INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2024, VOL. 70, NO. 1, PP. 145-152
Manuscript received November 15, 2024; revised March, 2024. doi: 10.24425/ijet.2024.149524

Authorial approach to the detection of selected psychological traits based on handwritten texts

Marek Woda, and Grzegorz Oliwa

***Abstract*—The study sought to use computer techniques to detect selected psychological traits based on the nature of the writing and to evaluate the effectiveness of the resulting software. Digital image processing and deep neural networks were used. The work is complex and multidimensional in nature, and the authors wanted to demonstrate the feasibility of such a topic using image processing techniques and neural networks and machine learning. The main studies that allowed the attribution of psychological traits were based on two models known from the literature, KAMR and DA. The evaluation algorithms that were implemented allowed the evaluation of the subjects and the assignment of psychological traits to them. The DA model turned out to be more effective than the KAMR model.**

***Keywords*—prediction; handwritten texts; personality traits; machine learning; neural networks**

I. Introduction

THIS research delves into the realms of digital image processing, deep machine learning, and graphology—a discipline often regarded as a pseudoscience, employed for deciphering hand-written texts to discern specific emotional attributes of their authors.

Graphology, a field characterised by its pseudo-scientific nature, is primarily concerned with the psychological study of handwriting. Its overall aim is to identify possible correlations between the way text is written and the personality traits of an individual. The results of graphological analysis can vary widely due to a variety of factors. In particular, graphologists often use different criteria when examining handwriting, leading to inconsistent interpretations. These criteria include elements such as the slant of the characters, the intensity of the pressure applied, the way in which connections are made between individual letters, the spacing between letters, words and lines, and the distinctive style of writing.

In view of the limited time available to us, this research manuscript focuses exclusively on the last aspect mentioned above, namely the study of an individual's letter-writing techniques. In the context of template-based graphological analysis, it is customary to undertake a comprehensive assessment of the entire textual corpus, except in cases of exceptional rarity. Consequently, it would be considered inappropriate to undertake a meticulous examination of individual letters in isolation. The intermediate procedures outlined in the previous subsection are therefore indispensable, for without their inclusion the integrity of this undertaking would be compromised at its very core.

While the subject of graphology is a daily pursuit for numerous highly qualified experts in this specialised field, recent years have seen significant challenges in conducting rigorous and reliable research into its efficacy. The development of a universally applicable model in this field represents an extremely formidable undertaking. Nonetheless, there is a consensus among a significant portion of the graphology community, often encompassing rather extreme viewpoints. One such viewpoint claims that it is indeed possible to identify individuals suffering from neurodegenerative diseases through graphological analysis.

Analogous to the principles of electrocardiograms, the establishment of a pre-defined norm becomes an essential precursor in the quest to identify anomalies or deviations that may indicate underlying problems.

This research can be divided into two distinct parts. The first section provides a comprehensive insight into the technological aspects and algorithms used in the realisation of the final software, allowing a comprehensive graphological analysis based on the chosen models. This section also delves into the effectiveness of extracting the constituent results of the various phases of the project, which together make up its entirety.

Conversely, the latter part of this thesis is devoted to addressing pertinent issues, including the methodology used to interview subjects, the research methodology and the interpretation of findings, culminating in the formulation of conclusive insights. In addition, the strengths and weaknesses of the approach used will be outlined, while potential avenues for further development of the research in question will be suggested.

Crucially, it should be emphasised that this research is based on the meticulous research carried out by respected professionals in the fields of graphology and psychology. Consequently, although this research is based on sometimes ambiguous and controversial material, the reliability of the features derived from handwritten texts is, by definition, considered acceptable within the wider graphological community.

II. Related work

This section is devoted to an exploration of the three most important works, as seen through the lens of the research discussed. These works serve as the basis for the development of an algorithm designed to deduce the characteristics of

Marek Woda is with Department of Computer Engineering, Wroclaw University of Technology, Poland (e-mail: marek.woda@pwr.edu.pl).

Grzegorz Oliwa, (e-mail: grzegorz.oliwa@outlook.com).

individuals whose handwriting has been graphologically analysed. The most important of these works, referred to as [1], is notable for its depth and breadth. It covers various facets of psychology and graphology, while providing concrete illustrations that make the intricacies accessible even to beginners. This study is structured into thematic sections focusing on various aspects, including the attributes of written language and the principles of graphology. Owing to the extensive scope of this research, some aspects have not been comprehensively addressed. Nevertheless, their influence is profound.

In this text, the researchers striveto find elements that have impact on the intelligence, integrity, tendency to secrecy, introversion, extroversion and other characteristics of individuals. However, the author refrains from presenting these personality traits in a comparative manner, which makes it difficult to accurately determine the consequences of different writing styles.

The next referenced item [2] claims that graphology has remained largely unchanged over the past few decades. In addition, extracts from this book are often used, both directly and indirectly, as the basis for various theories. Consequently, it has been chosen as the basis for the development of a model for identifying specific personality traits in the subjects studied. Nevertheless, its style is very similar to that of [1], often resembling a haphazard compilation of theories and handwriting assumptions rather than a well-structured guide to conducting graphological analysis with precision and reliability.

These references are based on extensive research carried out by expert level psychologists and graphologists, a crucial factor in their selection. However, the subject matter described in these books has not yet been sufficiently researched to shed its pseudo-scientific label, leaving its effectiveness unproven and undefined. Consequently, we decided to develop two different models, called DA [1] and KAMR [2] after the initials of the authors. In order to create a common framework for both models, we designed a questionnaire, which is described in detail in Chapter 4 on data collection methodology. To simplify the technical aspect, we chose a limited set of common signs rather than individual writing characteristics and their associated interpretations. Despite the differences in the implications of certain facts in the source texts, these differences were effectively addressed by a survey of the respondents in our study.

The next literature position [5], complements the conclusions of the previous two works. Originally published in 1991, it remains a seminal reference among psychologists and graphologists worldwide. The authors' approach transcends specific languages and aims to identify universal phenomena within a selected language family, with a preference for English and Chinese. Surprisingly, this source proved invaluable as it included interviews with individuals from Central Europe and Australia, totalling 50 writing samples and completed questionnaires from 42 and 8 participants respectively. Although not used in the development of graphological analysis software, it provided valuable insights.

For those entering the field of graphology, [8] is an essential resource. It offers guidance on the interpretation of common writing phenomena, echoing the themes discussed earlier, but with a broader scope. In addition, [3] presents groundbreaking research using neural networks in graphology tasks.

III. SYSTEM CONCEPT

The first stage is to collect participants' responses to the questionnaire (tinyurl.com/2m8ead6n), along with handwriting samples. This data forms the basis of a personality analysis, which is described in detail in Chapter IV.

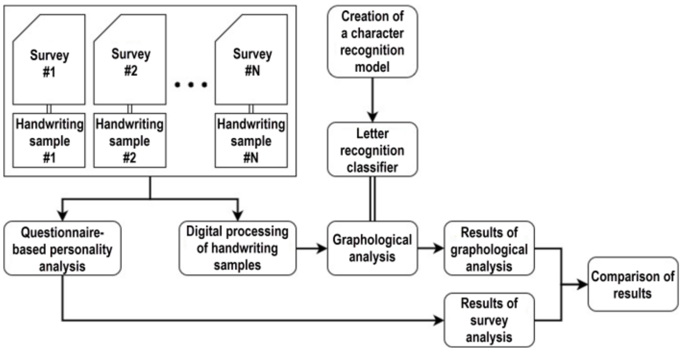

Fig. 1. Course of the research

Meanwhile, the handwriting samples are digitally processed (filtered, subject to thresholding and segmented), followed by further analysis. The data obtained from image analysis and questionnaire responses is juxtaposed to evaluate the efficacy of the developed system. Critical to the quality of the graphological analyser is the development of a character recognition model, discussed in section *A*, which focuses on choosing the best parameters for the neural network.

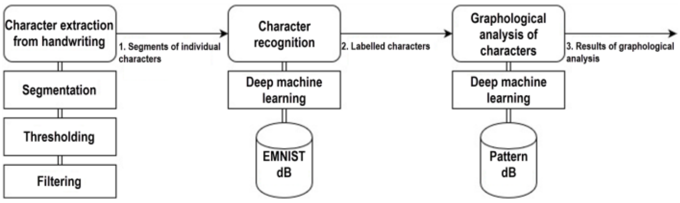

Fig. 2. Graphological analysis

The processes depicted in Figure 1 involve the incorporation of various intermediates, including individual character segments, labelled characters, and results derived from graphological analysis. These intermediates inherently introduce their own errors which, if not promptly mitigated, will be amplified. It is therefore imperative to understand the significance of each component.

IV. COURSE OF THE STUDY

The methodology adopted for this study mimicked a supervised machine learning framework. Initially, data collection focused on the psychological traits of different persons through two different approaches:

1. Using a survey with a fuzzy response scale (ranging from *definitely yes* to *definitely no*), with corresponding values as shown in Table III.
2. The use of a standardised written sample with consistent content for all participants.

TABLE I
ASSESSED CHARACTERS

Char	Features sought in DA model	Features sought in KAMR model
a	A constant sense of pressure, low self-confidence	A tendency to change one's mind, to be inspired by others, to improvise constantly
b	Ability to make new friends, responsibility, honesty and integrity	Scepticism, need for manipulation, desire to be in the centre of attention
g	Nervousness, tendency to worry, avoidance of responsibility	Acting only with the approval of others, generosity, openness, petulance
h	Tendency to lie to others, high self-control, preference for living alone	Taking quick and decisive action, low intelligence
k	Agreeableness, curiosity, self-control, egocentrism	Constant sense of pressure, blind stubbornness, gruffness
o	Deceitfulness, treachery, tendency to change one's mind frequently	High extrovertedness, willingness to talk to others, arrogance, taking criticism badly
t	Tendency to aggression, high level of self-confidence, eccentricity	Willingness to be aggressive, high levels of self-confidence, eccentricity
u	Constant rush, introverted lifestyle, frequent reliance on others	Lack of ability to make new friends, low self-confidence, nervousness
y	Openness to people and change, high level of intelligence, ruthlessness	Tendency to worry, critical evaluation of own actions

The survey consists of 50 pair-wise questions, following a widely used method referred to in [10, 11, 12, 14, 15], which facilitates the quantification of errors in respondents' answers. This is essential for the precise measurement of personality characteristics and reduces the tendency to under- or overestimate oneself. Question development is trait specific, aiming at optimal trait characterisation with subsequent categorisation. A fuzzy response scale is incorporated to accommodate the logarithmic function used in machine learning process.

TABLE II
SURVEY QUESTIONS VS. PERSONALITY CHARACTERISTICS

Char	Questions relevant to the analysis	
	DA	*KAMR*
a	8, 19, 36, 43	3, 4, 20, 30, 34
b	2, 4, 7, 11, 17, 27, 36, 39, 43, 46	7, 12, 23, 26, 31, 38, 45, 48
g	10, 18, 42, 49	11, 15, 22, 24, 26, 40, 43, 45, 46
h	12, 16, 26, 27, 31, 40, 41, 44	1, 9, 18, 20, 21, 28, 33, 48
k	4, 13, 22, 27, 31, 36, 41, 44, 49	3, 8, 14, 28, 42, 44
o	2, 5, 10, 12, 15, 16, 24, 25, 26, 31, 38, 40, 45	7, 25, 29, 35, 38, 39, 41, 44, 47, 50
t	3, 6, 14, 18, 33, 44, 48	3, 6, 14, 18, 33, 44, 48
u	10, 15, 19, 21, 27, 37	8, 10, 15, 18, 21, 22, 27, 32, 40
y	23, 34, 47, 50	8, 42, 43, 49

Each handwriting sample is first digitally processed to allow a comprehensive graphological analysis of its individual letters. The text, originally without Polish diacritics, was presented as *'Ochoczy kot skacze przez wysoki mur bo go lubi'* (in English: '*An eager cat jumps over a high wall because it likes to*'). This succinct but witty phrase remained the focus of the research throughout. During the sampling phase, a rigorous collection process was employed to gather a dataset comprising 50 pairs of questionnaires and handwritten fragments. The analysis of this remarkably short text proved sufficient for our graphological investigation, as it included all the letters relevant to our study. Moreover, certain letters recurred next to different characters, making it easier to extract them for subsequent comparative analysis.

The above traits can be identified by interpreting survey responses. However, their juxtaposition lacks balance, resulting in varying degrees of trait determination precision due to response scarcity. This variance contributes to algorithmic complexity and code size.

Some traits are inherently hard to explicitly describe. The overall occurrence coefficient of a trait is represented as the quotient of the cumulative score obtained from each response (Table III) divided by the number of relevant queries. For example, using the letter *'a'* in the DA model allows us to identify traits such as a constant feeling of pressure and reduced self-confidence. Assuming a respondent unambiguously affirms questions 8 and 19 while unambiguously denying questions 36 and 43, the overall probability of these attributes reaches a measure of 50%. However, this scenario is atypical as previous literature has shown that these attributes are strongly related. In most cases, a particular trait is manifested by a subset of the responses provided, which is incorporated into the algorithmic framework. Such a scoring scheme invariably yields values within the [0, 1] interval, facilitating the translation of these obtained values into a DNN score via the application of a (sigmoidal) function during the graphological analysis.

TABLE III
SURVEY RESPONSES EARN POINTS

Selected answer	Score
definitely yes	1.00
almost certainly yes	0.85
rather yes	0.65
rather not	0.35
almost certainly not	0.15
definitely not	0.00

A. Deep machine learning parameter selection

In the context of deep machine learning, critical parameters include the depth of the neural network (number of hidden layers), the configuration of each layer (number of nodes), the choice of activation function, and the duration of training, represented by the number of epochs.

A systematic approach was adopted in the search for optimal character classification from the EMNIST database. Extensive experimentation and a grid search method were used to determine the most effective network architecture. The optimal configuration consisted of five layers, with each successive layer having a decreasing number of nodes: 208, 156, 104, 52 and 26 nodes, respectively. The initial layer, corresponding to the input data, naturally had 784 nodes, reflecting the 28x28 pixel resolution of the EMNIST images. This resolution information was represented as a vector with a length equal to the product of the image width and height. The final layer contained 26 nodes, corresponding to the 26 letters of the English alphabet, representing different membership classes.

The network used a sigmoidal activation function and underwent 5000 training iterations (epochs). This approach resulted in an impressive classification effectiveness of 92%.

By switching the multilayer perceptron model to a convolutional counterpart using an interleave factor of 5, a significant increase in efficiency of up to 6% was achieved, raising it to a level of 98%. Consequently, this trained convolutional network was selected for subsequent use in handwritten letter identification within this thesis. An attempt to replace the logistic activation function with a unipolar threshold variant proved counterproductive, as it showed higher error rates in letter misclassification. The success rates of the deep neural networks (DNN) were subsequently recorded as 90% for the multilayer perceptron and 94% for the convolutional network.

The following phase of the study aimed to detect specific letter features. Initially, this was achieved using rudimentary techniques that relied on the OpenCV library's built-in algorithms (edge, circles, and other patterns detection). Over time, it became evident that employing a specialized neural network yielded expedited performance in this task while maintaining, if not surpassing, comparable efficiency levels. Through cross-validation, it was determined that the optimal approach was to aggregate all segments containing handwritten letters with specific patterns, using 80% of these segments for training and the remaining 20% for testing. This process was repeated iteratively across 10 different groups, increasing the reliability of the results. In cases where an insufficient number of patterns were available, manual supplementation was used. Remarkably, the efficiency of such a network proved to be remarkably high for the DA model, but considerably lower for the KAMR model discussed in the following section. Moreover, for the latter network, the logistic activation variant provided an additional benefit by allowing a stochastic interpretation of the results. The presence of a given characteristic could be determined from the network output both as a probability of occurrence and, consequently, allowed for increased inference and improved analytical capabilities. Table IV provides a succinct overview of the optimal parameters employed in the final neural network configuration.

TABLE IV
SUMMARY OF OPTIMAL *DNN* PARAMETERS

Rank	Activation function	Hidden layers	Max. effectiveness
1	Sigmoid	[208 – 156 – 104 – 52 - 26]	59%
2	Sigmoid	[200 – 150 – 100 – 50 - 26]	57%
3	Sigmoid	[150 – 100 – 75 - 50 - 26]	52%
4	Sigmoid	[250 - 200 - 150 - 100 - 50 - 26]	51%
5	Threshold unipolar	[208 – 156 – 104 – 52 - 26]	47%

The omission of a fixed epoch number of 5,000 from the table is due to empirical evidence supported by previous research and literature [4, 7, 9, 10]. It has been observed that as this epoch number increases, the network exhibits progressively better efficiency up to a certain threshold, beyond which further improvements become insignificant. The values in the last column refer exclusively to the DA model. Irrespective of the selected parameters, the KAMR model's performance remained below 24%. The investigation of the cause of this limitation was limited by the considerable computational requirements associated with model formation within *DNN* algorithm. Nevertheless, two plausible explanations are possible: firstly, potential errors in the survey methodology used with respondents, and secondly, the possibility that the algorithm itself was exceedingly effective but the underlying hypotheses upon which the KAMR model was constructed, were not correct.

Conversely, the DA model shows an exceptionally extraordinary level of efficiency. Under the optimal parameter configuration, an efficiency of 47% was achieved. However, careful investigation and analysis is essential to assess the accuracy of the personality traits identified. The following sections provide a comprehensive summary of these traits, with specific subsections for each model under review.

B. *DA model - results*

DA model was most effective in identifying personality characteristics that respondents rated as predominantly affirming, achieving a success rate of 69%. This score means that most questions relating to a specific characteristic were answered in accordance with the corresponding trait in reality. The following sections of this study are devoted to the examination of each model.

Conversely, the model had its lowest success rate in interpreting traits that fell into the category of responses categorised as *definitively yes*, with full agreement observed in only 12% of circumstances. Conversely, the DA model was highly effective when attempting to negate a specific feature, achieving an efficiency rate of 66%. Furthermore, the analysis of other results shows significant promise, with an efficiency of over 40%. A detailed breakdown of the percentage recognition of traits within the responses is given in Table VI.

TABLE V
DNN ALGORITHM RESULTS IN RELATION TO SURVEY RESPONSES (DA / KAMR)

Survey response	Recognition effectiveness	
	DA	KAMR
definitely yes	12%	5%
almost certainly yes	40%	21%
rather yes	69%	33%
rather not	51%	15%
almost certainly not	42%	47%
definitely not	66%	25%

TABLE VI
DA AND KAMR MODEL CHARACTER RECOGNITION

Char	Correct interpretation	
	DA	KAMR
a	12%	5%
b	40%	21%
g	69%	33%
h	51%	15%
k	42%	47%
o	66%	25%
t	66%	66%
u	52%	19%
y	41%	20%

The procedure was most effective in detecting features resulting from the interpretation of *o*, *t* and *u* characters. This observation is in line with expectations, as the distinctiveness of their notation facilitates relatively straightforward anomaly detection.

The previously presented results have been clarified using ranged values, referred to as standard thresholds (see Table VII). These thresholds indicate the presence of a numerical value in the interval [0, 1] at the output of the neural network, which is then translated into the corresponding threshold value. However, for the purposes of this study, we also investigated the performance of the algorithm when dealing with slightly underestimated and overestimated ranges.

TABLE VII
THRESHOLD VALUES TESTED

Description of characteristic	Default threshold	Underrated threshold	Overrated threshold
definitely yes	85 – 100%	80 – 100%	90 – 100%
almost certainly yes	70 – 84%	65 – 79%	75 – 89%
rather yes	50 – 69%	45 – 64%	55 – 74%
rather not	30 – 49%	25 – 44%	35 – 54%
almost certainly not	15 – 29%	10 – 24%	20 – 34%
definitely not	0 – 14%	0 – 9%	0 – 19%

These values were chosen empirically and produced interesting results for both the specific DA model and the KAMR, as detailed in the following subsection. An alternative approach to this analysis involves the use of extended and restricted interval methods, which would have a significant impact on an algorithm that was not originally designed with this in mind. The former approach would introduce ambiguity by associating a feature with multiple classes, while the latter would require the creation of an entirely new class, resulting in the feature not matching any of the options described in Table VII, and potentially leading to non-matching results.

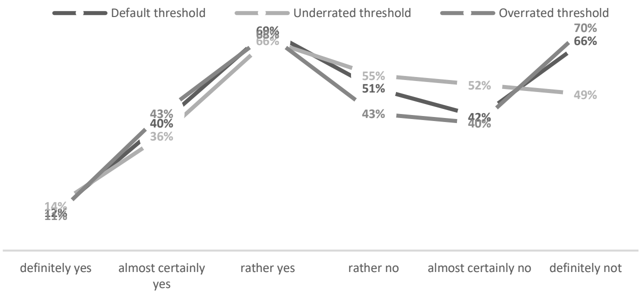

Fig. 3. Effect of change in feature classification threshold on respondent response coverage in the DA model

Lowering the threshold significantly reduces algorithm efficiency in classifying responses into the *definitely not* category. Consequently, it has a narrower operating range. Nevertheless, within this specific scenario, there is a noticeable decrease in effectiveness of up to 17% within the threshold range of 10 to 14%. Similar observations apply to the *definitely yes* category. In all other instances, precise quantification of effectiveness percentage change is unattainable, as solely the cumulative impact is observable.

When the responses of the characters were evaluated individually, changing the threshold showed different effects for different characters. Specifically, increasing the thresholds significantly improves the efficiency of interpretation for the character *k*, whereas the results for *g* and *t* characters are less favourable.

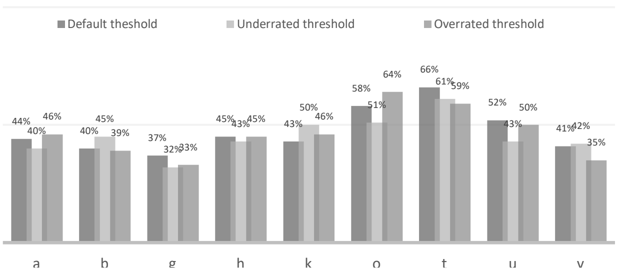

Fig. 4. DA Efficacy of in sign feature recognition across varied thresholds

This observation may be indicative of potential problems with either feature selection within the model or inherent imperfections within the algorithm.

C. KAMR model - results

Overall efficiency of this model was significantly lower and did not exceed 24%. It performed optimally when classifying responses as highly unlikely, while its worst performance was like that of the DA model, particularly when classifying responses as highly certain. In all other cases, the model's performance deteriorated significantly, making its use in broader predictions inadvisable. Nevertheless, it would be advisable to identify the distinguishing factors that contribute to its superior effectiveness in categorising responses as highly unlikely, and to integrate these findings into the DA model to create a novel variant (a hybrid one).

The model under analysis showed remarkably modest effectiveness (Table V) in practically all response categories. It could be up to three times less effective than the DA model, depending on the scenario. An assessment of the percentage recognition of the features found in the letters is detailed in Table VIII.

TABLE VIII
EFFICIENCY IN INTERPRETING RESPONSES ACROSS VARIOUS THRESHOLDS IN DA AND KAMR MODELS

Description of characteristic	Default threshold		Underrated threshold		Overrated threshold	
	DA	KAMR	DA	KAMR	DA	KAMR
definitely yes	12%	5%	14%	9%	11%	4%
almost certainly yes	40%	21%	36%	19%	43%	25%
rather yes	69%	33%	66%	29%	68%	34%
rather not	51%	15%	55%	17%	43%	10%
almost certainly not	42%	47%	52%	51%	40%	35%
definitely not	66%	25%	49%	14%	70%	32%
Average	47%	24%	45%	23%	46%	23%

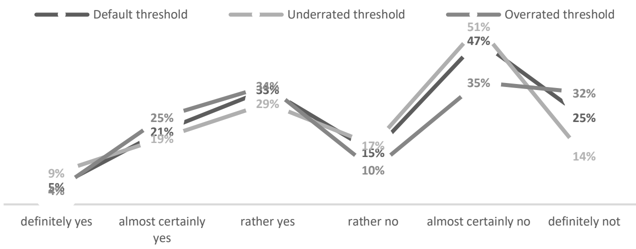

Fig. 5. Effect of change in feature classification threshold on respondent response coverage in the KAMR model

The only noteworthy aspect concerns the model's ability to identify features associated with the letter *t*. The performance of both models remains consistent, as they share a common underlying information base. This implies that the effectiveness of the KAMR model is below average, because of the inadequacies of the assumptions made by the authors of the paper. The recognition of other attributes is particularly poor. As with the previous model, we have chosen to assess the effectiveness of the KAMR model using the percentage thresholds shown in Table VIII. A summation of the results is presented in Table IX.

TABLE IX
EFFICIENCY OF CHARACTER RECOGNITION MODELS FOR DIFFERENT THRESHOLD VALUES

Character	Default threshold		Underrated threshold		Overrated threshold	
	DA	KAMR	DA	KAMR	DA	KAMR
a	44%	29%	40%	27%	46%	33%
b	40%	14%	45%	14%	39%	16%
g	37%	11%	32%	10%	33%	13%
h	45%	12%	43%	12%	45%	8%
k	43%	22%	50%	25%	46%	17%
o	58%	26%	51%	29%	64%	22%
t	66%	66%	61%	56%	59%	61%
u	52%	19%	43%	15%	50%	16%
y	41%	20%	42%	23%	35%	25%
Average	47%	24%	45%	23%	46%	23%

Like the previous model, the algorithm's behavior becomes evident when adjusting the threshold: it tends to provide a *definitely yes* response when the threshold is decreased and a *definitely not* response when it is increased. However, in contrast to the previously discussed scenario, a significantly greater departure from the default range is observed when the threshold is lowered. This is a positive advancement, as it implies that modifying the threshold for 'rather not' and 'almost certainly not' responses is not warranted. In other instances, the differences are too minor to draw substantial conclusions.

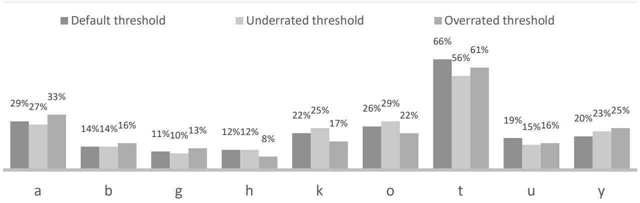

Fig. 6. KAMR Efficacy of in Sign feature recognition across varied thresholds

The limited efficacy of the model resulted in suboptimal outcomes at this juncture. Initially, a substantial qualitative advancement was anticipated, which would have negated the hypothesis of the model's inefficacy, attributing the subpar results to inadequate initial parameter selection. Regrettably, it is probable that the deductions made by the authors of the foundational paper underpinning the KAMR model were found to possess scant empirical support.

Sadly, a replication of this research at the particular subject level produced consistent results. The conclusions drawn confirm the limited validity of the KAMR model, which does not reliably identify high-quality personality trait inferences from handwriting.

D. Modernisation of the character segmentation algorithm

The initial version of the algorithm enabling character segmentation was unsatisfactory, as it worked on the basis of a simple contour detection algorithm provided with the OpenCV library. While it is true that it was able to find and identify almost all of the characters included in the analysed text, it also detected many others, such as nested loops and letter endings. In extreme cases, the number of characters detected using it was almost double (with 38 letters, 71 segments were created). With such low efficiency, it was decided to supplement it with the MSER (Maximally Stable Extremal Regions) algorithm. This resulted in a reduction of the maximum segmentation error from 87% to 47%, whereby it was determined from the formula *|[number of characters] - [number of segments]| / [number of characters]*. Among the texts examined, there was also one in which individual letters could not be determined due to the poor legibility of the handwriting. This was the one characterised by such a high error.

TABLE X
SUMMARY OF THE LARGEST SEGMENTATION ALGORITHM ERRORS

Sample #	*Contour detectionalg.error*	*MSER / contour detection alg. error*	*Improvement*
1	0.87	0.47	40%
2	0.74	0.37	37%
3	0.74	0.34	40%
4	0.71	0.34	37%
5	0.66	0.37	29%
6	0.66	0.32	34%
7	0.58	0.21	37%
8	0.55	0.11	44%
9	0.55	0.16	39%
10	0.55	0.21	34%
11	0.50	0.18	32%
12	0.45	0.21	24%
13	0.42	0.16	26%
14	0.42	0.24	18%
15	0.42	0.21	21%
16	0.42	0.26	16%
17	0.39	0.26	13%
18	0.37	0.21	16%
19	0.37	0.21	16%
20	0.34	0.18	16%

The values of the others oscillated around 24%. Unfortunately, this is not an exceptionally good value and certainly requires significant improvement, but due to the limited time and purpose of this thesis, it was decided to leave the algorithm in its current form and manually eliminate the erroneously indicated segments of the analysed digital image.

E. Char classification learning

The process of machine learning using a deep neural network to develop a model to recognise characters had to result in a very precise classifier. After all, it was up to it to ensure that specific letters were examined from the right angle. The use of a convolutional variant with an interleave of 5, made it feasible to obtain a classifier with an accuracy of 98%. The importance of this task was already known prior to the thesis, which is why the process was included in the thesis title. Figure 3, which is a compilation of the temperature map and the confusion matrix, shows the most frequently misclassified characters. As can be seen, most confusion occurred between letters such as n and m, or k and h. Everything else was indicated almost perfectly.

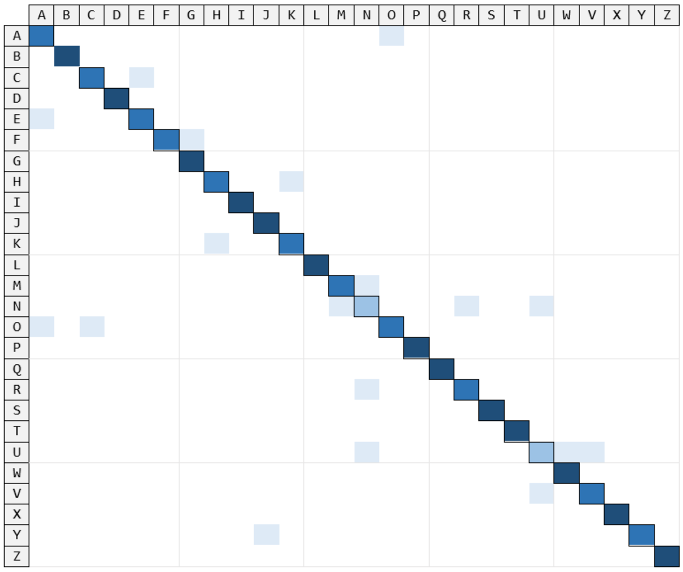

Fig. 7. EMNIST Temperature Confusion Matrix

The use of off-the-shelf networks, created along the lines of a multi-layer perceptron, only allowed correct classification in around 92% of cases. The network itself, on the other hand, used the ready-made EMNIST (Extended Modified National Institute of Standards and Technology) database for both learning and testing, consisting of, among other things, 145600 characters. They themselves were divided into 26 balanced classes of 5600 letters each. Cross-validation was applied to the learning process. All data - also in a balanced way - was divided into 10 sets of 14560 characters each. Then, within each set, a proportion was separated for learning (80%) and testing (20%). These proportions were determined based on the literature [2]. It was in this way, through labelled data, that the effectiveness of the classifier was established.

F. Model-based inference

By collecting the research samples in an appropriate manner, it was possible to label the data subjected to proper graphological analysis. It turned out that attempting to classify characters outside the EMNIST set, resulted in a success rate of 69%. This was not as satisfactory as expected, although still much better than guessing, given the membership of 1 of 26 classes. In order to increase the effectiveness of the classifier, it was decided to use an iterative approach. After each successive graphological analysis, the model was tuned using the newly acquired characters. Unfortunately, the end result of this test was a failure - the number of correct classifications was only improved by 3, which did not even translate into half a per cent effectiveness. This was most likely because the total amount of data was too small and the characters of the writing too varied to significantly affect the result. Ultimately, further attempts to improve the process were abandoned due to limited working time.

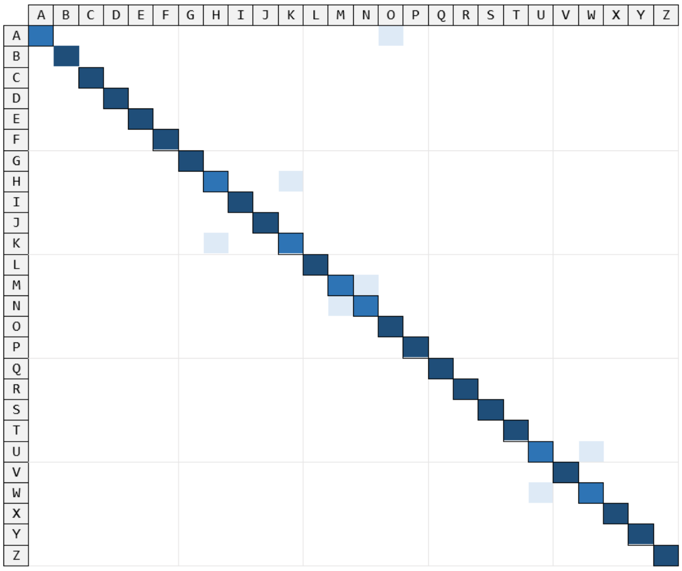

Fig. 8. Non-EMNIST Temperature Confusion Matrix

The most frequently misclassified letters were *'n'* and *'u'*. The remaining characters maintained a relatively high recognition rate, at above 85%. Unfortunately, it was not possible to establish better parameters for the *DNN*, with which it would have been possible to increase the effectiveness of the classifier. It is possible that the use of a different activation function, e.g. *ReLu* with wide applicability in constrained Boltzmann machines, would have contributed to this.

CONCLUSION

The use of the Deep Neural Network (DNN) model exceeded initial expectations, demonstrating a remarkable efficiency of 47%. This result underlines the practicality of using deep neural networks in conjunction with graphological analysis. If the model can effectively identify basic personality characteristics such as openness and honesty, it implies the feasibility of identifying more complex traits such as introversion and extroversion. Unfortunately, the same level of success cannot be attributed to the KAMR model, which performed significantly worse, achieving an effectiveness rate of only 24%. This disparity highlights the inherent challenges of graphology, which often relies on speculative hypotheses that lack empirical support.

It's important to note that this study was conducted on a relatively small cohort of participants from Central Europe and Australia. The results may be significantly different if replicated with individuals from other continents or countries [5].

The development and use of a proprietary *DNN* was resource intensive but yielded satisfactory results. The adaptability offered by the pick of activation functions facilitated the identification of the optimal sigmoidal (logistic) function. The use of a stochastic approach to network interpretation allowed the establishment of thresholds for acceptability of results. In particular, categorising results based on output values proved more effective than alternative methods. Given the use of a 6-point fuzzy response scale in the initial questionnaire, a consistent approach to result interpretation was adopted.

Different threshold values (Table XI) can result in significantly improved effectiveness for both the DA and KAMR models, albeit within marginal percentage variations. The research (Chapter 4) highlights the sensitivity of the results to small adjustments in the thresholds. Combining the optimal parameter ranges, the DA model achieved its highest accuracy, 69%, with responses categorised as '*rather yes*'. Conversely, the least successful identification occurred with features characterised as '*definitely yes*', with a recognition rate of only 12%. KAMR excelled at recognising features classified as *very weak*, achieving a recognition rate of 47%. However, it performed less well on features classified as "certain", with a recognition rate of 5%.

TABLE XI
METHOD FOR CONVERTING GRAPHOLOGICAL ANALYSIS FINDINGS INTO SURVEY DATA

Survey response	*Interpretation of characteristic*	*Suggested value of DNN threshold*
definitely yes	*certain*	*85 – 100%*
almost certainly yes	*very strong*	*70 – 84%*
rather yes	*strong*	*50 – 69%*
rather not	*weak*	*30 – 49%*
almost certainly not	*very weak*	*15 – 29%*
definitely not	*none*	*0 – 14%*

On the basis of the above data, the DA model shows superior performance and, as noted above, holds great promise for effectively identifying personality traits. It must be emphasised that the survey, although based on reputable sources in psychology and graphology, is of amateur origin. The critical endeavour at hand is the development of a much more accurate assessment tool designed to extract personality information from individuals more efficiently. If professional research is conducted in the areas explored in the literature regarding the potential applications of this discussed work, the use of analogous systems in the fields of medicine and forensics becomes highly rational [6]. Such use could potentially revolutionise the prevailing approaches to preventative testing, while significantly enhancing security measures.

In terms of the recommended direction for future work, the most important, and probably the most critical, improvement required to the current software is to extend the classifier's capacity to identify a wider range of features. Currently, this capacity remains relatively limited, resulting in a significant loss of information. Of course, in order to determine the accuracy of the classification, it is imperative to increase the pool of survey questions. It is advisable to delegate this task to an expert in the field. Only after successful development of software with these improvements should the next step involve the creation of a specialised module for more in-depth handwriting analysis.

REFERENCES

[1] Antony, D. John, and O. F. M. Cap. "Personality Profile Through Handwriting Analysis." In Anugraha (Tamil Nadu Capuchin Institute for Counselling, Psychotherapy and Research) Nochiodaipatti Post Dindigul-

624 003 Tamil Nadu, India Tel: 0451–2550100, 2550324, 255083 Email: anugrahacap@ eth. net. Anugraha Publications, 2008.

[2] Amend, K.K. and Ruiz, M.S., 2000. Handwriting analysis: The complete basic book. Red Wheel/Weiser. ISBN: 978-0-87877-050-2

[3] Champa, H.N. and AnandaKumar, K.R., 2010. Artificial neural network for human behavior prediction through handwriting analysis. International Journal of Computer Applications, 2(2), pp.36-41. last accessed 2023/1/17 https://pdfs.semanticscholar.org/9962/3bfbc10270210cc2a4caeda6ae98c49644b2.pdf

[4] Halilaj, E., Rajagopal, A., Fiterau, M., Hicks, J.L., Hastie, T.J. and Delp, S.L., 2018. Machine learning in human movement biomechanics: Best practices, common pitfalls, and new opportunities. Journal of biomechanics, 81, pp.1-11. doi: 10.1016/j.jbiomech.2018.09.009

[5] Kao, H.S., Hoosain, R. and Van Galen, G.P. eds., 1986. Graphonomics: Contemporary research in handwriting. Elsevier. ISBN: 978-0-444-70047-6

[6] Matozza F., Ortiz A., Levy D.: Handwriting analysis in cancer patients clinical-radiological and graphological correlation. last accessed 2023/1/17 https://www.slideshare.net/frankmatozza/HRA-ADOBE-english-presentation

[7] Mehta, Y., Majumder, N., Gelbukh, A. and Cambria, E., 2020. Recent trends in deep learning based personality detection. Artificial Intelligence Review, 53(4), pp.2313-2339. doi:10.1007/s10462-019-09770-z

[8] Obrębska A. Sekrety charakteru zawarte w piśmie. Jak wykonać ekspertyzę grafologiczną. Praktyczny podręcznik. Wydawnictwo Medium, Łódź, 2004. ISBN: 978-83-87025-38-0

[9] Samek, W., Montavon, G., Lapuschkin, S., Anders, C.J. and Müller, K.R., 2021. Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE, 109(3), pp.247-278. doi:10.1109/JPROC.2021.3060483

[10] Roberts, R. and Woodman, T., 2017. Personality and performance: Moving beyond the Big 5. Current opinion in psychology, 16, pp.104-108. doi:10.1016/j.copsyc.2017.03.033

[11] Woda, M. and Batogowski, J., 2020, June. Prediction of selected personality traits based on text messages from instant messenger. In International Conference on Dependability and Complex Systems (pp. 672-685). Springer, Cham. doi:10.1007/978-3-030-48256-5_66

[12] Woda, M. and Oliwa, G., 2023, July. Analysis of Handwritten Texts to Detect Selected Psychological Characteristics of a Person. In International Conference on Dependability and Complex Systems (pp. 327-341). Cham: Springer Nature Switzerland. doi:10.1007/978-3-031-37720-4_30

[13] Wright, T., 2012. Handwriting Recognition with Artificial Neural Networks and OpenCV. CS488-Senior Capstone. https://cs.earlham.edu/~twright09/paper.pdf, last accessed 2023/1/17

[14] Wojtowicz A. Grafologia dla początkujących (praktyczny podręcznik). Wydawnictwo Wolumen, Bytom, 2007. ISBN: 978-83-926158-0-4

[15] Zuckerman, M., 2002. Zuckerman-Kuhlman personality questionnaire (ZKPQ): An alternative five-factorial model. Big five assessment, pp.377-396.