U.P.B. Sci. Bull., Series C, Vol. 79, Iss. 4, 2017 ISSN 2286-3540

# 3-LAYER ARCHITECTURE FOR DETERMINING THE PERSONALITY TYPE FROM HANDWRITING ANALYSIS BY COMBINING NEURAL NETWORKS AND SUPPORT VECTOR MACHINES

Mihai GAVRILESCU[1]

*We propose a 3-layer architecture for determining the personality type of a subject by only analyzing handwriting. The proposed architecture combines Neural Network and Support Vector Machine approaches and it is tested in various configurations for determining which combination offers the best personality type classification results for each mixture of handwriting features. In order to test the system, we created a new training database based on Myers-Briggs Type Indicator (MBTI) questionnaire with the purpose of eliminating the inconsistencies of the experimental results compared to manual analysis. We present the architecture, the experimental results, as well as further improvements that could be brought to the current architecture.*

**Keywords:** neural networks, affective computing, personality recognition, bioinformatics

## 1. Introduction

Handwriting is one of the most important means of communication present in our lives for centuries. Although it was intensively used, only recently has it been correlated to the personality and emotional state of the writer and this is currently a disputed domain.

The current ways of analyzing handwriting are by means of a psychological analysis called graphology. Because it is though that the brain forms characters based on habits of the writer, it is considered that each neurological brain pattern forms a distinctive neuromuscular movement acting the same for individuals with the same type of personality and hence the writing of an individual is an accurate image of a person's brain [1]. Graphologists typically use different handwriting features in order to study the personality or emotional state of the writer, such features being: weight of the strokes [2], the way certain letters are written (letter "t" and letter "y" in [3]) as well as other patterns, such as, for example, the trajectory of the writing [4].

[1] PhD student, Department of Telecommunications, University POLITEHNICA Bucharest, Romania, e-mail: mike.gavrilescu@gmail.com

In terms of determining the personality of subjects, the current methods imply the use of specific questionnaires. However, the main disadvantage of a questionnaire is that it cannot be filled in too often and it can also be faked, being subjective, hence the need for a less-intrusive and more objective approach is needed, and this is the purpose of this paper, as we are trying to fill the gap between handwriting and personality types by building a system able to determine the personality type of a writer only by analyzing his writing.

## 2. Related Work

Because there is no standard in handwriting behavior prediction and graphology typically implies a subjective analysis done by specialized graphologists, researchers have tried to design automatic systems able to determine personality traits or emotional states from handwriting, as a way of standardizing the graphological analysis. There are various classification techniques used for determining the personality traits of the writer based on handwriting, but the most employed are the ones based on neural networks.

Researchers in [5], for example, present a system that acquires writings and drawings from pupils and by means of a Bayesian network-based model, it provides useful information for a child development psychologist to determine which strategy should be used in order to increase the performance of the child. On a similar note, a neural-network based system studying the behavior of children based on their handwriting is presented in [6], based on the judgment that infants are the best subjects to be used for such tests because they are not affected by cultural background and have a fast evolution of the cognition rate. The system showed over 78% accuracy in determining developmental disorders in children, results more than promising. Another system based on neural network classification of handwritten features in order to determine personality features (more specifically the active personality and the leadership abilities of subjects) is presented in [7], with the purpose of being used in the recruitment process. The handwriting features used are document layout, letter size, slant, line angles, and letter shape and the performance of such a system is also extremely promising. Multiple artificial neural networks (ANN) are used in [8] together with multi-structure algorithms in order to analyze handwriting samples and predict personality traits. The technique used is to divide an A4 paper in two areas: signature area (having 9 handwriting features, 5 of them being classified using ANNs and the others using multi-structure algorithms), and handwriting area (with 5 handwriting features classified using multi-structure algorithm as well as ANN for hill valley extraction based on baseline features). Researchers obtained accuracies ranging from 87% to 100%. As neural networks were shown to offer good classification accuracies only for some handwriting features, the motivation of the current work is to integrate neural networks with support vector machines and to determine what are the handwriting feature combinations that should be classified via neural networks and which should be classified by means of support vector machines (SVM) in order to achieve the highest accuracies in recognizing the personality traits of the writer. Moreover, we employ the Meyer-Briggs Type Indicator (MBTI) [9] as a standard in determining the personality type of the writer and we train the neural network in order to provide outputs in form of MBTI personality types. The motivation is therefore building a non-intrusive and practical way of determining personality types based solely on handwriting analysis in order to replace the MBTI questionnaires that are typically impractical to be filled in often enough, can be faked by the subjects taking the test and do not offer the results in a fast manner.

In the following chapters we will present the theoretical model and the architecture of our system, as well as the experimental results and conclusions drawn from them.

## 3. Theoretical Model

As previously mentioned, this research aims determining the MBTI personality types of writers only by analyzing the handwriting features using a combination of Neural Networks (NN) and Support Vector Machines (SVM).

MBTI refers to a psychometric questionnaire that is often used for measuring the personality traits of an individual, having as applications from career counseling and development, leadership training, to even recognizing personality shifts specific to personality disorders, such as schizotypal disorder [9]. As mentioned before, the MBTI personality types are typically determined by asking the subject to fill in a questionnaire, this having as disadvantage the fact that a questionnaire can sometimes be faked by subjects, it is cumbersome to be filled in as well as impractical if you are aiming for a real-time personality type monitoring of the subject. This is why we are aiming easier and faster ways to determine the personality type, in this case making use of the handwriting of the writer.

Typically, the MBTI is based on 4 categories [9], also called personality primitives:

- *Extraverted (E) vs. Introverted (I)*
- *Sensing (S) vs. Intuition (N)*
- *Thinking (T) vs. Feeling (F)*
- *Judging (J) vs. Perceiving (P)*

A subject's personality can be therefore defined as a combination of these personality primitives (e.g. INTJ refers to Introverted, Intuition, Thinking, and Judging).

In terms of handwriting, there are tens of handwriting features that can be used to analyze writing [10], but out of these in the current paper we will analyze only six of them which we considered as providing clues on the personality traits of an individual. Also, four of them (baseline, writing pressure, connecting strokes, and word slant) are the main handwriting features employed by graphologies in a graphological analysis [11]. We also limited the number of handwriting features employed to six in order to avoid overcomplicating the system and overfitting the neural network used. These six features are:

- ***Baseline*** (the line on which the writing flows): *Ascending* (associated with optimistic, happy persons), *Descending* (associated with pessimistic, over thinkers), *Leveled* (associated with self-control and reasoning) [10]. We considered that this correspondence could give clues on the personality aspect as well, such as helping differentiate between Introverted and Extraverted or Sensing and Intuition.
- ***Writing pressure*** (amount of pressure applied by the pen on the paper): *Heavy writer* (associated with emotional persons), *Medium writer* (associated with persons easily affected by trauma), *Light writer* (associated with persons easily coping with trauma) [10]. We considered that this link between writing pressure and the ability of the individual to get over traumas can help as well dichotomize between Thinking and Feeling, Judging and Perceiving as well as Sensing and Intuition.
- **Connecting strokes** (how letters are connected to form a word): *Non-connected* (associated with monotonous persons), *Medium connected* (associated with persons that like to change environments), *Connected* (associated with persons easily adaptable to change) [10]. We considered that the evaluation of a person’s ability to adapt to changing environments also gives information on the Extraverted vs. Introverted and Intuition vs. Sensing dichotomies.
- **Word slant** (inclination of the written words): *Vertical slant* (associated with persons who can control their emotions), *Moderate right slant* (associated with persons that easily exteriorize their emotions), *Extreme right slant* (associated with persons who lack self-control), *Moderate left slant* (associated with persons that find it hard to express emotions), *Extreme left slant* (associated with defensive persons suffering from self-rejection) [10]. As word slant gives information on the ability to exteriorize emotions, this can offer important information to discriminate between Extraverted and Introverted as well as Thinking and Feeling.
- **Lowercase letter “t”** (how the “t” bar is written on letter t): *Very low* (low self-esteem), *Very high* (high self-esteem) [10]. This low self-esteem assessment can offer information to discriminate between Extraverted and Introverted as well as Judging and Perceiving.

- **Lowercase letter "f"** (how the letter "f" is written): *Narrow upper loop* (associated with narrow minded people), *Angular loop* (strong reaction to obstacles), *Angular point* (associated with persons who are easily revolted), *Cross-like lowercase letter "f" (*associated with persons who have an increased level of concentration), *Balanced* (associated with persons having leadership abilities) [10]. This could provide information that can help distinguish between Judging and Perceiving as well as Thinking and Feeling.

Therefore, having both the MBTI questionnaire results as well as writing samples from each of the subjects, we will build a NN-SVM based system that will be trained on the handwriting samples in such way that the output will be the same as the one computed via the MBTI questionnaire, and the trained system should be able to determine the personality traits solely by analyzing handwriting.

## 4. Proposed Architecture

As we detailed in previous sections, we aim building a system able to determine the personality type of an individual based solely on his handwriting. To achieve this, we propose a 3-layer architecture that combines NN and SVM and analyzes what is the best combination that offers the highest accuracies. The 3 layers of this architecture are the following:

- 1st layer: determining handwriting features
- 2nd layer: determining the personality traits based on the handwriting features collected from the 1st layer, in parallel for the Neural Network block and the Support Vector Machines block. A selector is employed to determine which features are fetched to which block.
- 3rd layer: a k-nearest neighbor classifier used to determine the personality results based on the input received from the 2nd layer.

Each layer is detailed in the following sections and the overall architecture is displayed in Fig. 1.

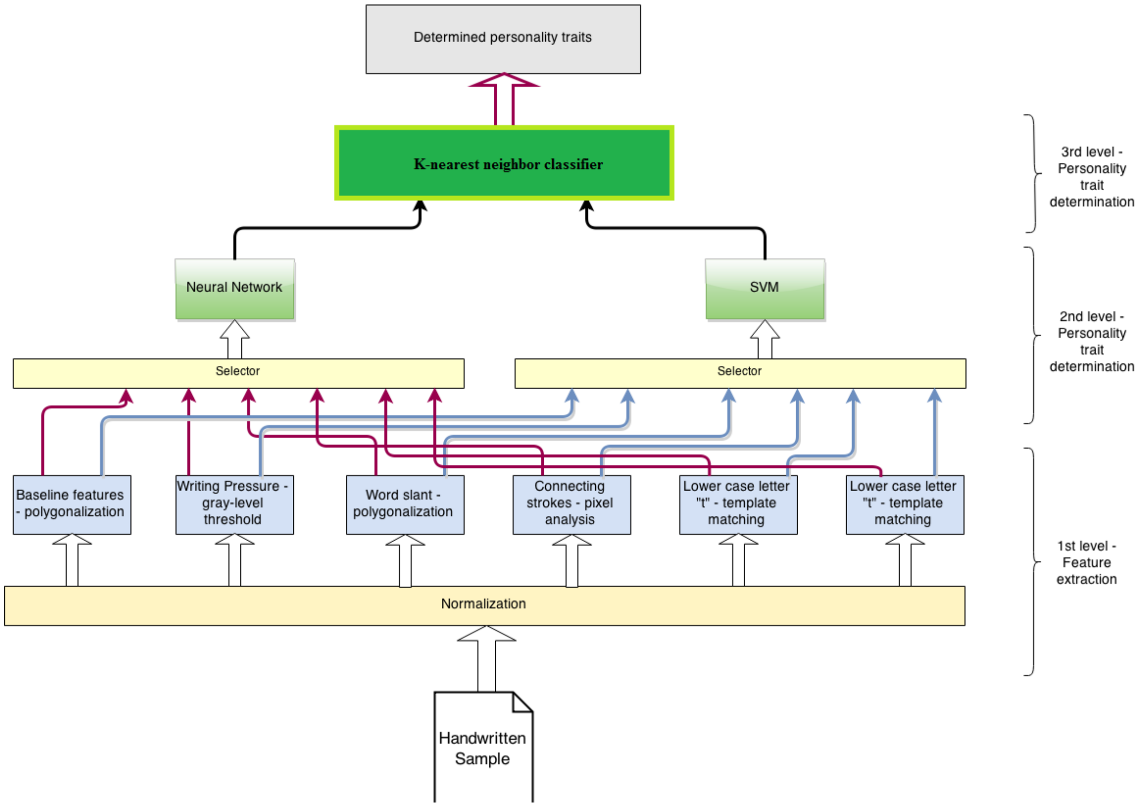



Fig. 1. 3-Layer architecture for personality type recognition from handwriting

### *A. The 1$^{st}$ layer*

The first layer has the main purpose of determining the handwriting features. In order to achieve this, the first step is to normalize the handwritten image, increase the contrast in order to make the characters more visible, isolate the regions containing handwriting, convert the image to a grey-scale image and remove any existing noise.

In order to isolate the handwriting text and reduce the noise, considering that the features are extracted directly from a greyscale image (as opposed to a binarized image) to reduce the information loss but which introduces additional noise, we made use of a two-dimensional Gabor filter to reduce it as it was proved robust to noise and insensitive to variations in line width [12]. The Gabor filter is applied at two different frequencies and four different orientations, the filter response is partitioned in 16x16 grids and the number of strong responses in each grid is concatenated in a 2048-dimensional vector. We conducted a part-of-word (PAW) classification experiment in order to evaluate the Gabor filter used and we obtained 99.8% accuracy at discriminating the handwritten text from the noise. It is important to mention that the writing was done on a clear white sheet of paper and it was previously evaluated by a human observer to ensure that the handwriting samples provided as input to the system don't have any visible noise.

After the handwriting text is delimited and the noise is removed, for some of the features the words must be split into letters. For this, each character is cropped; its edges are refined so that each character will be part of a bounding box. For this we used the Hough transform for character extraction as detailed in [13] employing the corresponding OpenCV function. The detected and processed characters are then fetched to the feature extraction step.

In the feature extraction step, for each handwriting characteristic that is being analyzed, the features are determined. As previously described, the handwriting features that are taken into consideration are: the baseline, the writing pressure, the word slant, the connecting strokes, the lowercase letter "t" and the lowercase letter "f". For baseline, the polygonalization algorithm [14] is employed. A polygon is determined that best delimits the writing and by studying the coordinates of the polygon compared to the overall page, the baseline is determined. The algorithm is depicted in Fig. 2 and was tested on all the handwriting samples from our own database comparing the results to those of a human observer, offering 99.1±0.3%. For writing pressure, the technique used is grey-level thresholding algorithm as detailed in [7] and the accuracy of the method was 98±0.5%. The connecting strokes are studied using a pixel rule-based algorithm [15] that studies the entire text and determines all the spaces between letters comparing the value with the total number of spaces that had to exist in the document if it was typewritten. The accuracy of the method is 98±0.5%. In terms of word slant, the same polygonalization technique employed for baseline is used with an accuracy of 98±0.4%.. In what it concerns the lowercase letter "t" and the lowercase letter "f", both were analyzed using template matching via the Hamming distance algorithm [16] and offered an average accuracy of 99±0.2% when tested on all the handwriting samples from our own database. All these handwriting features are subsequently fetched to the 2$^{nd}$ layer in order to determine the personality types. The algorithms used for the pre-processing steps as well as for feature extraction were the ones provided by OpenCV library which were adapted to the current handwriting recognition task.

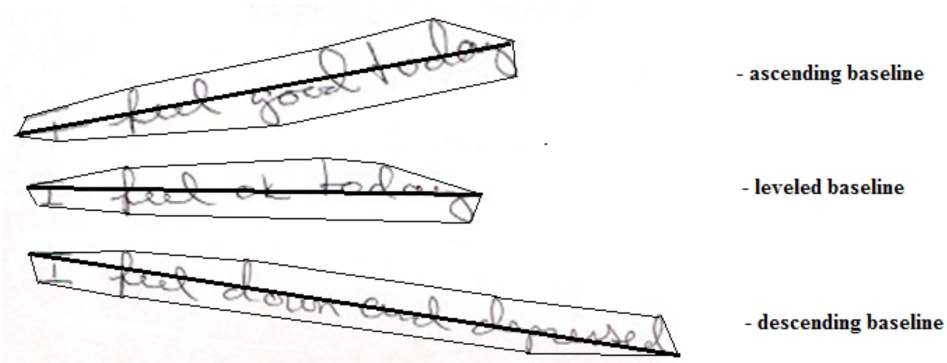



Fig. 2. Polygonization method for assessing the baseline type

### *B. The 2nd and 3rd layers*

As previously specified, the second layer will take the handwriting features from the first layer and will provide scores for each of the four personality primitive categories to the 3rd layer. The 2nd layer will therefore contain two parallel modules: a NN module and a SVM module, as detailed in Fig. 1. Because this is a pattern recognition task and our 3-layer architecture is a bottom-up one, for the neural network module we used a feed forward neural network, having 6 input nodes and 4 output nodes corresponding to the 4 MBTI dichotomies (Extraverted vs. Introverted, Sensing vs. Intuition, Thinking vs. Feeling, Judging vs. Perceiving). The Neural Network is trained using backpropagation comparing the results obtained when computing the handwriting features provided as inputs with the personality traits that were obtained via the MBTI questionnaire until the network is weighted and the error is minimized. We compute the Average Absolute Relative Error (AARE) in the training phase as the difference between what is expected and what is determined, and we tune the weights of the neural network until the output provides the best results. Through trial and error we determined that the optimal number of hidden nodes is 78 with an AARE of 0.003. We used gradient descent algorithm for learning the weights and biases of the NN and the Nguyen-Widrow weights initialization method for setting up the initial weights in the NN. The optimal learning rate obtained is 0.02, the optimal momentum 0.03 and the number of training epochs needed was 500. In order to introduce nonlinearity in the model as well as knowing that it determines a faster convergence in NNs trained with backpropagation, the activation function used is the log sigmoid.

For the **Support vector machine module** we used RBF kernel which showed to provide better results and less complexity than the polynomial kernel in several handwriting recognition tasks [17][18].

We made use of multiclass-SVM as there are many features that are received as inputs in the SVM classifier and we need to clearly separate them based on the personality primitives. The SVM multi-classification will be performed in the following steps:

- the multi-class SVM is trained with sample data in order to determine the feature space and to map them using RBF function;

- the multi-class SVM is then used for personality type prediction. The features coming from the 1st layer are mapped in the feature space using RBF kernel function and a division of the global hyperplane is done in order to separate the features one from another based on personality primitives.

The results from the NN and SVM modules are in a form of a percentage which is normalized in the [0,1] interval. These results are fetched to the 3rd layer containing a k-nearest neighbor classifier which takes the final decision regarding the personality type of the individual in the MBTI standard (e.g. INTJ). The k-nearest neighbor classifier optimal *k* value was determined to be 10. We have chosen the k-nearest neighbor because of its less complexity and fast calculation time, as well as due to the fact that we are faced with lower dimensionality (the number of input variables is low) and the data is already scaled in the [0,1] range [19]. The k-nearest neighbor classifier is trained using the 10-fold cross validation paradigm which is detailed in the next section.

In what it concerns the platform used for implementing this architecture; we used C++ programming with OpenCV library and the proposed system has a complexity of about 15.000 lines of code. Each of the feature detection blocks works in a special parallel thread. The selector is represented by a configuration file where we select which of the 1st layer features are taken into account in each of the two modules (SVM and NN) present in the 2nd layer. We use this selector because we want to test which combination of features are better treated with a neural network and which are providing better results as part of the SVM and hence determine the combination that offers the best accuracies. The SVM module as well as the NN module have a wait function in order for all the features to be calculated at 1st level before 2nd level process starts. In the same way the last k-nearest neighbor classifier has a wait function so that both the NN module and the SVM module provide their inputs. The neural networks, RBF-based SVM and the k-nearest neighbor classifier were implemented using the OpenCV library (for FNN we used *CvANN_MLP()* function, for SVM we used *CvSVM::RBF()* and for k-nearest neighbor classifier we used the *CvKNeaerst()* function) and adapting the out-of-the-box functions to the current handwriting recognition task by modifying the input .xmls.

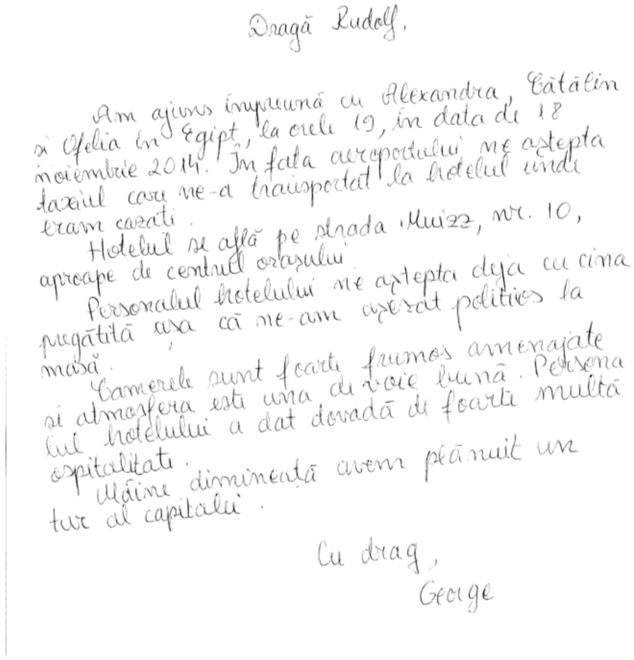
C. ***Training database and handwriting samples***

In terms of **training database**, we asked 64 subjects, according to the Helsinki Ethical Declaration, to take the MBTI questionnaire every 2 weeks for 2 months as well as provide a sample of 300 words containing the handwriting features that we are analyzing. The 64 subjects were chosen in order to be in accordance with the global statistics related to MBTI personality types, as follows: ISTJ - 11.6% - 7 subjects, ISFJ 13.8% - 8 subjects, INFJ 1.5% - 2 subjects, INTJ - 2.1% - 2 subjects, ISTP - 5.4% - 3 subjects, ISFP - 8.8% - 5 subjects, INFP - 4.4% - 3 subjects, INTP - 3.3% - 3 subjects, ESTP - 4.3% - 3 subjects, ESFP - 8.5% - 5 subjects, ENFP - 8.1% - 5 subjects, ENTP - 3.2% - 3 subjects, ESTJ - 8.7% - 5 subjects, ESFJ - 12.3 – 7 subjects, ENFJ - 2.8% - 2 subjects, ENTJ - 1.8% - 2 subjects.

In what it concerns the **handwriting samples**, we designed 3 letters of 100 words in Romanian language that were inspired from *The London letter*, a standard request exemplar used by graphologists in handwriting analysis [10], in such manner that they offer all the handwriting features we need. A sample of a handwritten letter is presented in Fig. 2.

In conceiving the handwriting text samples, we took into consideration the following points:

- **letter "t"** position in words: beginning (e.g. "**t**ransportat", "**t**axiul"), middle (e.g. "s**t**rada", "aş**t**epta"), end (e.g. "plănui**t**", "transporta**t**");
- **letter "f"** position in words: beginning (e.g. "**f**oarte", "**f**rumos"), middle (e.g. "atmos**f**era", "a**f**lă"), end (e.g. "Rudol**f**");
- **for connecting strokes** we made sure the sample tests contain the following cases that usually add difficulties in connecting letters when writing a word: words starting with uppercase (e.g. "**A**lexandra", "**E**gipt"), intercalating numbers (e.g. "în data de **18** noiembrie **2014**"), intercalating numbers and punctuation (e.g. "strada Muizz**,** nr**. 10**"), group of long words (e.g. "personalul hotelului", "multă ospitalitate"), use of letters that need an additional stroke - such as x, z, i or j - (e.g. "ta**xi**ul", "ca**z**aţ**i**", "a**j**uns"), words containing doubled letters (e.g. "Mui**zz**").



Dragă Rudolf,

Am ajuns împreună cu Alexandra, Cătălin și Ofelia în Egipt, la orele 19, în data de 18 noiembrie 2014. În fața aeroportului ne aștepta taxiul care ne-a transportat la hotelul unde eram cazați.

Hotelul se află pe strada Muizz, nr. 10, aproape de centrul orașului.

Personalul hotelului ne aștepta deja cu cina pregătită așa că ne-am așezat politicos la masă.

Camerele sunt foarte frumos amenajate și atmosfera este una de voie bună. Personalul hotelului a dat dovadă de foarte multă ospitalitate.

Mâine dimineață avem plănuit un tur al capitalei.

Cu drag,
George

Fig. 2. Example of a handwriting sample text

## 5. Experimental results

As stated in the previous chapter, because of the lack of a standard database containing handwritten samples and their MBTI results, we created a training database with samples collected from 64 subjects that took the MBTI questionnaire every two weeks for two months and also provided their handwriting samples for the three letters detailed in the previous chapter. As the number of samples is low, we use the 10-fold cross validation as it is known as the best validation technique when the data available is not sufficient for dividing it into training and test sets without losing modelling or testing abilities. The system was tested in all possible combinations, meaning that we combined SVM and NN selectively for each combination of handwriting features in order to determine which combination offers the best accuracy. We evaluate all these combinations by assessing the accuracy and the standard deviation. Accuracy is calculated as the proportion of true positive and true negative results from the entire number of results. Standard deviation quantifies the amount of variation of the results and determines how much results differ one from another.

The results are shown in *Table 1*.

*Table 1*

**Accuracy and standard deviation per different configurations**

(1) – BASELINE; (2) – WORD SLANT; (3) – WRITING PRESSURE; (4) – CONNECTING STROKES; (5) – LOWER CASE LETTER "T"; (6) – LOWER CASE LETTER "F"

| (1) | (2) | (3) | (4) | (5) | (6) | Accuracy (%) | Standard deviation (%) |
|---|---|---|---|---|---|---|---|
| NN | NN | NN | NN | NN | NN | 85.51 | ±0.31 |
| NN | NN | NN | NN | NN | SVM | 85.62 | ±0.23 |
| NN | NN | NN | NN | SVM | NN | 85.41 | ±0.33 |
| NN | NN | NN | SVM | NN | NN | 85.34 | ±0.42 |
| NN | NN | SVM | NN | NN | NN | 84.85 | ±0.31 |
| NN | SVM | NN | NN | NN | NN | 84.36 | ±0.35 |
| SVM | NN | NN | NN | NN | NN | 84.14 | ±0.51 |
| NN | NN | NN | NN | SVM | SVM | 85.23 | ±0.33 |
| NN | NN | NN | SVM | NN | SVM | 85.32 | ±0.42 |
| NN | NN | SVM | NN | NN | SVM | 85.41 | ±0.31 |
| NN | SVM | NN | NN | NN | SVM | 84.23 | ±0.32 |
| SVM | NN | NN | NN | NN | SVM | 84.24 | ±0.33 |
| NN | NN | NN | SVM | SVM | NN | 86.25 | ±0.31 |
| NN | NN | SVM | NN | SVM | NN | 86.23 | ±0.21 |
| NN | SVM | NN | NN | SVM | NN | 84.51 | ±0.43 |
| SVM | NN | NN | NN | SVM | NN | 83.83 | ±0.45 |
| **NN** | **NN** | **SVM** | **SVM** | **NN** | **NN** | **89.21** | ±0.12 |
| NN | SVM | NN | SVM | NN | NN | 84.13 | ±0.41 |
| SVM | NN | NN | SVM | NN | NN | 83.17 | ±0.51 |
| NN | SVM | SVM | NN | NN | NN | 82.85 | ±0.54 |
| SVM | NN | SVM | NN | NN | NN | 82.43 | ±0.55 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| SVM | VM | N | N | N | N | 79.41 | ±0.78 |
| NN | N | N | VM | VM | VM | 87.02 | ±0.2 |
| NN | N | VM | N | VM | VM | 87.43 | ±0.18 |
| NN | VM | N | N | VM | VM | 84.21 | ±0.45 |
| SVM | N | N | N | VM | VM | 83.93 | ±0.43 |
| SVM | N | N | VM | N | VM | 83.84 | ±0.43 |
| SVM | N | VM | N | N | VM | 84.13 | ±0.4 |
| SVM | VM | N | N | N | VM | 83.14 | ±0.48 |
| SVM | N | N | VM | VM | N | 84.26 | ±0.43 |
| SVM | N | VM | N | VM | N | 84.14 | ±0.34 |
| SVM | VM | N | N | VM | N | 83.23 | ±0.51 |
| SVM | N | VM | VM | N | N | 84.11 | ±0.32 |
| SVM | VM | N | VM | N | N | 83.03 | ±0.33 |
| SVM | VM | VM | N | N | N | 82.74 | ±0.34 |
| NN | N | VM | VM | VM | VM | 84.03 | ±0.35 |
| NN | VM | N | VM | VM | VM | 83.91 | ±0.43 |
| SVM | N | N | VM | VM | VM | 83.12 | ±0.51 |
| NN | VM | VM | N | VM | VM | 85.14 | ±0.34 |
| SVM | N | VM | N | VM | VM | 84.34 | ±0.41 |
| NN | VM | VM | VM | N | VM | 84.22 | ±0.42 |
| SVM | N | VM | VM | N | VM | 84.11 | ±0.41 |
| SVM | VM | N | VM | N | VM | 83.23 | ±0.44 |
| SVM | VM | VM | N | N | VM | 83.14 | ±0.46 |
| NN | VM | VM | VM | VM | N | 85.32 | ±0.32 |

| (1) | (2) | (3) | (4) | (5) | (6) | Accuracy (%) | Standard Deviation (%) |
|---|---|---|---|---|---|---|---|
| SVM | N | VM | VM | VM | N | 84.21 | ±0.42 |
| SVM | VM | N | VM | VM | N | 84.13 | ±0.41 |
| SVM | VM | VM | N | VM | N | 83.25 | ±0.5 |
| SVM | VM | VM | VM | N | N | 83.53 | ±0.48 |
| NN | VM | VM | VM | VM | VM | 85.82 | ±0.28 |
| SVM | N | VM | VM | VM | VM | 85.21 | ±0.3 |
| SVM | VM | N | VM | VM | VM | 84.23 | ±0.32 |
| SVM | VM | VM | N | VM | VM | 84.35 | ±0.31 |
| SVM | VM | VM | VM | N | VM | 84.83 | ±0.31 |
| SVM | VM | VM | VM | VM | N | 85.01 | ±0.25 |
| SVM | VM | VM | VM | VM | VM | 85.23 | ±0.24 |
| SVM + NN | VM + NN | VM + NN | VM + NN | VM + NN | VM + NN | 86.22 | ±0.32 |

We can observe that the best accuracy is not obtained when both the SVM and NN are used on all the handwritten features, but the system provides better results if NN is taking care of baseline, word slant and the lower case letters “t” and “f”, while the SVM is processing the pen pressure and connecting strokes obtaining an accuracy of 89.2% with the lowest standard deviation of 0.12%, compared to 86.2% accuracy obtained assessing all features with both NN and SVM modules. This shows that the two approaches can complement each other, NN being more suitable for handwriting features determined via polygonalization and template matching, while SVM offers better results for handwriting features determined via grey-level threshold classifiers.

Moreover we kept the configuration of *NN,NN,SVM,SVM,NN,NN* that showed to offer the best accuracy and we proceeded with determining the false positive and false negative rates for each personality primitive using the same 10-fold cross validation method. False positive rate is calculated as the number of false positives divided by the sum of false positives and true negatives. The false negative rate is calculated as the number of false negatives divided by the sum of true positives and false negatives. The results are detailed in *Table 2*.

*Table 2*

**False positive and false negative rates**

| Primitive | False positive rate (%) | False negative rate (%) |
|---|---|---|
| Extraverted | 1.1% | 1.2% |
| Introverted | 1.3% | 1.7% |
| Sensing | 3.6% | 4.1% |
| Intuition | 5.3% | 5.7% |
| Thinking | 1.5% | 2.6% |
| Feeling | 1.3% | 2.1% |
| Judging | 2.4% | 2.2% |
| Perceiving | 1.8% | 1.9% |

Analyzing the results, we can see that the highest error rates for both false positive and false negative are for Intuition vs. Sensing, which can be explained by the fact that the training database, built to simulate the global statistics, only has 21 subjects for Intuition, compared to 43 subjects for Sensing. The low false positive and false negative rates obtained for Extraverted vs. Introverted, Thinking vs. Feeling as well as Judging vs. Perceiving are extremely promising showing that this architecture can successfully be used for determining personality types only based on handwriting. For Introverted vs. Sensing we might consider increasing our training database in order to improve the results obtained for these personality primitives. With this in mind, we trained the system on 32 subjects and tested it on the other 32 and compared it with the results obtained when the system was trained on 48 subjects and tested on the remaining 16 subjects in order to confirm if increasing the number of subjects involved in training can increase significantly the false positive and false negative rates. Results are detailed in Table 3 and it can be observed that as the number of subjects involved in training increases, the false positive and false negative rates decrease, hence the accuracy of the system improves, even for Intuition vs. Sensing, which is an indication that increasing the number of subjects involved in training can also improve these two primitives even more.

*Table 3*

**False positive and false negative rates**

| Primitive | False positive rate (%) / trained on 48 subjects | False negative rate (%) / trained on 48 subjects | False positive rate (%) / trained on 32 subjects | False negative rate (%) / trained on 32 subjects |
|---|---|---|---|---|
| Extraverted | 1.1% | 1.3% | 2.3% | 2.4% |
| Introverted | 1.3% | 1.7% | 3.1% | 3.3% |
| Sensing | 3.7% | 4.1% | 5.5% | 7.5% |
| Intuition | 5.2% | 5.6% | 7.6% | 7.7% |
| Thinking | 1.6% | 2.6% | 3.3% | 4.3% |
| Feeling | 1.3% | 2.1% | 2.4% | 3.3% |

| | | | | |
|---|---|---|---|---|
| Judging | 2.5% | 2.1% | 3.4% | 3.4% |
| Perceiving | 1.8% | 1.9% | 2.6% | 2.4% |

In terms of performance, the personality type is computed in an average of 60 seconds when analyzing 100 words handwritten letter, showing that such a system is faster than the standard MBTI questionnaire and it is also more practical.

## 6. Conclusions

We proposed a three-layered architecture for determining the MBTI personality type of subjects only by analyzing their handwriting. The architecture is combining Neural Networks and Support Vector Machines and is trained on a database containing handwriting samples as well as MBTI questionnaire results from 64 subjects chosen in accordance with the global statistics regarding personality types. The system takes into consideration the following handwriting features: baseline, writing pressure, the connecting strokes, the word slant, the lower case letter “f” and the lower case letter “t”. We tested the system on various configuration and we showed that the best results are obtained when Neural Networks are used for classifying the combination of handwriting features related to baseline, word slant and lower case letters “t” and “f”, while SVM is used for connecting strokes and writing pressure. In this case the personality type accuracy was about 88.6%. We also determined the false positive and false negative rates for each MBTI personality primitives and showed that, as the number of subjects involved in training increases, the false positive and false negative rates are smaller, hence the accuracy is better. We obtained low false positive and false negative rates for Extraverted vs. Introverted, Thinking vs. Feeling and Judging vs. Perceiving, while for Intuition vs. Sensing the results were not satisfactory enough. This proves that we need to increase the number of subjects involved in training as well as explore other classification techniques for Intuition vs. Sensing cases, and this will be the subject of our next studies.

## R E F E R E N C E S

[1] *R. Plamondon*, "Neuromuscular Studies of Handwriting Generation and Representation", International Conference on Frontiers in Handwriting Recognition (ICFHR*)*, pp. 261, Kolkata, India, Nov. 2010.

[2] *Y. Tang, X. Wu, W. Bu*, "Offline text-independent writer identification using stroke fragment and contour based features", IEEE International Conference on Biometrics (ICB), pp. 1-6, Madrid, Spain, June 2013.

[3] *H. N. Champa, K. R. Anandakumar*, "Automated Human Behavior Prediction through Handwriting Analysis", 1st International Conference on Integrated Intelligent Computing (ICIIC), pp. 160-165, Bangalore, India, Aug. 2010.

[4] *M. Naghibolhosseini, F. Bahrami*, "A behavioral model of writing", International Conference on Electrical and Computer engineering (ICECE), pp. 970-973, Dhaka, Bangladesh, Dec. 2008.

[5] *I. Zaarour, L. Heutte, Ph. Leray, J. Labiche, B. Eter, D. Mellier*, "Clustering and Bayesian Network Approaches for Discovering Handwriting Strategies of Primary School Children", International Journal of Pattern Recognition and Artificial Intelligence, **vol. 18**, no. 7, pp. 1233-1251, Nov. 2004.

[6] *R. Sudirman, N. Tabatabaey-Mashadi, L. Ariffin*, "Aspects of a Standardized Automated System for screening children's handwriting", 1st international conference on Informatics and Computational Intelligence (ICI), pp. 48-54, Bandung, Indonesia, Dec. 2011.

[7] *R. Coll, A. Fornes, J. Llados*, "Graphological Analysis of Handwritten Text Documents for Human Resources Recruitment," 10th International Conference on Document Analysis and Recognition, pp. 1081-1085, Barcelona, Spain, July 2009.

[8] *E. C. Djamal; R. Darmawati; S.N. Ramdlan*, "Application image processing to predict personality based on structure of handwriting and signature," International Conference on Computer, Control, Informatics and Its Applications (IC3INA), pp. 163-168, Jakarta, Indonesia, Nov 2013.

[9] *I. B. Myers*, "Gifts Differing: Understanding Personality Type", Davies-Black Publishing, Reprint edition, May 1st, pp. 33-82, 1995, ISBN 0-89106-074-X.

[10] *R. N. Morris*, "Forensic Handwriting Identification: Fundamental Concepts and Principles", Academic Press, 2000.

[11] *K. M. Koppenhaver*, "Forensic Document Examination: Principles and Practice", Humana Press, November 2010.

[12] *Y. Hamamoto, S. Uchimura, K. Masamizu, S. Tomita,* "Recognition of handprinted Chinese characters using Gabor features", Proceedings of the Third International Conference on Document Analysis and Recognition, August 1997, Montreal, Canada.

[13] *S. Saha, S. Basu, M. Nasipuri, d. K. Basu,* "A Hough Transform based Technique for Text Segmentation", Journal of Computing, vol. 2, issue 2, February 2010.

[14] *E. C. Djamal, S. N. Ramdlan, J. Saputra*, "Recognition of Handwriting Based on Signature and Digit of Character Using Multiple of Artificial Neural Networks in Personality Identification", Information Systems International Conference (ISICO), pp. 410-415, Bali, Indonesia, Dec. 2013.

[15] *W. Huang, G. Rong, Z. Bian,* "Strokes recovering from static handwriting," Proceedings of the 3rd International Conference on Document Analysis and Recognition, **vol. 2**, pp. 861-864, Montreal, Canada, Aug. 1995.

[16] *H. Tian, H. Shen*, "Hamming Distance and Hop Count Based Classification for Multicast Network Topology Inference", Proceedings of the 19th International Conference on Advanced Information Networking and Applications (AINA), **vol. 1**, pp. 267-272, Tamkang, Taiwan, Mar. 2005.

[17] *S. Ramya, K. Shama*, "Comparison of SVM Kernel Effect on Online Handwriting Recognition: A Case Study with Kannada Script", Data Engineering andIntelligent Computing, pp. 75-82, June 2017.

[18] *A. Choudhary*, *R. Rishi*, "A Fused Feature Extraction Approach to OCR: MLP vs. RBF", Advances in Intelligent Systems and Computing, Springer, **vol. 248**, pp. 159-165, 2014.

[19] *N. S. Altman*, "An introduction to kernel and nearest-neighbor nonparametric regression", The American Statistician, **vol. 46**, issue 3: pp. 175–185, 1992.