

RESEARCH

Open Access



Predicting the Big Five personality traits from handwriting

Mihai Gavrilescu*  and Nicolae Vizireanu

Abstract

We propose the first non-invasive three-layer architecture in literature based on neural networks that aims to determine the Big Five personality traits of an individual by analyzing offline handwriting. We also present the first database in literature that links the Big Five personality type with the handwriting features collected from 128 subjects containing both predefined and random texts. Testing our novel architecture on this database, we show that the predefined texts add more value if enforced on writers in the training stage, offering accuracies of 84.4% in intra-subject tests and 80.5% in inter-subject tests when the random dataset is used for testing purposes, up to 7% higher than when random datasets are used in the training phase. We obtain the highest prediction accuracy for Openness to Experience, Extraversion, and Neuroticism (over 84%), while for Conscientiousness and Agreeableness, the prediction accuracy is around 77%. Overall, our approach offers the highest accuracy compared with other state-of-the-art methods and results are computed in maximum 90 s, making the approach faster than the questionnaire or psychological interviews currently used for determining the Big Five personality traits. Our research also shows there are relationships between specific handwriting features and prediction with high accuracy of specific personality traits and this can be further exploited for improving, even more, the prediction accuracy of the proposed architecture.

Keywords: Neural networks, Handwriting analysis, Personality classification, Feature classification

1 Introduction

Handwriting has been used for centuries as a way of communication and expression for humans, but only recently its links to the brain activity and the psychological aspects of humans have been studied. The psychological study of handwriting with the purpose of determining the personality traits, psychological states, temperament, or the behavior of the writer is called graphology and is still a debatable domain as it lacks a standard, most of the handwriting interpretations being done subjectively by trained graphologists.

However, there have been various research papers showing the link between handwriting and neurological aspects of humans, one such study being the one of Plamondon [1], where it was shown that the brain forms characters based on habits of writers and each neurological brain pattern forms a distinctive neuromuscular movement which is similar for individuals with the same

type of personality. Therefore, handwriting is, from this perspective, an accurate mirror of people's brain.

Graphologists currently analyze multiple handwriting features in order to assess the psychological aspects of the writer, such as the weights of strokes [2], the trajectory of writing [3], the way the letter “t” or “y” are written [4], as well as other features related to how letters or words are written or how the text is positioned on the page.

In the current paper, we aim to build the first architecture in literature that is able to automatically analyze a set of handwriting features and evaluate the personality of the writer using the Five-Factor Model (FFM). To test this architecture, we propose the first database that links the FMM personality traits to handwriting features, which is a novel aspect of this research paper. The proposed system offers an attractive alternative to the standard FMM questionnaire or psychological interviews that are currently used for evaluating personality, because it is easier to use, it involves less effort, and is faster as well as removes the subjectivity from both subject's (as usually the subject is asked to self-report on a

* Correspondence: mike.gavrilescu@gmail.com

Department of Telecommunications, University “Politehnica” of Bucharest, 1-3 Iuliu Maniu Blvd, 06107 Bucharest 6, Romania

specific questionnaire) as well as clinician's sides (as typically psychologists are reviewing the questionnaire results and share opinions regarding the personality of the individual, opinions which can sometimes be prone to bias such that different psychologists might provide different evaluations). We show that our proposed system offers the highest accuracy compared to other state-of-the-art methods as well as share our findings regarding the relationship between several handwriting features and specific personality traits that can be further exploited to improve, even more, the accuracy of such a system.

In the following section, we present the state-of-the-art in the area of handwriting analysis, focusing on papers related to predicting the psychological traits of individuals. We continue in the subsequent section with describing the two models used (FMM and graphology analysis) followed by a detailed presentation of the three-layer architecture, as well as the classifiers and the structure of the neural network used. Finally, we detail the experimental results and share our findings and conclusions on the results obtained.

2 Related work

As mentioned previously, currently, there is no standard developed in predicting behavior based on handwriting, the majority of graphological analysis being done by specialized graphologists. However, research was conducted in the area of computer science which aimed to create such systems in order to recognize the behavior from handwriting in an easier way and also to standardize the graphological analysis. In the next paragraphs, we present the state-of-the-art in this area as well as several studies which made use of handwriting to determine the psychological traits or mental status of individuals.

Behnam Fallah and Hassan Khotanlou describe in [5] a research with a similar purpose as the one conducted in this paper, aiming to determine the personality of an individual by studying handwriting. The Minnesota Multiphasic Personality Inventory (MMPI) is used for training their system and a Hidden Markov Model (HMM) is employed for classifying the properties related to the target writer, while a neural network (NN) approach is used for classifying the properties which are not writer-related. The handwriting image is analyzed by these classifiers and compared with the patterns from the database, the output being provided in the form of the personality of the writer on the MMPI scale. Their system offers over 70% accuracy at this task. Similarly, in [4], an instrument for behavioral analysis is described with the task of predicting personality traits from handwriting. The approach takes into account the following handwriting features: letter "t," lower loop of the letter "y," the pen pressure, and the slant of writing. A rule-based classifier is used to assess the personality trait of the writer on the Myers-Briggs Type Indicator (MBTI)

scale with also over 70% accuracy. The work of Chen and Tao [6] also provides an interesting exploratory study where they use combinations of Support Vector Machine (SVM), AdaBoost, and k-nearest neighbors (k-NN) classifiers for each of the seven personality dimensions in order to analyze a unique set of handwriting features. Their results are promising with accuracies ranging from 62.5 to 83.9%.

Although not aiming for personality traits, Siddiqi et al. [7] present a system that is able to predict the gender of individuals from scanned images containing their handwriting. A set of features is extracted from their writing samples, and artificial neural networks (ANNs) and Support Vector Machines (SVMs) are used to discriminate between the writing of a male and that of a female. The handwriting features employed are slant, curvature, texture, and legibility, computed in both local and global features. Evaluated on two databases under a number of scenarios, the system is able to predict with over 80% accuracy the gender of the writer. Similarly, in [8], it is proposed a way to describe handwritings based on geometric features which are combined using random forest algorithms and kernel discriminant analysis. The system is able to predict gender with 75.05%, age with 55.76%, and nationality with 53.66% when all the writers were asked to write the same text, and 73.59% for gender prediction, 60.62% for age prediction, and 47.98% for nationality prediction when each subject wrote a different text.

Another interesting research is the one conducted by Gil Luria and Sara Rosenblum [9] which uses handwriting behavior in order to determine the characteristics of both low and high mental workloads. They asked 56 participants to write three arithmetic progressions of different difficulties on a digitizer, and differences are seen in temporal, spatial as well as angular velocity spaces, but less in the pressure space. Using data reduction, they identify three clusters of handwriting types and conclude that handwriting behavior is affected by the mental workload. Zaarour et al. [10] show another interesting research where handwriting is employed to improve the performance of pupils through a system which takes as input different drawings and writings and, by means of a Bayesian network-based model, they can determine the writing style of the child which can be further analyzed by a child psychologist in order to advise parents on how to improve their child education. Similarly, Sudirman et al. [11] present a system that studies the behavior of children based on their handwriting, starting from the assumption that children are the best subjects to be analyzed in the context of handwriting as they are less influenced by cultural background and their cognition rate is evolving very fast. Therefore an automatic system is built which aims to determine the developmental disorders that the children

might be suffering from, with accuracies of over 78%, making the approach attractive for both teachers as well as therapists for patients' monitoring. Researchers in [12] present a system tasked with decreasing the time for job candidate selection in the pre-employment stage using automatic personality screening based on visual, audio, and lexical cues. The system extracts a set of relevant features which are used by a chain of machine learning techniques in order to predict candidates' scores on the Five-Factor Model scale and a classifier is used to combine the prediction results from all the three cues. The experimental results show promising results in terms of performance on first impression database.

Another direction for many studies involving handwriting analysis is the detection of deceit. Luria et al. [13] show such research where a non-intrusive system analyzes the handwriting in the context of healthcare with the purpose of detecting the false information that patients provide about their health. As current ways of determining deception are invasive and do not comply with a clinician-patient relationship, such an approach of using the handwriting as a tool is attractive from research perspectives. Subjects participating in the experiment were asked to write true/false statements about their medical condition on a paper linked to a digitizer. After this first step, the deceptive and truthful writings of all the subjects are compared and used to divide the subjects into three groups according to their handwriting profiles. It is found that the deceptive writing takes longer to write and is broader and the two types of writings show significant differences in both spatial and temporal vectors. In [14], similar research is conducted, based on the same assumption that for people it is easier to tell the truth than to lie; hence, we need to see changes in both velocity and temporal spaces when analyzing the handwriting features. Conducted in 11 languages, this research demonstrates the same point as in [13], with the specific purpose of helping managers pinpoint sudden emotional changes and decode handwritten messages to reveal the true meaning of those messages as well as detect lies.

Besides detecting deceit, the handwriting is also used for predicting physical diseases. Researchers in [15] present a study where diabetics' disease can be predicted with over 80% accuracy from handwriting. Similarly, in [16], the handwriting is used to predict micrographia (the decrease in the size of letters as well as the velocity and acceleration of writing) that is commonly associated with Parkinson's disease (PD). The system, tested on PD-diagnosed patients, offers over 80% accuracy on 75 tested subjects. The study described in [17] is another research analyzing the link between the handwriting and children with autism spectrum disorder (ASD), knowing the fact that children with ASD have several weaknesses

in handwriting. Boys aged 8–12 years and diagnosed with ASD were asked to take a digitized task in order to determine the handwriting performance using advanced descriptive methods. The study shows moderate to large links between handwriting performance and attention, ASD symptoms and motor proficiency, providing a relationship between handwriting and the ASD symptoms in terms of severity, attention, and motor behaviors.

Since handwriting analysis is a complex task requiring multiple techniques in order to analyze the multitude of handwriting features, there is a wide range of methods typically employed. For offline handwriting analysis, the normalization of the handwritten sample is the first step in order to ensure any possible noise is filtered out. As part of normalization phase, methods for removing the background noise (morphological approaches or Boolean filters are typically used [18]), sharpening (Laplace filters, Gradient masking or unsharp masking [19]), and contrast enhancement (unsharp mask filters [20]) are essential for ensuring the analysis of the handwriting is done with high accuracy. Also, as the contour of the written letters is essential for this task, methods for contour smoothing also need to be used, the most common ones being the local weighted averaging methods [21]. After all these processing steps are applied to the handwritten sample, the image needs to be compressed and converted to greyscale and different types of thresholding techniques can be employed for this step [22]. Post-compression, the written text needs to be delimited through page segmentation methods where techniques for examining the foreground and background regions are employed, the most common one being the white space rectangles segmentation [23]. One of the most challenging tasks is the one of segmenting the handwritten image into text lines and words. For this, the Vertical Projection Profile [24] method has shown the most promising results and this is the one that we use in this paper for both row and word segmentation. Regarding feature classification, different classifiers are used successfully for each of the handwriting features. For example, for lowercase letters "t" and "f," the most common method used is template matching, for writing pressure gray-level thresholding methods are employed [22], while for connecting strokes the Stroke Width Transform (SWT) has shown the best classification accuracy compared to other state-of-the-art methods. In the following sections, we present in detail the classifiers used for each of the handwriting features analyzed in the current paper.

With all these in mind, the current research proposes a novel non-invasive neural network-based architecture for predicting the Big Five personality traits of a subject by only analyzing handwriting. This system would serve as an attractive alternative to the extensive questionnaire

typically used to assess the FMM personality traits and which is usually cumbersome and non-practical, as well as avoid the use of invasive sensors. We focus our attention on handwriting because it is an activity familiar to almost everyone and can be acquired fast and often.

In the next section, we present the theoretical model and the architecture of our system.

3 Methods

3.1 Theoretical model

As mentioned in the previous section, our research is proposing a novel non-invasive neural network-based architecture for predicting the Big Five personality traits of an individual solely based on handwriting. Therefore, our study is based on two psychological tools: Big Five (Five-Factor Model—FMM) [17] and graphological analysis. We detail both these instruments in the next subsections.

3.1.1 Big Five (Five-Factor Model)

Big Five (Five-Factor Model) [25] is a well-known model for describing the personality of an individual. It is based on five basic personality traits which are grouped in sub-factors, as follows:

- *Openness to Experience*: refers to people who can easily express their emotions and have a desire for adventure, appreciation for art, and out-of-the-box ideas. Typically, on this scale, people are rated based on the dichotomy: *consistent* vs. *curious*;
- *Conscientiousness*: refers to people who are dependable, have a predilection towards behaviors which are carefully planned, and are oriented towards results and achievements. On this scale, people are rated based on the dichotomy: *organized* vs. *careless*;
- *Extraversion*: refers to people who easily express positive emotions, like other's people company, are assertive, and talkative. On this scale, people are rated on the dichotomy: *outgoing* vs. *solitary*;
- *Agreeableness*: refers to people who have a tendency to be compassionate instead of suspicious, as well as helpful, and tempered. On this scale, people are rated based on the dichotomy: *compassionate* vs. *detached*;
- *Neuroticism*: refers to people who lack emotional stability and control and tend to experience negative emotions easily, such as anger and anxiety, as well as a vulnerability to depression. On this scale, people are rated based on the dichotomy: *nervous* vs. *confident*.

FMM is successfully used on a wide variety of tasks. The research conducted in [26] shows that compared to

other methods for assessing the personality of an individual, FMM offers more stability over time, the Big Five personality types reaching their stability peak 4 years after starting work. FMM has also proved to be useful in determining personality disorders, such as depression or anxiety, and even substance use, and was shown to be an indicator for different physical diseases, such as heart problems, cancer, diabetes or respiratory issues [27]. It is also successfully used in the area of career development and counseling as well as team performance, but also for improving learning styles and the academic performances of students [28]. Because of its extensive use and broad perspective of applications we employ it in our current study.

3.1.2 Graphological analysis

Typically, when analyzing the handwriting of an individual, graphologists are looking for a specific set of features, each of them conveying a specific message [29]. The main handwriting features used and the ones that we explore in the current paper are the following: baseline, word slant, writing pressure, connecting strokes, space between lines, lowercase letter “t,” and lowercase letter “f.” Examples of each of these features and their types as explained in [30] can be observed in Table 1.

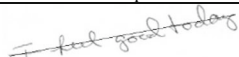
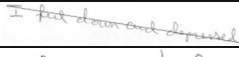
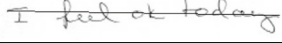

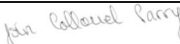

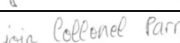
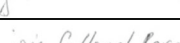

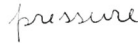

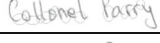
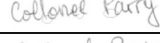
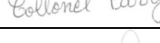
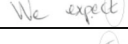

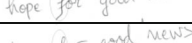
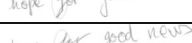
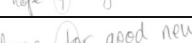
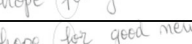

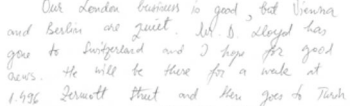
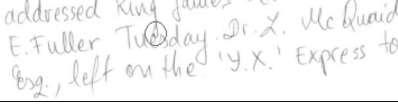
The *baseline* of the handwriting refers to the line on which the written words flow. It is further divided into *ascending baseline* (associated with optimistic people), *descending baseline* (associated with pessimistic people and over-thinkers), and *leveled* (associated with people with high levels of self-control and reasoning).

The *word slant* refers to how the words are written in terms of inclination/slant. Possible slant types are the following: *vertical slant* (associated with people who can easily control their emotions), *moderate left slant* (associated with people who find it hard to express emotions), *extreme left slant* (associated with people who want to be in permanent control and suffer from self-rejection), *moderate right slant* (associated with people who can easily exteriorize their emotions and opinions), and *extreme right slant* (associated with people who are impulsive and lack self-control).

The *writing pressure* refers to the amount of pressure that is applied to the pen on the paper: *light writer* (refers to people who hardly get affected by traumas), *medium writer* (refers to people who are usually affected by pain or traumas), and *heavy writer* (refers to people who are deeply affected by traumas and emotions).

Connecting strokes refer to how the letters composing words are connected to each other. These are dichotomized into *not connected* (refers to people that can hardly adapt to change), *medium connectivity* (refers to people who can adapt to change as well as like changing

Table 1 Handwriting features and their corresponding types [30]

Handwriting feature	Type	Example
Baseline	Ascending baseline	
	Descending baseline	
	Leveled baseline	
Word slant	Vertical slant	
	Moderate left slant	
	Extreme left slant	
	Moderate right slant	
	Extreme right slant	
Writing pressure	Light Writer	
	Medium Writer	
	Heavy Writer	
Connecting strokes	Not connected	
	Medium connectivity	
	Connected letters	
Lowercase letter "t"	Very low "t"-bar	
	Very high "t"-bar	
Lowercase letter "f"	Angular point	
	Angular loop	
	Narrow upper loop	
	Cross-like	
	Balanced	
Spaces between lines	Lines separated, evenly spaces	
	Lines crowded together with overlapping loops	

environments), and *connected letters* (refers to people who can quickly adapt to change).

Lowercase letter "t" typically refers to how the t-bar on the letter "t" is written. If it is written *very low*, it is an indication of low self-esteem, if it is written *very high* it is an indicator of high self-esteem.

Lowercase letter "f" refers to how the letter "f" is written. If it has an *angular point*, the person can be easily revolted, if it has an *angular loop*, the person has a strong reaction to obstacles, if it has a *narrow upper loop*

it is usually associated with narrow-minded people, if it is *cross-like* it is associated with an increased level of concentration, and if it is *balanced* it is an indicator of leadership abilities.

Spaces between lines refer to the space left by the writer between two consecutive lines. We can have *lines separated, evenly spaced* (associated with people who can organize work and have clear thoughts) or *lines crowded together with overlapping loops* (associated with people with confused thinking and poor organizational skills).

3.2 Proposed architecture

We design the architecture on three layers as follows: a base layer where the handwriting sample is normalized and the handwriting features are acquired, an intermediary layer where a Handwriting Map is built based on the handwriting features provided by the base layer, and a top layer where a neural network is used in order to determine the Big Five personality type of the writer. In the following subsections, we present each of these layers in detail.

3.2.1 Base layer

The base layer has the primary purpose of converting the scanned handwriting in the set of handwriting features mentioned in previous sections. A flowchart of the central processing blocks of this layer can be observed in Fig. 1.

The main steps are detailed below:

- Normalization:
 - *Noise reduction*: in order to remove the noise added by the scanning device or the writing instrument which typically cause distortion, disconnected strokes or unwanted lines or points, we use three filters. Boolean filters are used for removing the textured background as they were shown to outperform other morphological methods for cases when the text is written on highly textured backgrounds both in terms of accuracy and processing time [18]. For sharpening, we use the ramp width reduction filter as it is known as the most effective algorithm for ramp edge sharpening [19]. Adaptive unsharp masking is employed for adjusting the contrast [20] which is widely used as an effective method for contrast enhancement.
 - *Contour smoothing*: in order to reduce the possible errors that appear due to unwanted movement of writer's hand during writing we use an optimal local weighted averaging method [21] ensuring that these glitches are filtered out and only the strokes relevant for our analysis are kept. We opted for this algorithm as opposed to other less complex local weighted averaging methods because this method is known to provide more accurate estimations of contour point positions, tangent slopes, or deviation angles which are essential for our handwriting analysis task.
 - *Compression*: we used global thresholding in order to convert the color images to binary. We used the histogram modified by integral ratio [22] in order to determine the global threshold value as it was shown to provide better performance compared to other compression techniques.
 - *Isolation of handwriting in the page*: in order to only keep the handwritten text for the next steps of our handwriting analysis task, we use the white space thinning method [23] as it is a simple and fast method for this task; hence, we cut the page recursively on the two dimensions until only the handwritten text is delimited.
- *Row segmentation*: For row segmentation, we use the Vertical Projection Profile (VPP) method [24] as it was showed to provide the best classification accuracy compared to other row and word segmentation methods. We, therefore, analyze the sum of pixels for each row in the image and determine as row boundaries those with a sum lower than 8% of the highest pixel sum in the text sample. The threshold of 8% was chosen through trial-and-error after conducting tests on 100 handwriting samples using a leave-one-out approach and the average accuracy for correct row segmentation was 98.5%. Following this step, every row in the handwritten text has a corresponding bounding rectangle.
 - *Spacing between lines feature*: based on the bounding rectangles delimiting each row from handwriting, we determine the amount of overlap between two consecutive rows. If the overlap is higher than 15% of the sum of both row bounding rectangles' surfaces, we consider that the rows are *crowded together*, otherwise, they are considered *evenly spaced*. The 15% threshold was determined to be optimal for ensuring over 98% accurate classification of this handwriting feature.
 - *Baseline feature*: in order to determine the baseline features for each row, we use the method depicted in [31] where we study the pixel density of each segmented row rectangle and we rotate the rectangle within the -30° and $+30^\circ$ angle thresholds until the highest pixel density is horizontally centered. This method is broadly used for baseline feature extraction offering higher classification accuracy and faster convergence compared to other state-of-the-art methods. If the rotation needed to align the highest pixel density horizontally is within $[-5^\circ; +5^\circ]$, we consider that we have a *leveled baseline*, if it is within $[-30^\circ; -5^\circ]$, an *ascending baseline*, and within $[+5^\circ; +30^\circ]$ a *descending baseline*.
 - *Writing pressure feature*: we use the standard gray-level thresholding method that is widely used for the task of writing pressure classification [32] with high accuracy and fast convergence. We analyze the grayscale values for the segmented rectangle containing the row and we calculate the average for the segmented row. The result is

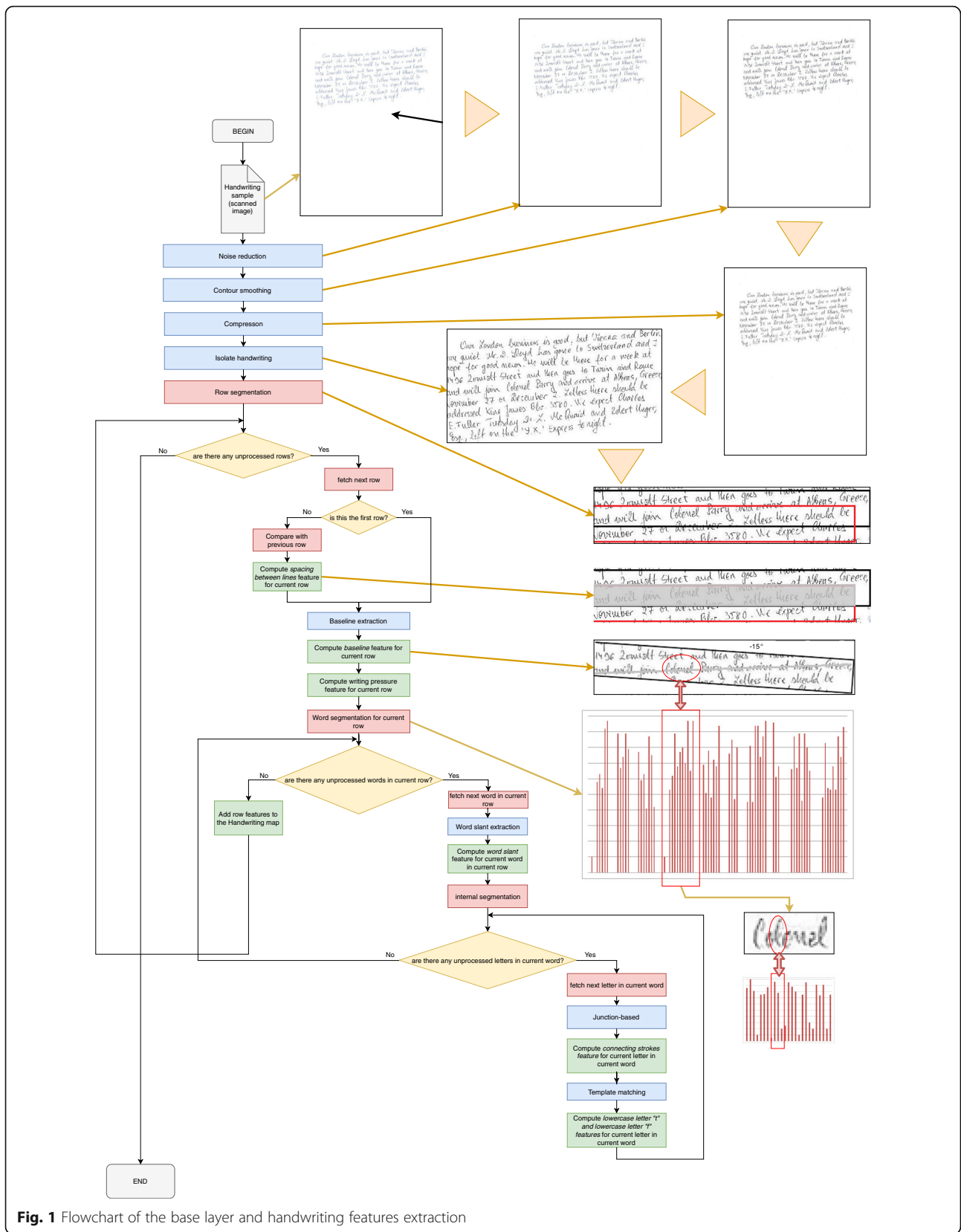


Fig. 1 Flowchart of the base layer and handwriting features extraction

classified as *light writer* for a value within 25 and 50%, *medium writer* for a value within 10 and 25%, and heavy writer for a value within 0% (absolute black) and 10%.

- *Word segmentation*: In order to further segment the words in a row, we use the same VPP method [24] that we employed for row segmentation as it was shown to provide better classification results than other state-of-the-art methods. We compute the height of the row first and use it for comparison purposes in order to determine whether a space between two strokes is indeed an inter-word space or not. We generate a vertical projection profile where we determine the pixel density for each vertical column and we determine the columns with low density, which are considered candidates for spaces between words. As there are cases when such gaps might not correspond to actual word separation spaces, we consider them spaces only if the number of consecutive columns with low density is not lower than 10% of the row height. The 10% threshold was determined through trial-and-error after testing the algorithm on 100 handwritten samples and obtaining the highest word segmentation accuracy of 98.2%. The segmented words are bounded by rectangles similarly as in the row segmentation case.
 - *Word slant feature*: in order to determine the word slant feature, we use the same technique described in [33]. We calculate the vertical pixel density histogram for each angle within $[-20^\circ; +20^\circ]$ and for each column in the histogram we determine the number of pixels and divide it with the highest and lowest pixel in the analyzed word segment. The values from all columns are then summed and the angle where the computed sum is the highest is considered to be the slant of the writing. We then classify the word slant as follows: if the angle is within $[-2.5^\circ; +2.5^\circ]$, it is a *vertical slant*; if it is within $[-7.5^\circ; -2.5^\circ]$, it is a *moderate left slant*; if it is lower than -7.5° , it is an *extreme left slant*; if it is within $[+2.5^\circ; +7.5^\circ]$, it is a *moderate right slant*; and if it is higher than $+7.5^\circ$, an *extreme right slant*.
- *Letter segmentation*: for segmenting the letters from each delimited word segment, we use the stroke width transform (SWT) [34] method for determining the average stroke width of the word. We use this operator because it is local and data dependent, making it faster and more robust than other methods that need multi-scale computations. We then create a projected profile for the word segment and determine the columns where the projection value is lower than 8% than the highest projected value in the word. For the identified

strokes, we determine their width and compare it with the word's average stroke width. If it is lower than 50%, we create a bounding box surrounding the character and we crop out the bounding box from the word segment. The 50% threshold was determined to be optimal after testing the method on 100 handwritten samples and obtaining 98.2% accuracy for letter segmentation. With the remaining part of the word segment, the process is repeated until all letters are identified.

- *Connecting strokes feature*: in order to compute the connecting strokes feature, we use the letter segmentation algorithm previously described and we compare each stroke width connecting two consecutive letter bounding boxes with the average stroke width of the word. If the stroke width is below 10% of the average stroke width of the word, we consider it as *not connected*; if it is above 30%, we consider it *connected*; and if it is between 10 and 30%, it is considered as having *medium connectivity*.
- *Lowercase letter "t" feature*: as letters are now delimited in corresponding bounding boxes, we use template matching to compare each letter to a set of predefined templates of letter "t" from the Modified National Institute of Standards and Technology (MNIST) database [35]. The templates were previously divided into the two categories of letter "t" (*very low "t" bar* and *very high "t" bar*), and we use Euclidean similarity to measure the letter matching to the chosen MNIST prototypes. The threshold matching determined as optimal through trial-and-error is 0.88 and the accuracy for detecting the right letter "t," tested on 100 handwriting samples with a leave-one-out approach, is 98.2%.
- *Lowercase letter "f" feature*: we use the same method depicted for letter "t" with the difference that the letter "f" templates from the MNIST database are divided into five categories corresponding to the ones analyzed (angular point, angular loop, narrow upper loop, cross-like and balanced). The threshold, in this case, is 0.92 corresponding to an accuracy of 97.5%.

3.2.2 Intermediary layer (Handwriting Map)

As we previously mentioned, the base layer offers as inputs to the intermediary layer the handwriting feature types for each letter in the exemplar. These are coded in the Handwriting Map (HM) using a binary code. Therefore if, for example, connecting strokes have medium connectivity, the code for this is 010 (0—connected, 1—medium connectivity, 0—not strongly connected). Typically, for each analyzed letter, we have the following possible codes associated with

each of the seven handwriting features that all compose one row in the HM:

- *Baseline: position 1 to 3*: possible values are 100—ascending, 010—descending, 001—leveled;
- *Connecting strokes: position 4 to 6*; possible values are 100—not connected, 010—medium connectivity, 001—strongly connected;
- *Word slant: position 7 to 11*; possible values are 10000—vertical slant, 01000—moderate left slant, 00100—extreme left slant, 00010—moderate right slant, 00001—extreme right slant;
- *Writing pressure: position 12 to 14*; possible values are 100—light writer, 010—medium writer, 001—heavy writer;
- *Lowercase letter “t”: position 15 to 16*; possible values are 10—very high; 01—very low; 00—not a lowercase letter “t”;
- *Lowercase letter “f”: position 17 to 21*; possible values are 10000—cross-like, 01000—angular loop, 00100—angular point, 00010—narrow upper loop, 00001—balanced; 00000—not a lowercase letter “f”;
- *Space between the lines: position 22 to 23*; possible values are 10—evenly spaced, 01—crowded together.

Therefore any row entry in the map has the following structure: $[100][010][00010][100][00][00010][10]$ (which means ascending *baseline*—100, *medium strokes connectivity*—010, *moderate right slant*—00010, *light writer*—100, *not a lowercase letter “t”*—00, *Narrow Upper Loop on lowercase letter “f”*—00010, *evenly spaced lines*—10).

Two observations should be made about the above-constructed mapping:

- For *baseline*, we might have the same code for all letters;
- For *space between the lines*, we might have the same code for all letters that are associated with a row in the handwritten sample.

Therefore, each letter in the handwriting sample generates a row in the HM in the form of a binary code which is then used in the top layer in a pattern recognition task in order to determine the Big Five personality traits.

3.2.3 Top layer

As we have detailed earlier, we have an HM that contains for each letter its handwriting features in the form of a binary code. Therefore, the HM is a matrix containing all the letters in the handwriting exemplar together with their coded features and based on this the system should be able to determine the Big Five personality trait of the writer.

As the task is a pattern recognition task and also considering that our architecture is bottom-up with no feedback loops, we use a feed-forward neural network. Also, with the same premises in mind, the training method used is *backpropagation*, which has proven to be very effective and offers fast learning in similar cases [36].

We define only one neural network that is called the Five-Factor Model–Neural Network (FFM-NN). In order to avoid overfitting it by fetching all the letters from the exemplar, we fetch them by rows and we consider that we do not have more than 70 letters on each row. If a row in the handwritten sample has more than 70 letters, only the first 70 are analyzed. More than this, this approach offers the ability to have multiple tests done on the neural network and we can average the results in order to reach more conclusive ones. As we have 23 entries for each row in the HM, in total we have 1610 input nodes in FFM-NN.

The output layer contains five nodes for each of the five dimensions of FMM. Each node computes a 0 if the subject is found on the lower side of the analyzed dimension, and 1 if it is found on the higher side of the dimension (e.g., a 1 for Openness to Experience means that the subject is more curious than consistent, while a 0 for Neuroticism means that the subject is more inclined towards being nervous than confident).

If we consider N_{in} the number of input training vectors and an N-dimensional set of input vectors for the FFM-NN neural network $X^{FFM-NN} = \{x^{FFM-NN}_n\}$, $n = 1, 2, \dots, N_{in}$, so that $x^{FFM-NN} = [x^{FFM-NN}_1, x^{FFM-NN}_2, \dots, x^{FFM-NN}_N]^T$, and a K_{out} the number of output vectors and K-dimensional set of output vectors $Y^{FMM-NN} = \{y^{FMM-NN}_k\}$, $k = 1, 2, \dots, K_{out}$ so that $y^{FMM-NN} = [y^{FMM-NN}_1, y^{FMM-NN}_2, \dots, y^{FMM-NN}_K]^T$, and if we denote the matrix of weights between input and hidden nodes, $W^{FFM-NNH}$, the matrix of weights between the hidden nodes and the output nodes $W^{FMM-NNO}$ with L the number of hidden nodes, and f^{FMM-NN}_{1a} and f^{FMM-NN}_{2a} the activation functions, the expression form for the output vectors can be written as follows:

$$y^{FFM-NN}_k = f^{FMM-NN}_{2a} \left(\sum_{l=0}^L w^{FMM-NNO}_{lk} f^{FMM-NN}_{1a} \left(\sum_{n=0}^{N_{in}} w^{FMM-NNH}_{nl} x^{FMM-NN}_n \right) \right) \quad \text{with } k = 1, 2, \dots, K_{out} \quad (1)$$

The input features for each letter on a row is fetched to the input nodes which then distributes the information to the hidden nodes and computes the weighted sum of inputs sending the result to the output layer through the activation function. In backpropagation stage, the Average Absolute Relative Error (AARE) (2) is calculated as the difference between what is expected (y^{FMM-NN}_e) and what is determined

(y_p^{FMM-NN} with $p = 1, 2 \dots N_{in}$) and $W^{FMM-NNH}$ and $W^{FMM-NNO}$ weight matrices are calibrated in order to minimize the $AARE^{FMM-NN}$:

$$AARE^{FMM-NN} = \frac{1}{N_{in}} \sum_{p=1}^{N_{in}} \left| \left(\frac{y_p^{FMM-NN} - y_e^{FMM-NN}}{y_e^{FMM-NN}} \right) \right| \quad (2)$$

With the purpose of +/- balance in the hidden layer, the activation function chosen for the input layer is tanh, also considering it offers fast convergence and has a stronger gradient than the sigmoid function. Because the final task of the neural network is a predictive one, we use sigmoid as activation function for the hidden layer, taking into account its non-linearity and that its output is in the range of [0,1]. Conducting various tests, through trial-and-error, we determined that the optimal number of hidden nodes in order to avoid overfitting is 1850. The optimal learning rate is determined as 0.02, the optimal momentum is 0.4, and 200,000 training epochs are needed to train the system in an average of 8 h on an Intel i7 processor computer. We use Gradient Descent to learn the weights and biases of the neural network until AARE is minimized and, in order to ensure an even spread of the initial weights, we use the Nguyen-Widrow weight initialization. The structure of the neural network can be observed in Fig. 2.

3.3 Overall architecture

3.3.1 Training database and handwriting text samples

For testing the above-described architecture, we create our database containing both handwritten exemplars as

well as the FMM personality trait of the writer. In collecting this, we involved 128 individuals, out of which 64 were males and 64 females, with ages between 18 and 35, all of them participating to this experiment in accordance and aware of the Helsinki Ethical Declaration.

Each of the 128 subjects was asked to take the FMM questionnaire as well as provide six handwriting samples. The FMM questionnaire results were analyzed by specialized psychologists to assess their results on the five personality dimensions. In what it concerns the six handwriting samples, two of them are a predefined text representing the *London Letter* [32], a standard exemplar broadly used by graphologists for handwriting analysis, while the others are minimum 300 words texts that subjects could write freely and randomly. All text samples are collected in the English language.

To summarize, for each subject involved in training we have their corresponding FMM personality dimensions results as well as six handwriting samples, out of which two are the London Letter.

In Fig. 3 we can observe an example of the London Letter collected from one of the subjects. The London Letter is chosen because of the handwriting features that we are collecting, such that lowercase letter "t" is assessed at the beginning (e.g., "to", "then", "tonight"), middle (e.g., "Switzerland", "Letters"), and end (e.g., "quiet", "expect") of words, lowercase letter "p" is analyzed at the beginning of words (e.g., "for") or intercalated (e.g., "left") as well as other situations that pose difficulties to writers and help us better discriminate

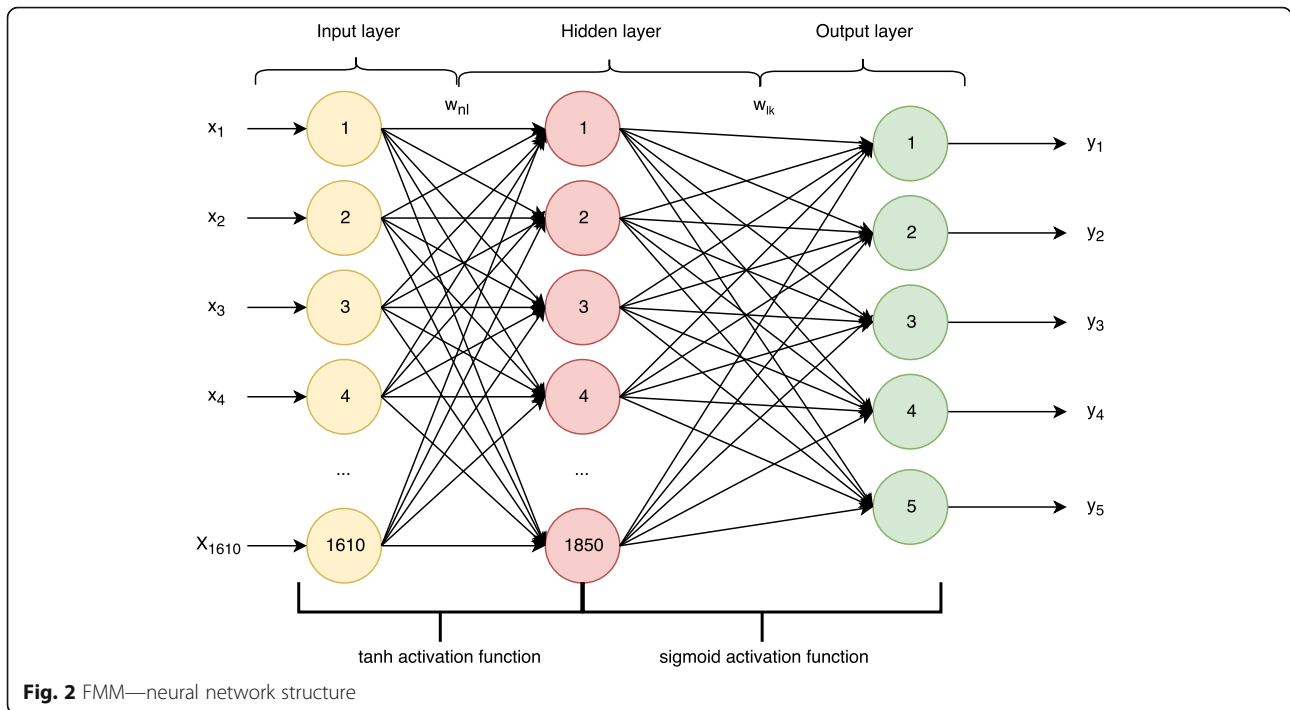


Fig. 2 FMM—neural network structure

between other handwriting features, such as: words starting with uppercase (e.g., Zermott Street), group of longer words (e.g., “Athens, Greece, November”), words containing doubled letters (e.g., “Greece”, “Zermott”), use of letters that need additional strokes (such as x, z, i, j; e.g., “Express”, “Switzerland”, “Vienna”, “join”), and intercalating numbers and/or punctuation (e.g., “King James Blv. 3580.”).

In the following section, we present the training as well as testing stages and how they use the above-described database.

3.3.2 Training and testing phases

The proposed architecture is built using 55,000 code lines in Scala programming with Spark Library. The testbed is functioning on an i7 processor with 8GB of RAM and it is designed to work in two stages: training and testing. The overall architecture can be seen in Fig. 4.

In the *training stage*, the FMM-NN needs to be trained to learn the handwriting patterns and compute the right values for the five personality dimensions. We, therefore, use a set of handwriting samples as training samples that are fetched to the base layer. The handwriting samples are first normalized, then the words are split into letters and the handwriting features for each letter are extracted and sent to the intermediary layer. In the intermediary layer, the HM is built which contains a row for each letter from the handwritten sample in the form of binary codes as previously presented. Every time we have handwritten features collected for 70 new letters, these are fetched to the FMM-NN which is trained via backpropagation so that its output is the one obtained from the FMM questionnaire. When AARE is low enough and the training samples are finished the system is considered trained and can be tested.

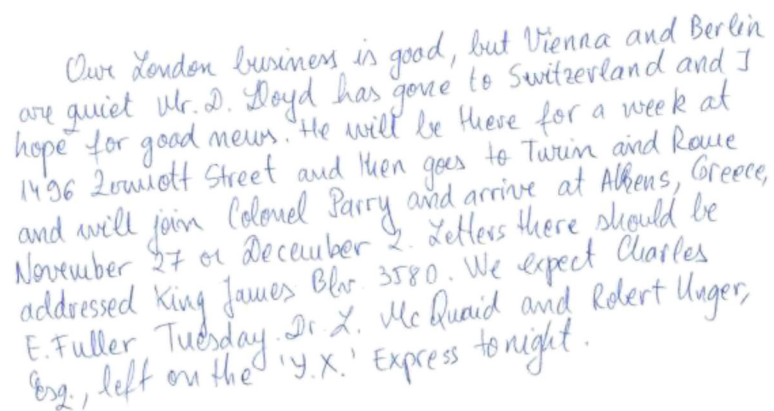
In the *testing stage*, the analyzed handwriting exemplar is also normalized and split into letters in the base layer. The letters are then analyzed and their features are

determined and sent to the intermediary layer which computes the HM. When 70 new letters are computed in HM, these are sent to the FMM-NN which provides an output representing its predicted FMM personality dimensions in the form of five binary codes, as previously explained. When on five consecutive rows (five sets of 70 letters) we have the same binary codes, the system considers that those are the personality dimensions of the writer and outputs the final result. If there are no five consecutive rows generating the same binary output (meaning that different personality traits are detected in any five consecutive rows), the result is flagged as *Undefined*. We chose five consecutive rows as they correspond to an average sized word (of five letters) and we determined that reducing or increasing this threshold results in lower system accuracies.

In the next section, we show the experimental results after testing the architecture as well as a comparison with state of the art.

4 Experimental results and discussion

As we described previously, due to the lack of a publicly available database that would relate the handwriting features with FMM, we built our database to support this study. The database contains handwritings collected from 128 subjects (64 females and 64 males), with ages between 18 and 35 years old as well as their results after filling in the FMM questionnaire which was subsequently analyzed by specialized psychologists to ensure the FMM personality traits are evaluated correctly. For testing the degree of generalization of the proposed approach when dealing with random handwritten text and the influence of the predefined handwritten text in both training and test phases, the database is divided into two main datasets: controlled dataset (consisting of handwriting samples where subjects were asked to write a predefined text—the *London letter*), and the random dataset (consisting of handwriting samples where subjects wrote a



Our London business is good, but Vienna and Berlin are quiet. Mr. D. Lloyd has gone to Switzerland and I hope for good news. He will be here for a week at 1436 Lombard Street and then goes to Turin and Rome and will join Colonel Parry and arrive at Athens, Greece, November 27 or December 2. Letters here should be addressed King James Blv. 3580. We expect Charles E. Fuller Tuesday. Dr. X. McQuaid and Robert Unger, Esq., left on the 'Y.X.' Express tonight.

Fig. 3 Handwritten sample of The London Letter

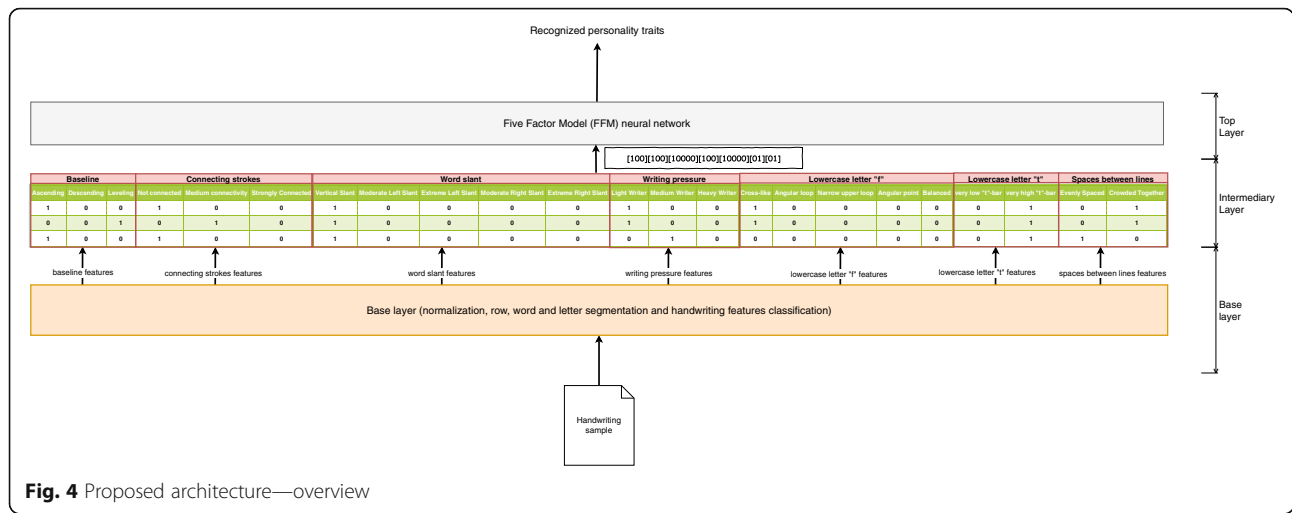


Fig. 4 Proposed architecture—overview

minimum 300 words text freely). Also for testing purposes, in order to determine the ability of the proposed approach to recognize the FMM features of a writer that was not involved in training, we divide the database in writer-specific datasets which contain handwritings only from one specific writer. Each sample from the database is therefore tagged with both the type of dataset to which it pertains (controlled or random) as well as a unique code specifying the writer. The tests conducted in both the intra-subject and inter-subject methodologies are presented in the following sections.

4.1 Own database tests

4.1.1 Intra-subject methodology

Intra-subject methodology refers to training and testing the system on handwriting samples coming from the same writer. We, therefore, use *n*-fold cross-validation for each writer-specific dataset taking also into account the handwriting type (controlled or random). For example, for determining the accuracy of the method when the controlled dataset is used both in test and training phases, since we have only two samples for each writer, we use leave-one-out cross-validation where one of the samples is involved in training and the other is used for testing and vice-versa. Similarly, for determining the accuracy of the method when the random dataset is used for training and the controlled dataset of testing, we train the system on the writer-specific random dataset (containing four samples) and we test it on the writer-specific controlled dataset (containing two samples) via *n*-fold cross-validation. The tests are repeated for all 128 users and the results are averaged and are detailed in Table 2.

We observe the highest prediction accuracy when the system is trained and tested on the controlled dataset reaching 85.3% prediction accuracy, however when we

use the same controlled dataset for training and we test the proposed approach on samples from the random dataset the accuracy does not decrease by much, reaching 84.4%. This is an important observation as it shows that the need for predefined handwritten texts is only for training purposes, while for testing we can use random texts which perform roughly similar to the predefined one. Similarly, when the controlled dataset is used for training, the cases where the personality type is flagged as *Undefined* is the lowest (0.2%), also sustaining the idea that the controlled dataset adds more value to the prediction accuracy when used in training stage as opposed to the random one. This indicates that if the text exemplar used for training handwriting samples is adequately chosen in order to train the neural network on all the analyzed features, using such an application we do not need a standard text for testing and we can ask the subject to write any text they like, making the approach more flexible and easy to use.

The highest prediction accuracies are obtained for Openness to Experience (88.3%—when the system was trained on the controlled dataset and tested on the random dataset), followed by Extraversion (87.4%), Neuroticism (85.3%), while for Conscientiousness and Agreeableness the results are lower, around 80%.

The average number of rows needed to compute the FMM personality types is 9 for the case where the controlled dataset is used in training and the random one for testing and maximum 14 when the random dataset is used for training. Typically, for a row to be computed it takes an average of 5 s, hence the system provides the FMM personality type in no more than 45 s when the controlled dataset is used in training, making the approach fast and attractive for clinicians as an alternative to the FMM questionnaire or psychological interviews.

Table 2 Big Five personality prediction accuracies and average number of rows in intra-subject tests

Type of training samples	Type of test samples	Number of training samples	Recognition rate (%)—Openness to Experience	Recognition rate (%)—Conscientiousness	Recognition rate (%)—Extraversion	Recognition rate (%)—Agreeableness	Recognition rate (%)—Neuroticism	Recognition rate (%)—overall (Undefined excluded)	Average % of cases where personality traits were Undefined	Average number of handwritten rows analyzed until the neural networks converge to a result
C*	C*	1	89.2 ± 0.1	82.3 ± 0.1	88.2 ± 0.05	81.2 ± 0.15	88.4 ± 0.05	85.3 ± 0.15	0.1	4
C*	R*	2	88.3 ± 0.1	80.2 ± 0.1	87.4 ± 0.07	79.8 ± 0.15	86.3 ± 0.07	84.4 ± 0.15	0.2	9
R*	R*	4	79.3 ± 0.2	71.4 ± 0.25	78.3 ± 0.1	70.5 ± 0.2	76.5 ± 0.1	75.2 ± 0.2	7	14
R*	R*	3	75.4 ± 0.2	70.3 ± 0.25	75.5 ± 0.1	68.4 ± 0.15	74.2 ± 0.1	72.1 ± 0.2	10	16
C* + R*	C* + R*	3	87.5 ± 0.15	79.5 ± 0.2	86.5 ± 0.15	78.5 ± 0.15	85.8 ± 0.15	83.4 ± 0.15	0.3	9

C* controlled, R* random

4.1.2 Inter-subject methodology

In inter-subject methodology, we train the system with handwriting samples coming from different writers than those used for testing in order to determine the ability of the proposed approach to extrapolate the trained data to new writers. We used n -fold cross-validation, keeping the database division in controlled and random datasets, and ensuring that handwritings from the writer tested have not been used for training. For example, for training the accuracy of the system when trained on handwritings containing a predefined text (controlled dataset) and testing on handwritings with random text in inter-subject methodology, we use the controlled handwritings from all subjects except the one used for training (2 controlled samples/subject \times 127 subjects used in test = 254 samples), and we test using n -fold cross-validation on random handwriting samples from the remaining subject (four samples). The tests are repeated until all subjects and all their samples are used in the testing phase and the averaged results are detailed in Table 3. To note that we also conduct several tests where we decreased the number of subjects involved in training in order to analyze the change in accuracy when the number of subjects is increased.

Similarly to the intra-subject methodology, the highest prediction accuracy is obtained when the controlled dataset is used for both training and testing and when the system is trained on the highest number of subjects. In this case, the overall prediction accuracy is 84.5%. It is interesting to observe that reducing the number of subjects involved in training does not result in many decreases in terms of prediction accuracy, such that if we use only 96 subjects in training the prediction accuracy is 1.8% lower and when we use 64 subjects in training it decreases with about 1.6% more. This little decrease as well as the fact that high accuracies are obtained when the controlled dataset is used for training and the random dataset for test (78.6%) compared to when the random datasets are used for both training and testing (when the prediction accuracy was 6% lower) provides the same conclusion as in the intra-subject methodology, that the controlled dataset adds more value to the performance of the system if used in the training stage, helping the system learn better the handwriting features. Once learned, for testing purposes random texts can be used in the handwriting sample, providing only 5% lower accuracy, but making the system more practical (in the sense that the subject can write freely whatever text he/she wants). Similarly as in the case for intra-subject methodology, the fact that the number of cases where the personality type is flagged as *Undefined* is lower when the controlled dataset is used for training, with a maximum of just 0.7%, is another indicator that using the controlled dataset in

the training stage improves the prediction accuracy by improving system's ability to discriminate between different FMM personality types.

As in the intra-subject tests, in inter-subject ones, the highest prediction accuracy is obtained for Openness to Experience (88.6%), Extraversion (87.1%), and Neuroticism (86.3%), while lower accuracies are obtained for Consciousness and Agreeableness, roughly around 80%. When controlled datasets are used for training, the average number of rows needed to determine the personality types is 12 taking around 60 s which supports the idea that the proposed approach is fast and can be an attractive alternative to the FMM questionnaire or psychological interviews commonly used for evaluating the FMM personality types.

4.1.3 Relationship between the handwriting features and FMM

We conduct the next experiment in order to see which handwriting feature is associated with each of the five personality traits in FMM. In order to accomplish this, we create a background application that checks the HM and counts each occurrence of all the handwriting feature classifications against each of the five personality traits. This is acquired with the system trained on controlled datasets for 127 subjects and tested on the random datasets for the remaining subject with n -fold cross-validation, averaging the results. The results obtained are highlighted in Table 4.

It can be observed that there are several links between the five personality types and the handwriting features, such that extreme left word slant, descending baseline, and cross-like lowercase letter "P" are associated with Conscientiousness, while medium connected strokes, medium right word slant, and balanced lowercase letter "P" are associated with Openness to Experience. These findings are significant as they can be used to optimize the proposed architecture such that the neural network is trained and tested only on the handwriting features that have relevant information about the personality traits that are investigated, the others being filtered out.

4.2 Comparison with state-of-the-art

As currently there is no standard public database that is broadly used for testing and comparing different architectures and methods for evaluating personality evaluation based on handwriting, we test the most common methods for assessing personality from handwriting on our database and compare the results with those obtained from our proposed approach. As it can be observed, our approach offers 84.4% accuracy for intra-subject tests and 80.5% accuracy for inter-subject tests, surpassing the rule-based classifier approach of Champa and AnandaKumar [4] with 12.5%, as well as the SVM, k-NN, and Ada-Boost combination of classifiers employed by Chen and Lin in [6], with 7.2%,

Table 3 Big Five personality prediction accuracies and average number of rows in inter-subject tests

Type of training samples	Type of test samples	Number of subjects involved in training	Recognition rate (%)—Openness to Experience	Recognition rate (%)—Conscientiousness	Recognition rate (%)—Extraversion	Recognition rate (%)—Agreeableness	Recognition rate (%)—Neuroticism	Recognition rate (%)—Overall (Undefined excluded)	Average % of cases where personality traits were Undefined	Average number of handwritten rows analyzed until the neural network converges to a result
C*	C*	64	86.3 ± 0.15	78.3 ± 0.2	84.4 ± 0.15	77.3 ± 0.2	83.4 ± 0.15	82.1 ± 0.15	0.6	12
C*	C*	96	87.2 ± 0.12	79.1 ± 0.15	85.4 ± 0.12	77.7 ± 0.15	84.4 ± 0.12	82.7 ± 0.12	0.4	11
C*	C*	127	88.6 ± 0.1	80.4 ± 0.12	87.1 ± 0.05	79.6 ± 0.12	86.3 ± 0.05	84.5 ± 0.1	0.3	10
R*	R*	64	74.4 ± 0.25	67.3 ± 0.3	73.3 ± 0.2	65.6 ± 0.3	72.3 ± 0.2	70.6 ± 0.2	9	20
R*	R*	96	76.5 ± 0.22	68.6 ± 0.3	75.4 ± 0.2	67.8 ± 0.25	75.5 ± 0.17	72.7 ± 0.17	5	18
R*	R*	127	78.6 ± 0.2	70.7 ± 0.22	77.4 ± 0.15	69.9 ± 0.2	76.7 ± 0.15	74.8 ± 0.17	3	16
C*	R*	127	86.7 ± 0.12	78.6 ± 0.17	85.6 ± 0.12	77.6 ± 0.15	84.4 ± 0.12	80.5 ± 0.1	0.6	12
R*	C*	127	80.7 ± 0.25	73.4 ± 0.25	79.7 ± 0.17	72.2 ± 0.25	79.1 ± 0.17	76.7 ± 0.17	3	16
C* + R*	C* + R*	127	86.65 ± 0.1	77.4 ± 0.1	84.2 ± 0.05	76.7 ± 0.1	83.1 ± 0.05	81.3 ± 0.07	0.4	11

C* controlled, R* random

Table 4 Correlation between the handwriting features and the Big Five personality types

Big Five personality type	Most present three handwriting features in HM
Neuroticism	<i>Baseline</i> : leveled, <i>Word Slant</i> : Moderate Right Slant, <i>Lowercase letter "f"</i> : angular point
Openness to Experience	<i>Lowercase letter "f"</i> : balanced, <i>Connecting Strokes</i> : Medium Connectivity, <i>Word Slant</i> : Medium Right Slant
Extraversion	<i>Connecting Strokes</i> : weak connectivity, <i>Baseline</i> : Ascending, <i>Word Slant</i> : Extreme Right Slant
Agreeableness	<i>Connecting Strokes</i> : strongly connected, <i>Word Slant</i> : Extreme Left Slant, <i>Lowercase letter "f"</i> : very low
Conscientiousness	<i>Word Slant</i> : Extreme Left Slant, <i>Baseline</i> : descending, <i>Lowercase letter "f"</i> : cross-like

respectively. Similarly, the proposed approach performs slightly better at the task of determining the FMM personality traits based on handwriting compared with the HMM-NN combination employed by Fallah and Khotanlou [5]. The results are detailed in Table 5.

5 Conclusions

We described the first non-invasive three-layer architecture in literature that aims to determine the Big Five personality type of individuals solely by analyzing their handwriting. This novel architecture has a base layer where the handwritten sample in the form of a scanned image is normalized, segmented in rows, words, and letters and based on the computed segments the handwriting features are determined; an intermediary layer where a Handwriting Map (HM) is computed by binary coding the handwriting feature type of each letter; and a top layer where a feed-forward neural network is trained via backpropagation to learn the patterns from the HM map and compute the FMM personality traits.

In order to train and test this novel architecture, due to lack of any database that would link the FMM personality traits with handwriting samples, we create the first such database containing the FMM personality traits of 128 subjects and six handwriting samples from each of them with both predefined text (referred to as

controlled dataset) as well as random text freely chosen by subjects (referred to as random dataset). We test our novel architecture on this database in both intra-subject and inter-subject methodologies and we obtain the highest prediction accuracies when the controlled dataset is used in the training stage, which shows that choosing a predefined text to be used for training the system is an important point in order to reach high accuracies, while testing can be done on random texts with no essential need for predefined texts to be used. This is an essential finding for real applications of such a systems, as it provides flexibility to the end-user, such that he/she will not have to write a predefined text every time, instead writing it only at the beginning in order to train the system, and then, to evaluate his/her personality traits at any given moment of time, he/she can use any random text he/she wants. In intra-subject tests, when the controlled dataset is used for training and random dataset for testing, we obtain an overall accuracy of 84.4%, while in inter-subject tests with a similar test-case we obtain an overall prediction accuracy of 80.5%. The highest prediction accuracies are obtained for Openness to Experience, Neuroticism, and Extraversion, reaching above 84%, while for Agreeableness and Conscientiousness we only obtained roughly around 77%. Overall, the prediction accuracy of the system is higher than that of any other state-of-the-art method tested on the same database. Another significant finding is that we determined several relationships between the prediction with high accuracy of specific FMM personality traits and the handwriting features analyzed which can be further exploited to improve the accuracy of the system. The accuracy of the system can also be further improved either by analyzing other handwriting features together with the seven ones already analyzed in our study or grouping these features based on the relevant information they offer in this task and filter out the irrelevant ones for each of the five personality traits. This will be the direction of our future research.

The proposed system computes the results in no more than 90 s which makes it faster than the current ways of determining personality traits through extensive self-report questionnaires, usually more cumbersome and time-consuming to fill in and involving more effort from both subject's and psychologist's side which will have to post-process the questionnaire results and evaluate the five personality traits; this shows that our current approach could be used as an attractive, faster, and easier to use alternative to these commonly used personality evaluation techniques.

Abbreviations

AARE: Average Absolute Relative Error; ANN: Artificial neural networks; ASD: Autism spectrum disorder; C: Controlled Dataset; FFM-NN: Five-Factor Model-Neural Network; FMM: Five-Factor Model; HM: Handwriting Map;

Table 5 Comparison with state-of-the-art

Work/Year	Method	Prediction accuracy
Champa and AnandaKumar [2010] [4]	Set of rule-based classifiers	68%
Chen and Lin [2017] [6]	Support Vector Machines, k-nearest neighbors, and AdaBoost classifiers	72.8%
Fallah and Khotanlou [2016] [5]	Hidden Markov Models and neural networks	78.9%
Current work [2017]	Feed-forward neural networks	84.4% (intra-subject)/ 80.5% (inter-subject)

HMM: Hidden Markov Model; K-NN: K-nearest neighbors; MBTI: Myers-Briggs Type Indicator; MMPI: Minnesota Multiphasic Personality Inventory; MNIST: Modified National Institute of Standards and Technology; NN: Neural network; PD: Parkinson's disease; R: Random Dataset; RAM: Random Access Memory; SVM: Support Vector Machines; SWT: Stroke Width Transform; VPP: Vertical Projection Profile

Availability of data and materials

Data is not shared publicly. Please contact the author for data requests.

Authors' contributions

MG contributed to the state-of-the-art research, implementation of the neural network-based testbed and methods employed in Scala using Spark library, testing the proposed architecture and discussion around the results and conclusions. NV contributed to the discussion around the results obtained and conclusions. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 20 April 2018 Accepted: 25 June 2018

Published online: 10 July 2018

References

- R Plamondon, Neuromuscular studies of handwriting generation and representation. *International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 261 (2010) Kolkata, November 2010
- Y Tang, X Wu, W Bu, Offline text-independent writer identification using stroke fragment and contour-based features. *2013 IEEE International Conference on Biometrics (ICB)*, 1–6 (June 2013)
- M Naghibolhosseini, F Bahrami, A behavioral model of writing. *International Conference on Electrical and Computer Engineering (ICECE)*, 970–973 (December 2008)
- HN Champa, KR Anandakumar, Automated human behavior prediction through handwriting analysis. *2010 First International Conference on Integrated Intelligent Computing (ICIIC)*, 160–165 (August 2010)
- B Fallah, H Khotanlou, in *Artificial Intelligence and Robotics (IRANOPEN)*. Identify human personality parameters based on handwriting using neural networks (April 2016)
- Z Chen, T Lin, Automatic personality identification using writing behaviors: an exploratory study. *Behav Inform Technol* **36**(8), 839–845 (2017)
- I Siddiqi, C Djeddi, A Raza, L Souici-Meslati, Automatic analysis of handwriting for gender classification. *Pattern. Anal. Applic.* **18**(4), 887–899 (November 2015)
- S Maadeed, A Hassaine, Automatic prediction of age, gender, and nationality in offline handwriting. *EURASIP Journal on Image and Video Processing* **2014**, 10 (December 2014)
- G Luria, S Rosenblum, A computerized multidimensional measurement of mental workload via handwriting analysis. *Behav. Res. Methods* **44**(2), 575–586 (June 2012)
- I Zaarour, L Heutte, P Leray, J Labiche, B Eter, D Mellier, Clustering and Bayesian network approaches for discovering handwriting strategies of primary school children. *Int. J. Pattern Recognit. Artif. Intell.* **18**(7), 1233–1251 (2004)
- R Sudirman, N Tabatabaey-Mashadi, I Ariffin, Aspects of a standardized automated system for screening children's handwriting. *First international conference on Informatics and Computational Intelligence (ICI)*, 48–54 (December 2011)
- J Gorbova, I Lusi, A Litvin, G Anbarjafari, Automated screening of job candidate based on multimodal video processing. *Computer Vision and Pattern Recognition Workshops (CVPRW) (2017) IEEE Conference on*, July 2017
- G Luria, A Kahana, S Rosenblum, Detection of deception via handwriting behaviors using a computerized tool: Toward an evaluation of malingering. *Cogn. Comput.* **6**(4), 849–855 (December 2014)
- TLP Tang, Detecting honest People's lies in handwriting. *J. Bus. Ethics* **106**(4), 389–400 (April 2012)
- SB Bhaskoro, SH Supangkat, An extraction of medical information based on human handwritings. *2014 International Conference on Information Technology Systems and Innovation (ICITSI)*, 253–258 (November 2014)
- Drotar, P., Mekyska, J., Smekal, Z., Rektorova, I., Prediction potential of different handwriting tasks for diagnosis of Parkinson's, *2013 E-Health and Bioengineering Conference*, Pages 1–4, November 2013.
- N Grace, PG Enticott, BP Johnson, NJ Rinehart, Do handwriting difficulties correlated with core symptomology, motor proficiency and attentional behaviors. *Journal of Autism and Developmental Disorders*, 1–12 (January 2017)
- WL Lee, K-C Fan, Document image preprocessing based on optimal Boolean filters. *Signal Process.* **80**(1), 45–55 (2000)
- JG Leu, Edge sharpening through ramp width reduction. *Image Vis. Comput.* **18**(6), 501–514 (2000)
- SCF Lin et al., Intensity and edge based adaptive unsharp masking filter for color image enhancement. *Optik - International Journal for Light and Electron Optics* **127**(1), 407–414 (2016)
- R Legault, CY Suen, Optimal local weighted averaging methods in contour smoothing. *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 690–706 (July 1997)
- Y Solihin, CG Leedham, Integral ratio: a new class of global thresholding techniques for handwriting images. *IEEE Trans. Pattern Anal. Mach. Intell.* **21**, 761–768 (Aug. 1999)
- Kai Chen, Fei Yin, Cheng-Lin Liu, Hybrid page segmentation with efficient whitespace rectangles extraction and grouping. *Document Analysis and Recognition (ICDAR) 2013 12th International Conference on*, pp. 958–962, 2013.
- V Papavassiliou, T Stafylakis, V Katsouro, G Carayannis, Handwritten document image segmentation into text lines and words. *Pattern Recogn.* **43**, 369–377 (2010)
- Costa, P.T. Jr., McCrae, R.R., Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual, Psychol. Assess. Resources, Odessa, FL, 1992.
- BW Roberts, D Mroczek, Personality trait change in adulthood. *Curr. Dir. Psychol. Sci.* **17**(1), 31–35 (2008)
- M Jokela, C Hakulinen, A Singh-Manoux, M Kivimaki, Personality change associated with chronic diseases: pooled analysis of four perspective cohort studies. *Psychol. Med.* **44**, 2629–2640 (2014)
- SJ Karau, RR Schmeck, AA Avdic, The big five personality traits, learning styles, and academic achievement. *Journal on Personality and Individual Differences* **51**(4), 472–477 (September 2011)
- Morris, R. N., *Forensic Handwriting Identification: Fundamental Concepts and Principles*, 2000.
- K Amend et al., *Handwriting Analysis: The Complete Basic Book* (Borgo Press, San Bernardino, California, 1981)
- MB Menhaj, F Razzazi, A new fuzzy character segmentation algorithm for Persian/Arabic typed texts. *International Conference on Computational Intelligence, Fuzzy Days 1999: Computational Intelligence*, 151–158 (1999)
- R Coll, A Fornes, J Lladós, Graphological analysis of handwritten text documents for human resources recruitment. *12th International Conference on Document Analysis and Recognition*, 1081–1085 (July 2009)
- EM Hicham, H Akram, S Khalid, Using features of local densities, statistics and HMM toolkit (HTK) for offline Arabic handwriting text recognition. *Journal of Electrical Systems and Information Technology* **4**(3), 387–396 (2017)
- B Epshtein, E Ofek, Y Wexler, Detecting text in natural scenes with stroke width transform. *Computer Vision and Pattern Recognition (CVPR)*, 2010 IEEE Conference on (August 2010)
- L Deng, The MNIST database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Process. Mag.* **29**(6), 141–142 (November 2012)
- L Xiaoyuang, Q Bin, W Lu, A new improved BP neural network algorithm. *Second International Conference on Intelligent Computation Technology and Automation*, 19–22 (October 2009)