personality traits from handwriting features, leveraging characteristics like baseline, margin, word slant, and t-bar height [6]. Ma et al. introduced a transformer deep learning model for sequence learning, analyzing the tremor symptoms of signatures, demonstrating a 97.8% improvement in efficiency and achieving an accuracy ratio of 95.4% in the validation and verification process [8].

One of the cornerstone works in predictive analysis is the aforementioned study by Morera et al. (2018) published in Hindawi's Complexity journal. Their research centers on leveraging convolutional neural networks (CNNs) for classifying gender and handedness from handwriting samples. They utilized publicly available datasets like IAM, which includes English texts, and KHATT, for Arabic texts, to validate their models.

The Hindawi paper posited CNNs as superior to other feature-extraction methods in handling the complexity of handwriting variability. They propose a unique CNN architecture consisting of multiple convolutional and pooling layers followed by fully-connected dense layers, enabling the model to extract and classify features hierarchically. This architecture achieved notable accuracy in binary classification tasks, such as gender (80.72% overall accuracy) and handedness (70.91% overall accuracy).

Moreover, Morera et al. introduced the combined gender-and-handedness classification problem, an aspect not previously covered in the literature. This combination, although complex, aimed to unveil any underlying correlations between gender and handedness features in handwriting. Their method involved treating the problem as a multiclass classification task, which reportedly achieved better results (70.84% overall accuracy in 4-class gender/handedness problem) than solving the two binary problems independently [1].

1.2. Our Approach

Building on their achievements, our research proposes to further expand the demographic and linguistic scope of handwriting analysis. One novel aspect is the exploration of native language influence on English handwriting patterns—a hypothesis that such influence could manifest as a "written accent." We suggest that understanding these nuances could dramatically sharpen the precision of demographic predictions. It would also expand on the complexity of the multi-class classification problem performed by Morera et al.

To implement these improvements, we will incorporate advanced architectures such as self-attention mechanisms within CNNs. Transformers, known for their capability to handle long-range dependencies in sequences, can be efficiently integrated into CNNs to improve feature extraction from handwriting samples. Our hybrid approach seeks to balance the specificity of CNNs in local pattern recognition with the ability of self-attention mechanisms to capture global interdependencies.

2. Methods and Models

2.1. Dataset and Preprocessing

The preprocessing of the dataset from the IAM On-Line Handwriting database, which contains over 1700 handwriting samples from 221 individual writers, involved several crucial steps to ensure the data's suitability for subsequent analysis and modeling. This database includes 1,539 pages of scanned text, capturing a broad spectrum of linguistic expressions and textual genres, as well as 13,353 labeled text lines, amounting to 86,272 word instances from an 11,059-word dictionary. We utilized a predefined subset of these samples, divided into training, testing, and two validation datasets.

Our first step was to access the labeled lines of text, as these were the most straightforward images available for download. We utilized four text files containing the sets of sample IDs for the training, validation, and test sets. Each sample ID (e.g., "a01-020x") corresponds to specific lines of text images stored in a structured directory (e.g., lineImages/a01/a01-020/a01-020x-01.tif to a01-020x-07.tif). To compile the corresponding sets of images, we traversed the lineImages directory based on these IDs. Each sample is linked to an author through the forms.txt file, which maps sample IDs to author IDs. Additionally, using the writers.xml file, keyed by author ID, we extracted the labels for each image. This organization process allowed us to generate a comprehensive set of labeled images ready for further processing.

The dataset was split as follows:

- **Training Set**: 776 samples containing 5,365 lines from 119 authors.
- **Test Set**: 217 samples containing 1,526 lines from 37 authors.
- **Validation Set 1**: 193 samples containing 1,301 lines from 31 authors.
- **Validation Set 2**: 545 samples containing 3,926 lines from 68 authors.
- **Total**: 1,731 samples containing 12,118 lines from 221 authors.

To ensure uniformity in our dataset, we standardized the image sizes to 300 x 1000 pixels. This process involved resizing and padding the images using the OpenCV library. The resizing process maintained the aspect ratio, and padding was added to fit the target size, thereby mitigating variations in aspect ratios and enhancing model performance.

2.2. Feature Extraction

Each handwriting sample was associated by metadata that included the writer's ID, gender, native country, native language, handedness, and birthdate. For model compatibility, it was necessary to convert these categorical labels into a numerical format. We applied label encoding, a technique where each category value is assigned a unique integer. In our classification tasks for the 17 native languages of the set of writers, we used a multi-class classification model with cross-entropy loss and Adam optimization. For the gender classification task, we used a binary classification model with binary cross-entropy loss and Adam optimization.

In order to fine tune and optimize our feature extraction, we used the grid search technique, which involves systematically evaluating the model's performance across various combinations of hyperparameters. In our case, we varied the number of filters applied in each of the convolutional layers (denoted as 'a' and 'b'). By exhaustively searching through the specified grid of parameter values and employing cross-validation to assess performance, GridSearchCV identified the parameter values that yielded the highest validation accuracy. This approach allowed us to fine-tune the model's hyperparameters systematically, maximizing its performance without resorting to manual trial and error. Additionally, we utilized grid search to fine-tune other necessary parameters, such as the number of epochs, batch size, and learning rate, in our models.

2.3. Convolutional Neural Networks

Our CNN architecture, as per prior research, consists of six trainable layers organized into two stacks of convolutional and subsampling (or max-pooling) layers, followed by two dense fully-connected layers. Dropout layers with probability p values of 0.25 and 0.5 were added after each max-pooling layer. These dropout layers serve to mitigate overfitting by randomly dropping a fraction of the neurons during training, thereby promoting better generalization to unseen data. The network processes input images with a resolution of 300 x 1000. Experimentation led us to utilize kernels of size 5x5 for the convolutional layers and 2x2 for the subsampling layers, optimizing performance. Parameters such as the number of feature maps for each convolutional layer and the number of output neurons for the final layer were determined through previously mentioned grid search techniques.

2.4. Transformers and Self-Attention CNNs

In handwriting analysis, capturing the sequential and spatial relationships between strokes and characters is critical as they can provide more comprehensive feature insights. Traditional CNNs excel at extracting local features but may struggle with capturing these long-range dependencies.

Transformers address the challenge of capturing long-range dependencies through the use of self-attention mechanisms. This allows each element of the input sequence to attend to all other elements, enabling the model to focus on relevant parts of the sequence regardless of their distance from one another. While transformers offer advantages, they also come with substantial computational costs, especially when applied to image data. Transformers require the input to be tokenized into patches, which can be computationally expensive and memory-intensive. For our task, which involves processing high-resolution handwriting images, this approach would be impractical.

However, the self-attention mechanism, a key component of transformers, can still be beneficial. By integrating self-attention into CNNs, we can capture the long-range dependencies without the full computational overhead of a transformer. This hybrid approach leverages the strengths of both CNNs and self-attention, making it well-suited for our handwriting analysis task.

Similar to our CNN architecture described in section 2.3, our CNN with self-attention contains two stacks of convolutional and max-pooling layers. After the second max-pooling layer, the feature maps are reshaped and passed through a multi-head self-attention block, enabling the model to focus on multiple aspects of the input simultaneously. Following this, the output is passed through fully connected layers, including a dropout layer to prevent overfitting, similar to the base CNN.

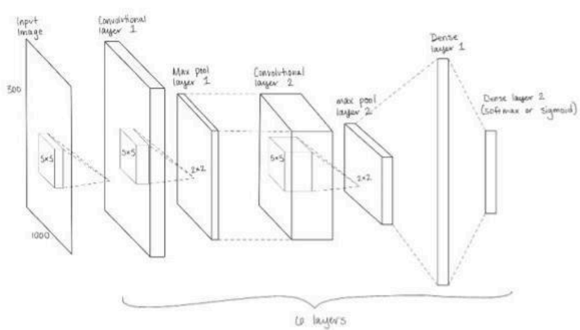

Figure 1. Architecture of the CNN, including 2 stacks of convolutional and sampling layers followed by dense fully-connected layers

2.5. Results

The Performance of all classifiers was evaluated using Validation Set 2 (VS2), which encompasses contributions from 68 authors, totalling 545 writing samples and 3926 images of lines of writing. All models were trained on the Training Set, which consists of 5365 images of lines of writing.

2.5.1 Gender

Within VS2, there were 48 male and 20 female authors contributing 2818 and 1108 images, respectively. Through fine-tuning, we identified the optimal hyperparameters for the Base CNN (a=32, b=64) and the Self-Attention CNN (a=32, b=64, utilizing 4 self-attention heads).

Base Convolutional Neural Network

Male (n=2818)	Precision	92.9
	Recall	83.4
Female (n=1108)	Precision	84.3
	Recall	77.6
	Overall Accuracy	81.8

With Self Attention

Male (n=2818)	Precision	95.2
	Recall	75.5
Female (n=1108)	Precision	82.8
	Recall	69.6
	Overall Accuracy	82.5

Figure 2. Results for gender classification using both architectures

2.5.2 Native Language

VS2 encompasses native speakers representing 9 of the 17 languages from the training dataset. Following fine-tuning, we determined the optimal hyperparameters for the Base CNN to be a=64 and b=128, while for the Self-Attention CNN, a=32, b=64, with 8 self-attention heads.

Base Convolutional Neural Network			With Self Attention		
Swiss German (n=2917)	Precision	94.9	Swiss German (n=2917)	Precision	95.2
	Recall	70.3		Recall	75.5
German (n=345)	Precision	81.5	German (n=345)	Precision	82.8
	Recall	63.8		Recall	69.6
French (n=229)	Precision	77.8	French (n=229)	Precision	78.9
	Recall	61.1		Recall	65.5
Spanish (n=91)	Precision	68.8	Spanish (n=91)	Precision	70.6
	Recall	60.4		Recall	65.9
Chinese (n=88)	Precision	57.1	Chinese (n=88)	Precision	50.5
	Recall	45.5		Recall	41.1
Persian (n=84)	Precision	66.7	Persian (n=84)	Precision	73.3
	Recall	59.5		Recall	65.8
Slovakian (n=78)	Precision	71.4	Slovakian (n=78)	Precision	73.3
	Recall	64.1		Recall	70.5
Hungarian (n=50)	Precision	57.1	Hungarian (n=50)	Precision	55.1
	Recall	40.0		Recall	44.0
Serbian (n=44)	Precision	77.8	Serbian (n=44)	Precision	77.8
	Recall	79.5		Recall	79.5
	Overall Accuracy	68.1		Overall Accuracy	72.3

Figure 3. Results for native language classification using both architectures

3. Conclusion

This study demonstrates the potential of integrating self-attention mechanisms within convolutional neural networks (CNNs) for the task of handwriting analysis, specifically in predicting demographic attributes such as gender and native language. Our results indicate that self-attention improved the accuracy of native language classification significantly, suggesting that the model could better capture the complex, long-range dependencies inherent in handwriting patterns influenced by linguistic backgrounds. However, this improvement was not mirrored in the gender classification task, where the predictiveness for female samples decreased, indicating that the benefits of self-attention may be task-specific.

The accuracy enhancement in native language classification was particularly notable for more prevalent languages, which suggests that the model was better at learning features from languages with larger sample sizes. This phenomenon highlights the importance of balanced datasets for training effective predictive models. Despite the computational cost associated with self-attention mechanisms, our hybrid approach demonstrated the feasibility and advantages of incorporating these advanced techniques in handwriting analysis.

3.1. Future Work

Building on the promising results of this study, future research should explore the full potential of transformers for handwriting analysis. This would involve efficient tokenization of input images to manage the computational load and enhance performance, as transformers can significantly improve the capture of sequential and spatial relationships in handwriting. Additionally, access to improved hardware resources is crucial. Enhanced GPU capacity would allow for more extensive fine-tuning, deeper and more complex convolutional layers, and overall better model optimization, thus potentially increasing accuracy and robustness.

Moreover, future studies should leverage the additional demographic information available in the dataset, such as profession and education level, to expand classification tasks and uncover new predictive patterns. Testing models on more diverse datasets, including non-English handwriting samples, could also generalize findings and improve model applicability across different languages and cultures. By addressing these areas, we can further advance the field of handwriting analysis, contributing to forensic science, medical diagnostics, and beyond, ultimately pushing the boundaries of what predictive modeling can achieve in understanding handwriting patterns.

3.2. Individual Contributions

Liang: Conducted gender binary classification experiment. This includes developing both base and Self-Attention CNNs, training, fine-tuning, and testing.

Jovanovic-Hacon: Conducted native language multi-class classification experiment. This includes developing both base and Self-Attention CNNs, training, fine-tuning, and testing.

All: Contributed to development of project idea, acquisition of pertinent datasets, data processing infrastructure, and writing the final paper.

References

[1] Ángel Morera, Ángel Sánchez, José Francisco Vélez, Ana Belén Moreno, "Gender and Handedness Prediction from Offline Handwriting Using Convolutional Neural Networks", *Complexity*, vol. 2018, Article ID 3891624, 14 pages, 2018. https://doi.org/10.1155/2018/3891624

[2] "TextOracle: Using AI to Support Forensic Hand-writing Analysis." *HTX Corporate*, www.htx.gov.sg/news/featured-news-textoracle-using-ai-to-support-forensic-handwriting-analysis

[3] Mohammed, H., Helman-Wazny, A., Colini, C., Beyer, W., Bosch, S. (2022). "Pattern Analysis Software Tools (PAST) for Written Artefacts." *Uchida, S., Barney, E., Eglin, V. (eds) Document*

Analysis Systems. DAS 2022. Lecture Notes in Computer Science, vol 13237. Springer, Cham. https://doi.org/10.1007/978-3-031-06555-2_15

[4] Hosaguthi Vishwanath, Yashaswini, "Auto -mating Graphology Using Computer Vision." Thesis, Georgia State University, 2020. doi: https://doi.org/10.57709/18660387

[5] Patil, Vishal & Mathur, Harsh. (2021). "Personality Prediction and Handwriting Recognition Using Machine Learning." 10.1002/9781119792345.ch8.

[6] Joshi, Prachi, et al. "Handwriting Analysis for Detection of Personality Traits Using Machine Learning Approach." *International Journal of Computer Applications*, vol. 130, no. 15, Nov. 2015, pp. 40–45, https://doi.org/10.5120/ijca2015907189.

[7] Zhao, H., Li, H. "Handwriting identification and verification using artificial intelligence-assisted textural features." *Sci Rep 13*, 21739 (2023). https://doi.org/10.1038/s41598-023-48789-9

[8] Ma, Chenbin, et al. "A Feature Fusion Sequence Learning Approach for Quantitative Analysis of Tremor Symptoms Based on Digital Handwriting." *Expert Systems with Applications*, vol. 203, Oct. 2022, p. 117400, https://doi.org/10.1016/j.eswa.2022.117400.

[9] T. M. Maruyama, L. S. Oliveira, A. S. Britto and R. Sabourin, "Intrapersonal Parameter Optimization for Offline Handwritten Signature Augmentation," in *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1335-1350, 2021, doi: 10.1109/TIFS.2020.3033442.

[10] Ly, N.T., Nguyen, H.T., Nakagawa, M. (2021). 2D "Self-attention Convolutional Recurrent Network for Offline Handwritten Text Recognition." *Lladós, J., Lopresti, D., Uchida, S. (eds) Document Analysis and Recognition –* ICDAR 2021. ICDAR 2021. Lecture Notes in Computer Science(), vol 12821. Springer, Cham. https://doi.org/10.1007/978-3-030-86549-8_13