UNIVERSITY „POLITEHNICA" BUCHAREST

ETTI-B DOCTORAL SCHOOL

Ph.D. THESIS SUMMARY

MULTIMODAL ARCHITECTURE FOR PREDICTING HUMAN AFFECTIVE PROCESSES

Ph.D. Student:
Ing. Mihai Gavrilescu

Scientific coordinator:
Prof. Dr. Ing. Nicolae Vizireanu

BUCHAREST 2019

Table of contents

CHAPTER 1 – INTRODUCTION

Given the significant impact of emotions in essential cognitive processes and human interactions, as well as the importance of personality and temperament for understanding humans, for a computer to interact efficiently, naturally and in an intelligent manner with humans, it is vital for it to be able to recognize and express affective processes. These processes have a crucial role in understanding more complex phenomena, such as *attention*, *memory* or *esthetics*, and have applications in a broad spectrum of areas, from education [1], communication [2], entertainment [3], and design [4] to medical science [5] or improving human-computer interactions [6]. Considering these research needs, a new branch emerged called Affective Computing (AC), an interdisciplinary research area that spans from information technology and computer science to psychology and cognitive sciences. It is evident that, given the technological progress, this domain will be more and more researched. AC has already started to become an active area, examining ways in which devices for predicting emotions or other psychological traits can be developed as well as how these devices should respond to simulate *empathy*, but also a so-called *affective mediation*, a group of techniques used to facilitate the communication between humans who reveal their emotions [7]. These research paths which were developed mostly in the last decade are joining other research in assistive technology to improve the neglected area of human-computer and computer-mediated communication [7]. In this thesis, we propose non-invasive architectures designed to predict emotions, personality, temperament, and emotional states. The thesis starts with a unimodal methodology by analyzing separately three types of features (face, speech, and handwriting), culminating with a multimodal analysis in which the optimal fusion methods are determined for each analyzed affective process to ensure the highest prediction accuracy.

CHAPTER 2 – PSYCHOLOGICAL TOOLS USED

Although there is a large number of ways in which human affective processes can be defined and evaluated, in this thesis, we focus on the most effective and with proved practical use. For **defining emotions**, we start with the study conducted by Clynes [8] who emphasizes the idea of pure basic sentic states and we rely on Paul Ekman's research [9] who determines six basic emotions which we use in this study as well: *anger, surprise, fear, happiness, sadness*, and *disgust*. We use the Facial Action Coding System (FACS) [9] to analyze microexpressions translated in sets of Action Units (AU), having the main advantage of determining real emotions, even in cases when the subject tries to simulate other emotions [9]. In the latest FACS revision [10], the face is divided into 46 AUs and six intensity levels for each AU [10]. For **defining personality,** we use the Five-Factor Model (FFM) [11], analyzing five fundamental personality dimensions, as follows [11]: *Openness (O), Consciousness (C), Extroversion (E), Agreeableness (A),* and *Neuroticism (N).* FFM is successfully used in a large number of applications, from personal and professional development to predicting personality disorders, predicting substance use [12] or several physical diseases which have as symptoms different FFM dimensions patterns [13]. For **defining temperament**, we use the Fundamental Interpersonal Relations Orientation-Behavior (FIRO-B) [14] based on three fundamental interpersonal needs that determine a person to act or not when they are part of a group: *Openness, Control,* and *Inclusion* [14]. These three fundamental needs are evaluated through a specific questionnaire, on a scale from 0 to 9, on two coordinates: *Expressed* and *Wanted*

[14]. The results are mapped into five fundamental temperaments: melancholic, choleric, sanguine, supine, and phlegmatic. FIRO-B is primarily used for efficient team building, career counseling, and professional development [15]. For **defining emotional states**, we use the Depression Anxiety Stress Scale (DASS) [16]. DASS proposes a Self-Analysis Questionnaire (SAQ), which allows for the assessment of 14 elements associated with the three emotional states [16]. The result obtained from completing the questionnaire depicts the severity level of the emotional state, from *Normal* to *Extreme Severe*. DASS is successfully used for determining the severity of depression, anxiety, and stress [16], as well as for employee psychological evaluation [16], adaptive learning [17] or diagnosing physical diseases with one or more of these acute emotional states as symptoms [18].

CHAPTER 3 – DATABASES AND GENERAL ARCHITECTURAL DETAILS

We want to compare the architectures proposed in this thesis with other methods in the literature. Therefore, we use a set of public databases which are evaluated in similar studies. These are mentioned in the below table.

Affective process	Facial analysis	Speech analysis	Handwriting analysis	Multimodal analysis
Emotion	CK [19], CK+ [20], JAFFE [21], MMI [22], RU-FACS [23], Belfast [24], AFEW [25]	AVIC [28], eNTERFACE [29], McGilloway [30], structured Belfast [24], naturalistic Belfast [24], SALAS [30]	-	eNTERFACE [29], AVIC [29], naturalistic Belfast [24], SALAS [30]
Personality	-	SSPNet SPC [31]	-	-
Temperament	-	SSPNet SPC [31]	-	-
Emotional state	ANUStressDB [26] (stress), AVEC2014 [27] (depression)	SUSAS [32] (stress), AVEC2014 [27] (depression), France et al. [34] (depression)	EMOTHAW [33]	AVEC 2014 [27] (depression)

Because we need a database that correlates all three types features (face, speech, and handwriting) with all four affective processes and there are no state-of-the-art databases of this kind, we created our own database called *MENTAL* (**M**ultimodal - **E**motion perso**N**ality Temperament - **A**ffective state **L**abeling), by involving 128 Caucasian subjects (64 males, 64 females), with ages between 18 and 35 years, chosen in order to keep a homogenous distribution of the five temperaments as well as to display all DASS levels. We collect handwritten samples and audio-video frontal face recordings in scenarios where no emotion is induced, as well as when each of the six basic emotions is induced. In each of these sessions, the results obtained after filling in the four questionnaires (Discrete Emotions Questionnaire - DEQ, FFM, FIRO-B, and SAQ) is also collected. The questionnaire results are evaluated by trained psychologists to exclude the samples which are not correctly completed. For inducing the six basic emotions, we use videos from LIRIS-ACCEDE database [35]. MENTAL is divided in a *controlled dataset* (DSC) where samples are acquired when emotion is induced or the handwritten text is predefined, and a *random dataset* (DSR) where samples are acquired when no emotion is induced or the handwritten text is freely chosen by the subject.

Although there's a broad range of neural networks that can be used, in this thesis, we evaluate exclusively feed-forward neural networks (FFNN). For the majority of FFNN-based systems, we don't need more than two hidden layers [36]. Therefore we test the architectures using only FFNNs with one or two hidden layers. As activation functions, we evaluate *tanh*, *sigmoid*, *ReLU*, and *softmax*. As a training method, we use backpropagation [37], and we use gradient descent for optimizing the weights and biases and minimize the Average Absolute Relative Error (AARE) to a value lower than 0.02. We use the Nguyen-Widrow initialization method [38] to initialize the weights in the proposed FFNNs. In order to create homogenous unimodal models

that can be easily fused, all proposed architectures have a three-layer structure unique in literature: *a base layer* where the features are extracted and classified, *a middle layer* where a features' matrix is built based on the classifications provided by the base layer, and *a top layer* where an FFNN-based structure analyses the matrix built in the middle layer in a pattern recognition task and predicts the affective process. Because we use FFNNs, all proposed architectures have two phases: training and testing. In **training phase,** samples are provided to the base layer where they are normalized, and a set of filters and transformations are applied, depending on the type of analysis (face, speech or handwriting), classifying the analyzed features. The classification results are sent to the intermediary layer, which builds the features' matrix. In the top layer, an FFNN-based structure uses the features' matrix for training in order to offer the same results as the ones obtained based on the questionnaire. The training finishes when the AARE is minimized to a value lower than 0.02 or when the training samples are exhausted. The **testing phase** follows similar steps. In the base layer, the samples are normalized, and specific transformations and filters are applied. Then the analyzed features are classified, and the classification result is sent to the middle layer where the features' matrix is built. A previously trained FFNN-based structure from the top layer analyzes the matrix in a pattern recognition task and offers a result in [0;1] interval representing the probability of occurrence for the evaluated affective process. We use a Rule-based classifier (RBC) to compute the final result as follows: if each FFNN outputs the same result for five consecutive seconds, the architecture outputs that result as the final result; otherwise, the RBC marks the final result as *Undefined.*

CHAPTER 4 – FACE FEATURES ANALYSIS

For face features analysis, we use the dynamic approach and the FACS model. From the 46 AUs from the latest FACS revision, we analyze only *AU1, AU2, AU4-AU7, AU43,* and *AU45* from the upper part of the face and *AU8-AU20, AU22- AU28* from the lower part of the face. To enrich the emotional information from the cheeks area, we also analyze *AU33*, *AU34*, and *AU35*. The architecture is depicted in Fig. 4.1. Initially, we establish a statistical model for the skin color using a method similar to the one presented in [38], classifying pixels in facial pixels and non-facial pixels which are used for face detection [39]. We apply filters for noise removal [40] and morphological operations (erosion, dilatation, and filling gaps) [41], obtaining candidate faces. After determining the candidate faces, we evaluate the position of eyes and mouth [42] through a face scaling method and by calculating the interocular distance. Based on the position of eyes and mouth, the face is divided into multiple blocks. We then extract the facial features using Active Shape Models (AAM) [43]. We use Principal Component Analysis (PCA) to produce a parameterized model that describes the faces used in the training phase and estimates new faces [44]. Learning the correlations between the model's parameters and the candidate face is done using steepest descent (SD) and Jacobian matrices [43][45]. To identify the facial features, we delimit equally sized strips centered in a predefined set of facial landmarks for each facial feature [46]. We use Stochastic Gradient Descent (SGD), and we modify the gradient in the opposite direction until we reach a minimum, obtaining, therefore, a set of non-rigid parameters. We use Support Vector Machines (SVM) classifiers with six classes [47], receiving the non-rigid parameters as input and being trained to determine the intensity of each of the 31 analyzed AUs. Evaluating the algorithm in cross-database tests using CK+ [20], MMI [21], JAFFE [22], Belfast [24], AFEW [25], and RU-FACS [23], the proposed method classifies all AUs with over 90% accuracy. For each frame from the video sample, each AU is classified with an intensity level

from O to E, then normalized in [0;1] interval. Based on these classifications, in the middle layer, we build a Face Matrix (FM), each line *n* in FM containing the classification results for all AUs evaluated in frame *n.*

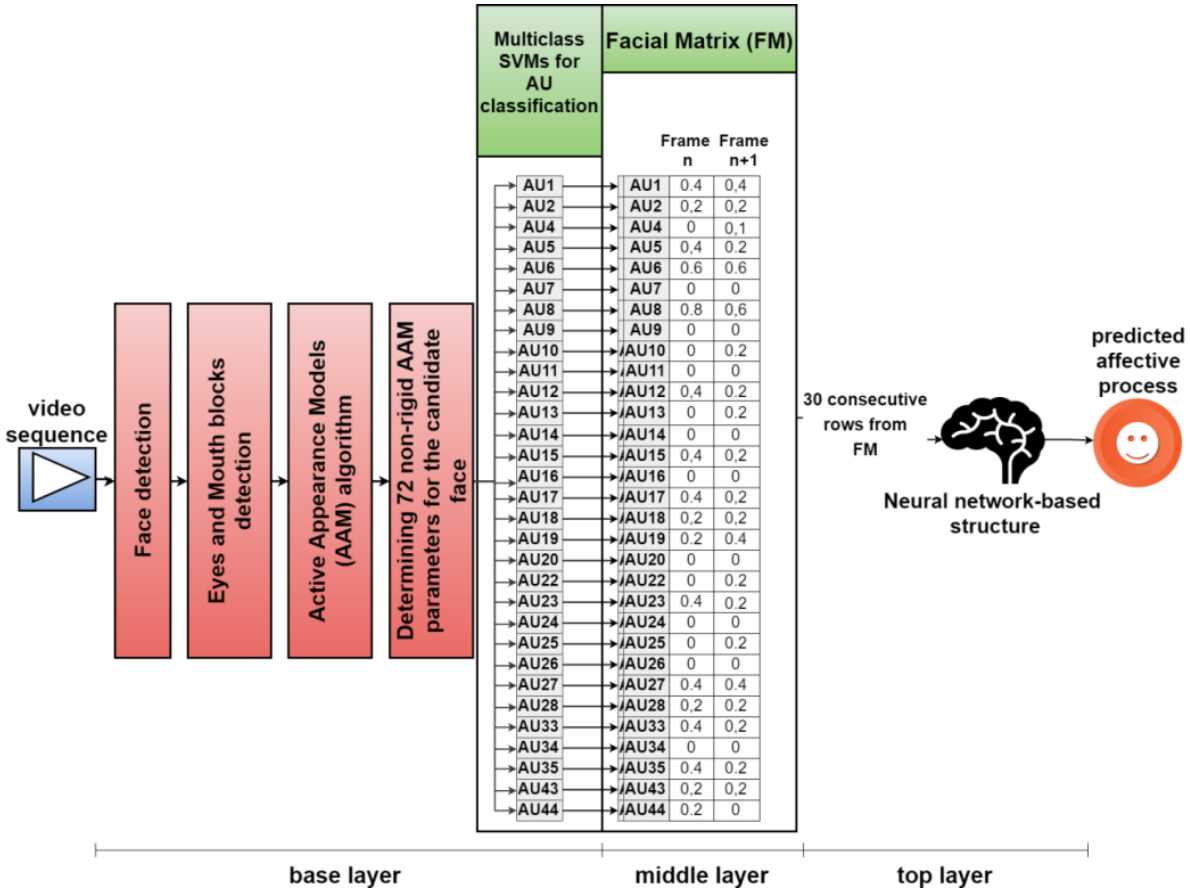

Fig. 4.1 *General architecture for predicting human affective processes based on face feature analysis*

For **emotion prediction based on face features analysis,** the optimal FFNN-based structure is comprised of 6 FFNNs dedicated for each of the six emotions; therefore, from a facial point of view, emotions are weakly correlated. The highest accuracy is obtained when the system is trained and tested using DSC (87.91%), while, when we use DSC only for the training phase, and we test using DSR, the accuracy decreases with maximum 2%. We show that DSC ads more value in the training phase, which makes the system practical as the facial expressions can be evaluated in random conditions (no emotion elicited), without a significant impact on accuracy. In this case, considered the most practically relevant, the maximum time needed to converge to a stable prediction is 33 seconds, significantly faster than the time required to fill in DEQ (12.1 minutes); hence, the method can be used successfully for replacing DEQ. The highest accuracy is obtained for *happiness*, followed by *disgust*, *sadness*, and *surprise*, while for *anger* and *fear*, the accuracy is lower than 86%. We analyze the AUs present at high intensity when each emotion is predicted correctly, and we determine a set of correlations between AUs and emotions which are equivalent to those from FACS [20], showing that the architecture is robust. We use these correlations for predicting the corresponding emotions directly when the associated AUs are present at high intensity, without being processed by the FFNN-based structure, and the accuracy increases with up to 5%. Compared with other state-of-the-art methods, the one proposed in this thesis offers the highest accuracy. This research was published in [48] and extended in [49] for predicting fatigue based on facial features, as well as in [50] and [51] by fusing it with methods for analyzing body posture and hand gestures, improving the system's accuracy with 7% compared to when only facial features are analyzed.

For **predicting FFM dimensions based on face features analysis**, the optimal FFNN-based structure contains a single FFNN which models all five dimensions. Therefore, from a facial point of view, the five FFM dimensions are strongly correlated. We obtain the highest accuracy when DSC is used for both training and testing phases (77.4%), while, if we keep DSC for training phase and we test the system using DSR, the accuracy decreases with less than 0.5%. The maximum time required for predicting FFM dimensions is 150 seconds, considerably lower than the time needed for filling in the FFM questionnaire (21.2 minutes), showing that the proposed architecture can replace the questionnaire method, also having the advantage of removing the subjectivity usually associated with questionnaire interpretation and allowing for real-time monitoring. The highest accuracy is obtained for *Extraversion* (83.5%) and *Openness for experience* (79.94%), while for other dimensions the accuracy is lower than 77%. We also determine a set of correlations between FFM dimensions and analyzed AUs, in the first study of this kind in the literature. Integrating these correlations leads to an increase in accuracy of up to 9%. Comparing other state-of-the-art methods with the proposed architecture, we obtain over 5% higher accuracy than the method based on Exaggeration Mapping (EM) [52] and the system based on Artificial Neural Networks (ANN) [53], reaching similar accuracy to the method based on CERT [54]. This research was published in [55] and extended in [56] for predicting the Sixteen personality factor (16PF).

For **predicting temperament based on face features analysis**, the highest accuracy is obtained when we use dedicated FFNNs for each of the five temperament types. Hence, the temperament types are not correlated from a facial point of view, analyzing them together, leading to a decrease in accuracy. We obtain the highest accuracy of 80.78% when we use DSC in the training as well as testing phases, the accuracy decreasing with over 13% when DSR is used for both phases. By keeping DSC only for the training phase, the accuracy decreases by less than 1%. The maximum time needed for predicting the temperament type is 95 seconds, significantly faster than the time needed to fill in the FIRO-B questionnaire (14.1 minutes). The sanguine temperament is predicted with the highest accuracy (85.12%), followed by supine (83%), and melancholic (80.9%) temperaments. We identify, in a first study of this kind in literature, a set of correlations between temperament types and FACS AUs. By modifying the proposed architecture to integrate these correlations, we obtain an increase in accuracy of 7%. We compare the proposed architecture with the ANN-based system described in [53], and we observe that our method offers an accuracy of up to 5% higher.

For predicting **emotional states based on face features analysis**, the optimal FFNN-based structure contains a single FFNN modeling all three emotional states which shows that they are strongly correlated from a facial point of view. The highest prediction accuracy is obtained by using DSC for both training and testing phases: 81.47% (stress), 79.83% (depression), and 68.65% (anxiety). Similar to the previous architectures, if we maintain DSC for training phase and we use DSR for testing phase, the accuracy decreases with less than 2%, therefore inducing emotions is needed only for training, while for testing the system, frontal face video recordings collected in non-emotion eliciting conditions are as efficient as those collected when emotions are induced. The highest accuracy for all three emotional states is obtained for the *Normal* and *Extreme Severe* levels, while for the intermediary levels (*Mild*, *Moderate*, and *Severe)*, the accuracy is lower and they are most often mistaken with their neighboring levels. Therefore, the proposed architecture offers the possibility of determining if a subject suffers from extreme depression, anxiety or stress, but, for a better assessment of the

actual severity of the emotional state, other methods need to be explored, such as including other AUs or multimodal analysis. The maximum time required for predicting the three emotional states is 64 seconds, significantly faster than the time needed for filling in the SAQ. The proposed architecture could replace SAQ if the prediction accuracy for the intermediary levels is improved. We also conduct the first study in the literature that identifies correlations between analyzed AUs and the three emotional states. By modifying the architecture to integrate these correlations, the accuracy increases with up to 5%. For predicting stress, the proposed architecture offers higher accuracy than other state-of-the-art methods on both MENTAL and ANUStressDB [26] databases. For predicting depression, the proposed architecture provides a 1.5% higher accuracy compared with the method based on Local Binary Patterns in Three Orthogonal Planes (LBP-TOP) and SVM [57]. This research was published in [58].

CHAPTER 5 – SPEECH FEATURES ANALYSIS

Although largely used in the last decades, analyzing speech is still a complex task because of its linguistic and semiotic variance, as well as that related to emotion or other affective processes [59], but also because of noise and reverberation, most architectures being tested using neutral utterances, recorded in the studio [59], in the absence of these conditions the results being significantly less efficient. The problem of choosing the right set of features from a broad spectrum of possible speech features is often complicated and done through TAE. Because of all these reasons, analyzing speech features for predicting affective processes is difficult.

As they are largely used in similar research papers [59], but also considering their robustness and the fact that they can be classified with high accuracy using computationally inexpensive methods, in this thesis we use the following speech features: pitch contour (number of peaks per second - *PMAXCOUNT*, median – *PMAXAVG*, variance – *PVAR*, gradient – *PGRAD*), intensity contour (number of peaks per second – *IMAXCOUNT*, median – *IMAXAVG*, variance – *IVAR*, gradient – *IGRAD*), Speech rate (SR), Pause rate (PR), Zero-crossing rate (ZCR), short-time energy (position, average value – *EAVG*, standard deviation – *ESTDEV*), jitter (JIT), shimmer (SHIM), fundamental frequency (FF), and 33 MFCC coefficients (*MFCC0 – MFCC32*). Fig. 5.1 depicts the structure of the base and middle layers. All architectures proposed in this chapter are based on these two layers, the difference consisting in the FFNN-based structure from the top layer. **The base layer** has the purpose of acquiring the speech signal representing the utterance and classifying a set of speech features. It is divided into three blocks. In the **normalization block** we use spectral subtraction for noise reduction [60], then we normalize the energy using Cepstral Mean Normalization (CMN) with Bayesian networks [61] and, finally, we normalize pitch using the method based on semitones [62]. We also use the methods published in [63], [64], and [65] for enriching the speech signal. In the **utterance-level speech feature analysis block** we analyze the last 7 seconds of the utterance every time a new frame is provided as input, the speech features analyzed in this block being: *PMAXCOUNT, PVAR, PGRAD, IMAXCOUNT, IVAR, SR, PR, STE, EAVG, ESTDEV, JIT, SHIM și FF*. In **frame-level speech feature analysis block,** we classify the following features for each new frame provided as input: *PMAXAVG, IMAXAVG, ZCR, JIT, SHIM, MFCC0, ..., MFCC32*. Therefore, for each frame from the utterance, we classify 50 speech features, and these are provided to the middle layer where the Speech Matrix (SM) is built, containing the normalized values for each feature in the current frame. When we have 30 new unprocessed rows in SM (corresponding with 30 new frames), these

are provided to the top layer which analyzes them in a pattern recognition task for predicting the analyzed affective process.

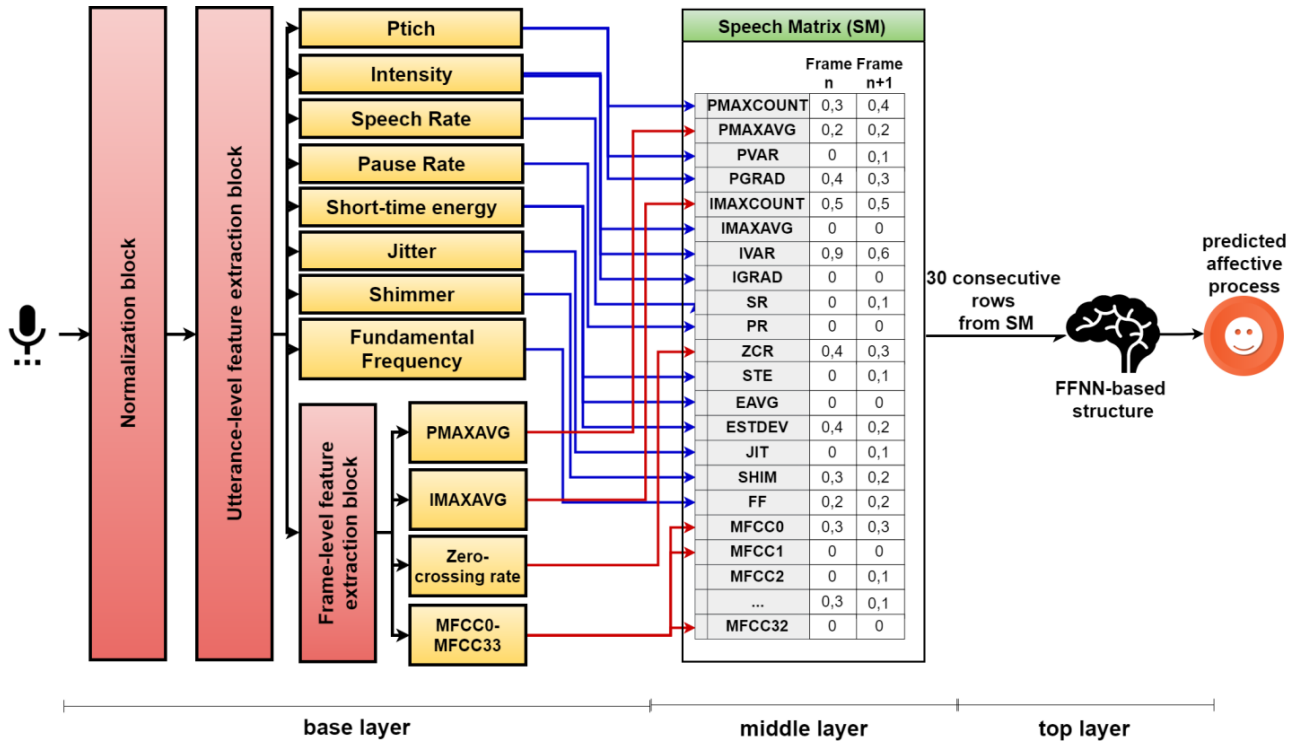

Fig. 5.1 *General architecture for predicting human affective processes based on speech features analysis*

For **predicting emotions based on speech features analysis**, the optimal FFNN-based structure contains a single FFNN used for predicting all six emotions. Compared with face features analysis where emotions are discrete and independent, from a speech point of view, the emotions are strongly correlated. We obtain an accuracy of 80.75% when the system is trained and tested using DSC and results 12% lower when we use only DSR for both training and testing phases. If we keep DSC for training phase, and we test the system using DSR, the accuracy decreases with less than 1.5%. The maximum time needed for predicting emotion is 78 seconds, significantly faster than the time needed to fill in DEQ (12 minutes), making the proposed architecture attractive for replacing the questionnaire, with the advantage of being suitable for real-time use. We obtain the highest accuracy for *happiness* (88.1%), *sadness* (81.8%), *fear* (80.5%), and *anger* (77.6%). For the first time in literature, we identify a set of correlations between emotions and the analyzed speech features which, used to enhance the architecture, lead to an increase of accuracy of up to 6%. Comparing the proposed architecture with other methods in the literature, we obtain an accuracy higher than the methods based on Particle Swarm Optimization (PSO) [66], Gaussian Mixture Models (GMM) [67][68], ANN and Hidden Markov Models (HMM) [69], SVM and K-Nearest Neighbors (KNN) [70], as well as Biogeography-based Optimization (BBO) and SVM [71]. Compared to the method based on Multi-scaled Sliding Window (MSW-AEVD) and HMM [72], our method offers better results except for the case where SALAS database is used, when the proposed architecture offers a lower accuracy, mainly because the method proposed in [72] is evaluated exclusively using induced emotions, while SALAS database contains only utterances collected in these conditions. This research was published in [73].

For predicting **FFM dimensions based on speech features analysis**, the optimal FFNN-based structure contains a single FFNN for predicting all five FFM dimensions, similar to the facial features analysis, which

shows that from a speech perspective the five FFM dimensions are also strongly correlated. We obtain an accuracy of 76.1% when DSC is used for training and testing phases, while, if we keep DSC for the training phase and we test using DSR, the accuracy decreases with less than 2%. The highest accuracy is obtained for *Extroversion* (78.4%), followed by *Openness for experience* (77.2%) and *Agreeableness* (76.3%). The maximum time needed for prediction is 120 seconds, a lot faster than the FFM questionnaire (21.1 minutes). Therefore, the proposed architecture can replace the FFM questionnaire and can also be used for real-time monitoring. We also determine a set of correlations between speech features and FFM dimensions, unique in literature, which leads to an increase in accuracy of up to 7%. Compared with other state-of-the-art methods, the proposed architecture offers a higher accuracy than the method based on linear kernel SVM [74], but the methods based on Wavelet transform with Convolutional Neural Networks (CNN) [75] and Frequency-domain linear prediction based on SVM [76] offer an accuracy with 0.7% and 3.7% higher.

For **predicting the temperament type based on speech features analysis**, the optimal FFNN-based structure is comprised of five dedicated FFNNs for each temperament type, showing that, similar to the face features analysis, temperament types are weakly correlated. The highest accuracy (79%) is obtained when the system is trained and tested using DSC, while, if we keep DSC for the training phase and we test the system using DSC, the accuracy decreases with less than 4%. Therefore, DSC is essential only for the training phase, while, for testing, we can use utterances collected in random conditions, showing that the proposed architecture is practical. The maximum time needed for predicting with high accuracy the temperament type is 78 seconds, significantly lower than the time needed to fill in the FIRO-B questionnaire, making the proposed architecture attractive for replacing the questionnaire-based method, with the advantage of being faster, more practical, and removing the subjectivity associated to questionnaire interpretation. We obtain the highest accuracy for sanguine and melancholic temperaments. For each temperament type we identify, for the first time in literature, sets of speech features to which they are are correlated and, by modifying the proposed architecture to use these correlations, the accuracy increases with up to 7%. Compared with other state-of-the-art methods, the proposed architecture offers the highest accuracy.

For **predicting emotional states based on speech features analysis**, the optimal structure has only one FFNN which evaluates all three emotional states, similar to the face features analysis. The highest accuracy is obtained when DSC is used for both training and testing phases: 86,3% (depression), 81,6% (anxiety), and 83,3% (stress). When we use DSC for the training phase and DSR for the testing phase, we obtain a decrease in accuracy of only 2%. The highest accuracy is obtained for the *Normal* and *Extreme Severe* levels, while for intermediary levels the accuracy is significantly lower and these are most often mistaken with the neighboring levels. The maximum time needed for predicting the three emotional states is 96 seconds, faster than the time needed to fill in SAQ. The proposed architecture could replace SAQ if the prediction accuracy of intermediary levels is improved. For the first time in literature, we determine a set of correlations between DASS levels and the analyzed speech features which we integrate into the proposed architecture, leading to an increase in accuracy of 3%. Comparing the proposed architecture with other state-of-the-art methods, we obtain the highest accuracy for stress prediction, while, for depression, methods based on ANN and SVM [77] or Dynamic Convolutional Neural Networks (DCNN) [78] offer a higher accuracy.

CHAPTER 6 – HANDWRITING FEATURES ANALYSIS

Creating systems that are capable of recognizing affective processes automatically based on handwriting, without needing a human observer or assessor, could provide graphology the relevance needed to be a field studied with more confidence [79]. These systems could prove useful for psychological analysis and diagnosis in both the psychology field [80], as well as medicine [81]. During a standard graphological analysis, experts analyze handwriting to determine a set of features, each containing information related to the emotional state, personality traits as well as other psychological aspects related to the writer [82]. The main handwriting features and those used in this paper are presented in Fig. 6.1 [83], alongside the structure for the base and middle layers used for the architectures proposed for handwriting features analysis.

The base layer has the main purpose of converting the scanned image containing the handwritten exemplar in a set of handwriting features identified based on a set of classification scores. The **normalization block** is responsible with noise reduction (for which three filters are used: boolean filter [83] for removing texturized background, a ramp width reduction filter for sharpening, and an adaptive unmask sharping filter for contrast adjustment [84]), contour smoothing for reducing possible errors caused by the unwanted movement of subject's hand while writing and for which we use the optimal local weighted averaging method [85], image compression for which we use the integral modified histogram [86], and isolating the handwriting through white space thinning method [87]. For **line-level segmentation and analysis,** we use the Vertical Projection Profile (VPP) method [88]. The *spacing between lines* feature is classified by determining the number of pixels that overlap between two bounding boxes delimited for two consecutive lines. The *baseline* feature is determined using the method described in [89] by studying the pixel density in each box that bounds a segmented line, and by rotating the box until the highest pixel density is centered horizontally. The *pen pressure* feature is determined using grey-level thresholding [90], analyzing values from the spectrum of the bounding box that delimits the analyzed line and computing the average for the segmented line. For **word-level segmentation and analysis,** we generate VPP [88] to determine the pixel density for each vertical column, and we identify the columns with the lowest density, these being possible inter-word spaces [88]. For *letter slant* feature we use the technique detailed in [90], calculating the probability density function of vertical pixels for different angles and for each column in the histogram determining the number of pixels and dividing it to the highest and lowest pixel from the analyzed word segment, the values from these columns being summed up and the angle where the calculated sum is the highest representing the letter slant. For **letter-level segmentation**, for each delimited word, we use Stroke Width Transform (SWT) to determine the median height of the strokes and we create a VPP for the word segment by determining the columns where the projected value is lower than 8% from the maximum projected value of the analyzed word (threshold determined through TAE). For *letter connection* feature we use the algorithm for letter segmentation anteriourly described, and we compare the width of each stroke connecting two bounding boxes of two consecutive letters with the average width of the strokes in the word. For the *lowercase letter "t"* and *lowercase letter "f"* features we use template matching (TM), and we compare each letter with a set of predefined models from MNIST database [91], classifying the matching using Euclidean similarity [90].

Applying the methods described previously, each handwriting feature is classified with an accuracy higher than 90%. The base layer offers classification scores for each handwriting feature, normalized in [0;1] interval. For each letter in the handwritten exemplar, the classification scores obtained for each feature are provided as input to **the middle layer** and are stored in a new line in the Handwriting Matrix (HM).

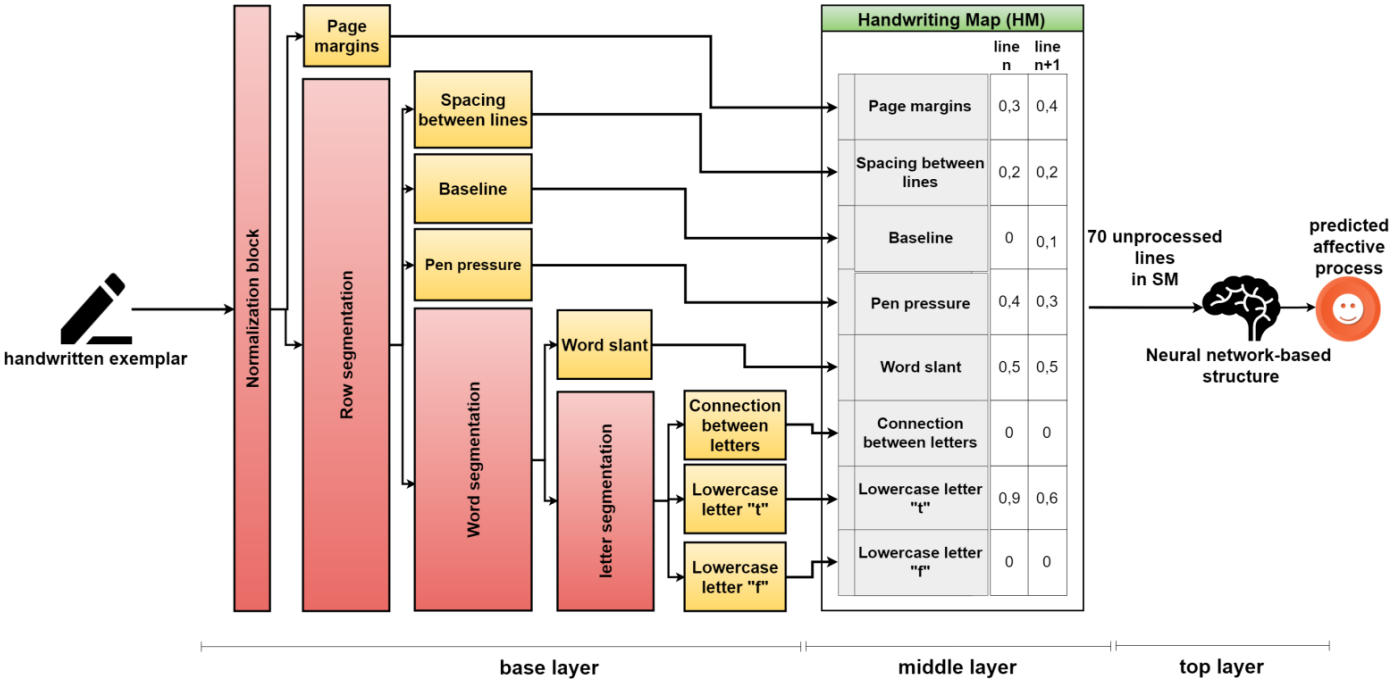

Fig. 6.1 *General architecture for predicting human affective processes based on handwriting features analysis*

For **predicting emotions based on handwriting features analysis**, the optimal FFNN-based structure is obtained when we use a single FFNN to model all six emotions. From a graphological point of view, emotions are not discrete, analyzing them in a correlated manner leading to higher accuracy. The highest accuracy is obtained when DSC is used for both training and testing phases (68.83%), and, similar to previous architectures, we observe a decrease of 1.5% accuracy when DSC is maintained for training phase, and DSR is used for testing the architecture, showing that handwritten exemplars with predefined text are needed only for training, while end-users can write a freely chosen text without significantly impacting the performance of the system. The time needed to predict the subject's emotion accurately is 9 minutes, lower than the time necessary to fill in the DEQ questionnaire. Hence, the architecture can replace the questionnaire method if the prediction accuracy is improved. We obtain the highest accuracy for *sadness* (72.92%), followed by *happiness* (71.49%), other emotions being predicted with an accuracy lower than 70%. We determine, for the first time in literature, a set of handwriting features that are associated with each of the six emotions. Modifying the architecture to integrate these correlations leads to an increase in accuracy with 6%. We use MENTAL to test other state-of-the-art methods, results showing that the proposed architecture offers similar results. This architecture was published in [93] where it was used for predicting blood pressure based on handwriting, reaching a prediction accuracy of over 90%.

For **predicting FFM dimensions based on handwriting features analysis**, the optimal FFNN-based structure uses a single FFNN for predicting all five dimensions. The highest accuracy is obtained when DSC is used for both training and testing phases (82.34%). The accuracy decreases with less than 2% when DSC is used for the training phase and DSR for the testing phase, showing that exemplars with predefined text are vital only for training the system. The highest accuracy is obtained for *Openness for experience* (84.6%), followed

by *Extraversion* (83.58%), and *Neuroticism* (82.39%). The maximum time needed for predicting FFM dimensions is of approximately 8 minutes, faster than the FFM questionnaire, and, hence, making the proposed architecture suitable for replacing the questionnaire-based method. We identify a set of correlations between the five FFM dimensions and the analyzed handwriting features which are unique in literature. By integrating these correlations in the proposed architecture, the accuracy increases with up to 6%. We use MENTAL for testing different state-of-the-art methods, and we show that the proposed architecture offers the highest accuracy in the literature, the only method that achieves similar results being the one based on a combination of SVM, AdaBoost and KNN classifiers [94]. The research was published in [95] and extended in [96] for predicting the Myers-Briggs type indicator, as well as in [97] by fusing the FFNN-based method with SVMs, improving the accuracy with up to 2%.

For **predicting temperament type based on handwriting features analysis**, the optimal structure is one in which FFNNs dedicated for each of the five temperament types are used. From a graphological point of view, similar to the results obtained when analyzing face and speech, temperament types are weekly correlated. We obtain the highest accuracy when DSC is used for both training and testing phases (81.7%). Similarly, we observe a decrease in accuracy of only 4% when DSC is maintained for the training phase, and DSC is used for testing. The time needed for accurately predicting the temperament type is 7 minutes, lower than the one needed for filling in the FIRO-B questionnaire. The proposed architecture is, therefore, faster and more efficient than the standard questionnaire, and can be successfully used for replacing it. We obtain the highest accuracy for supine, sanguine, and melancholic temperament types. We determine, in this case as well, for the first time in literature, a set of correlations between analyzed handwriting features and temperament types which lead to an accuracy increase of 6%. We use MENTAL to test other state-of-the-art algorithms, and we show that our proposed architecture offers the highest accuracy.

For **predicting emotional states based on handwriting feature analysis**, the optimal structure uses only one FFNN for modeling all three emotional states. Hence these are strongly connected from a graphological point of view, confirming previous studies that identify this possible connection [98]. The highest accuracy is obtained when DSC is used for both training and testing phases: 82.69% (depression), 79.84% (anxiety), and 81.53% (stress). If we use DSC only for training and we test the system on DSR, the accuracy decreases with less than 2%. Therefore, subjects can freely choose the text they write, the system offering results as precise as when the handwritten samples contain a predefined text on which the architecture was anteriourly trained. We obtain the highest accuracy for *Normal* and *Extreme Severe* levels, while, for intermediary levels, the accuracy is significantly lower. The maximum time needed for prediction is 10 minutes, faster than filling in the SAQ, offering also the possibility of monitoring the three emotional states at any given moment in time without the need of psychologists, also removing the subjectivity associated with SAQ. We identify, for the first time in literature, a set of correlations between the analyzed handwriting features and the three emotional states. By integrating these correlations in the proposed architecture, the accuracy increases with 3%. We compare the proposed architecture with the method based on Random Forest Models (RFM) [33] using MENTAL and EMOTHAW databases. We observe that the proposed method offers higher accuracy than the RFM-based one in both cases, proving to be robust as we obtain consistent results on two different databases.

CHAPTER 7 – MULTIMODAL ANALYSIS

In previous chapters, we evaluated the possibility of predicting different human affective processes based on several unimodal analyses, studying the face, speech, and handwriting features. In this chapter, we aim to identify the optimal ways in which the previously described unimodal architectures can be fused to improve the prediction accuracy and make such architectures more attractive for day to day use, as well as for replacing the specific questionnaires which have the main advantage of being often prone to bias. Because we aim for a real-time prediction, the handwriting features are used only for improving the prediction accuracy of the other two modalities (face and speech) because acquiring and evaluating handwriting cannot be done in the same time as face and speech analysis. As we showed, handwriting features offer high accuracy for predicting personality traits and temperament types, so we use a handwritten exemplar collected from them analyzed subject for determining a set of behavioral parameters stored in a vector called Behavioral Stamp (BS) and which are used to weight the results from the other two modalities. Therefore, we propose five types of fusion**: feature-level fusion for face and speech (FLF-FS)** by using a FFNN or set of FFNNs to analyze the combined vector of face and speech features, **feature-level fusion for face and speech with behavioral stamp (FLF-FS-BS)** in which we use FLF-FS and we add an additional layer in which the results provided by FLF-FS are weighted based on a BS provided by another FFNN, previously trained on exemplars collected from multiple subjects and which receive as input an exemplar handwritten by the analyzed subject, **score-level fusion for face and speech (SLF-FS)** which computes the maximum of the two scores provided by the two unimodal architectures, **score-level fusion for face and speech with behavioral stamp (SLF-FS-BS)** in which we modify SLF-FS by calculating the weighted maximum of the scores provided by the two unimodal analyses, the weights being provided by the BS generated by a pre-trained FFNN which analyses a handwritten exemplar from the current subject, and **decision-level fusion for face and speech** which is activated if, for 5 consecutive seconds, the predictions provided by the two unimodal architectures (for face and speech analysis) offer the same result. Otherwise, the result is marked as *Undefined.* The fusion techniques based on Behavioral Stamp are unique in literature, being first presented in this thesis.

For **predicting emotion based on multimodal analysis**, the highest accuracy is obtained when we use FLF-FS-BS, the handwriting features adding over 5% more accuracy compared to FLF-FS, which shows that BS ads value to the emotion prediction accuracy. The optimal FFNN-based structure contains a common FFNN with a single hidden layer, having *tanh* as the activation function for the hidden layer and *sigmoid* for the output layer. This FFNN has 2430 input nodes and six output nodes normalized in [0;1] interval and representing the predicted emotion. The optimal number of neurons determined through TAE is 980, the optimal learning rate is 0.2, the optimal momentum is 0.03, and 25000 epochs are needed for training. The optimal configuration for the FFNN used for determining BS contains one hidden layer with *ReLU* as activation function and *sigmoid* as the activation function for the output layer. The optimal number of hidden neurons determined through TAE is 230, the optimal learning rate is 0.2, the optimal momentum is 0.04, and we need 12000 training epochs. As in the case of unimodal analyses, the highest accuracy is obtained when DSC is used for both training and testing the system: 91.84%. The accuracy decreases with less than 1.5% when we keep DSC for training the system, and we test it using DSR. The end-user only needs to provide a handwritten exemplar once to calibrate the

system, while any subsequent assessment of his/her emotion can be done in random conditions without significantly impacting the accuracy. We need, on average, 94 seconds for predicting with high accuracy subject's emotion (90.46%), substantially faster than the time needed for filling in the DEQ questionnaire, making the system suitable for replacing the questionnaire-based method, as well as offering the possibility of monitoring the subject in real-time with high accuracy. We obtain the highest accuracy for *sadness* (92.84%), followed by *happiness* (91.84%), *surprise* (90.12%), and *disgust* (90.03%). We determine that the accuracy is now balanced across each emotion compared to the unimodal architectures, which shows that FLF-FS-BS grips the emotionally-relevant features from each modality to provide robust and precise results. The correlations determined in this case between the analyzed features and emotions are similar to those obtained through unimodal analyses, enhancing the idea that the proposed architecture is robust. By using these correlations to improve the architecture, we obtain an accuracy of over 91%. Comparing state-of-the-art methods with the proposed architecture on different databases using FLF-FS (as other databases don't contain handwritten exemplars), we observe that our method offers an accuracy higher than all other architectures in literature.

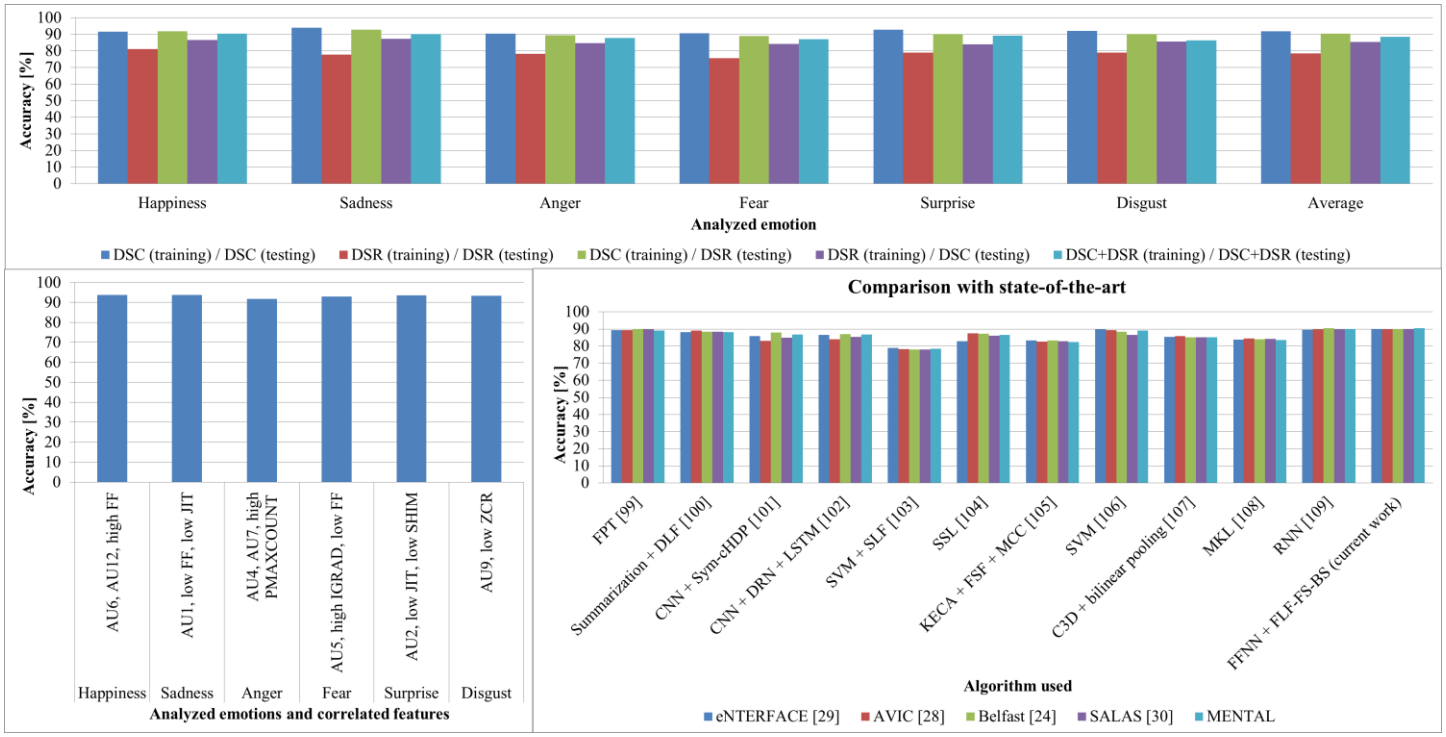

Fig. 7.1 *Emotion prediction based on multimodal analysis*

For **predicting FFM dimensions based on multimodal analysis**, the optimal fusion method is SLF-FS-BS. The optimal FFNN for generating BS has two hidden layers, *tanh* as the activation function for the two hidden layers, and *sigmoid* for the output layer. The optimal number of neurons determined through TAE is 55 for the first layer and 35 for the second layer, the optimal learning rate is 0.02, the optimal momentum is 0.02, and we need 12000 training epochs. A bias is needed in the second hidden layer. The highest accuracy is obtained when DSC is used for both training and testing the system (88.7%), while, by using DSC for training and DSR for testing, the accuracy decreases with only 4%, showing, in this case as well, that the architecture is practical. The time needed for accurately predicting FFM dimensions is 115 seconds, significantly lower than that required for filling in the FFM questionnaire (21.2 minutes). Hence the architecture is fast and suitable for real-time monitoring. The highest accuracy is obtained for *Extraversion* (90.25%), followed by *Openness to experience* (86.51%), and *Agreeableness* (86.57%). We determine correlations similar to the ones obtained

through unimodal analyses, showing that the proposed architecture is robust. By using these combinations of face and speech features to optimize the proposed architecture, the accuracy increases with 5%. We test the proposed architecture in comparison with other state-of-the-art methods on the MENTAL database, and we show that our method offers the highest accuracy, 4% higher than the most accurate method in literature based on ANN with score level fusion (SLF) [112].

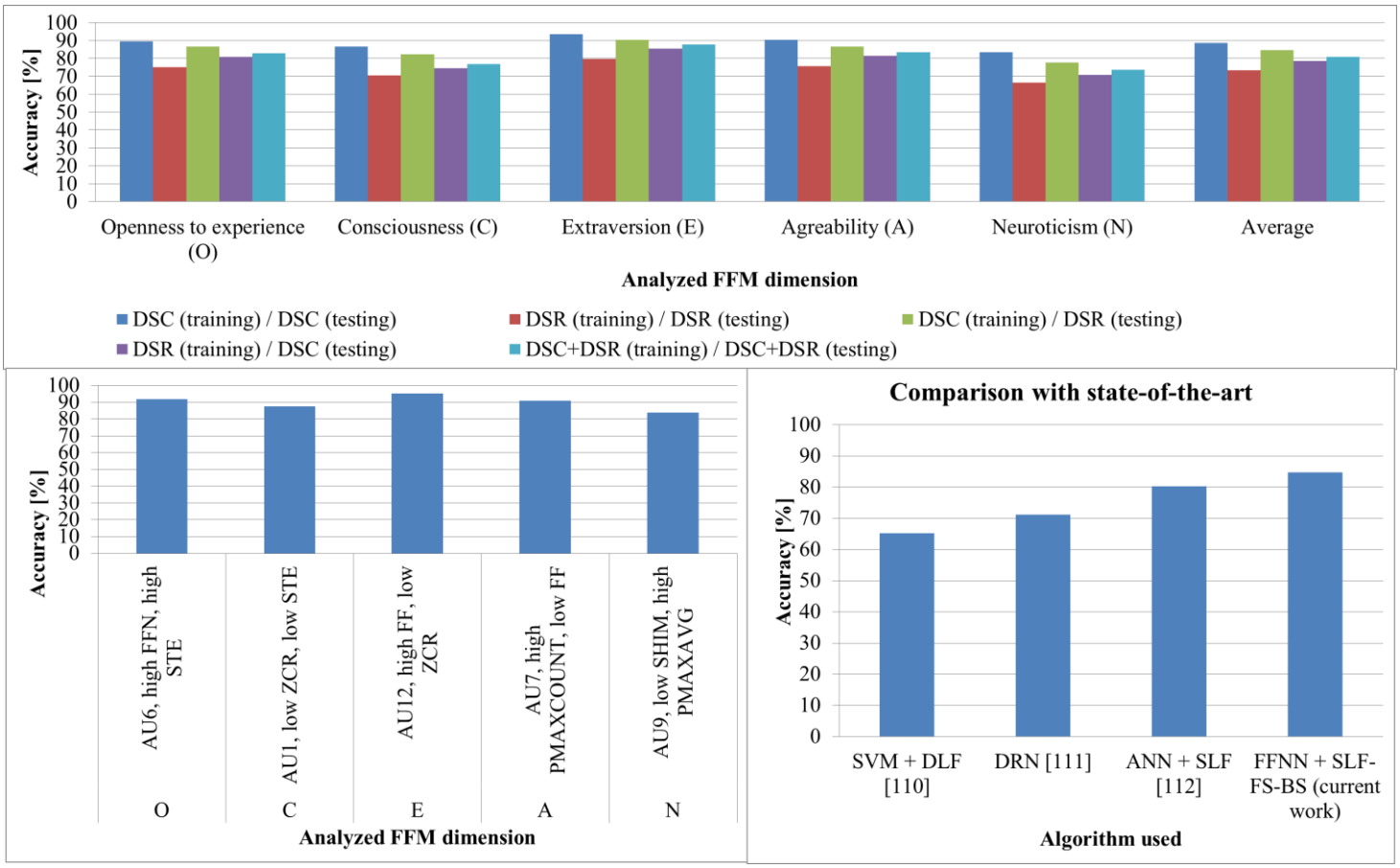

Fig. 7.2 *FFM dimension prediction based on multimodal analysis*

For **predicting temperament type based on multimodal analysis**, the most accurate fusion method is SLF-FS-BS, similar to the multimodal architecture for predicting FFM dimensions. For generating the BS, the optimal FFNN has two hidden layers, with *sigmoid* as the activation function for the first hidden layer and the output layer, and *tanh* for the second hidden layer. The optimal number of hidden neurons is 65 for the first layer and 65 for the second layer, the optimal learning rate is 0.3, the optimal momentum is 0.02, and 8000 training epochs are required. The highest accuracy is obtained when DSC is used for both training and testing phases (90.41%), with 1.2% lower than when we keep DSC for training and DSR is used for testing the system, showing that inducing the emotion is necessary only for training the system, while testing can be done in naturalistic conditions without degrading system's accuracy. The maximum time needed for predicting the personality type is 117 seconds, considerably lower than the time needed for filling in the FIRO-B questionnaire (14.2 minutes). Being fast and accurate, the proposed architecture can be successfully used for replacing the FIRO-B questionnaire. We obtain an accuracy of over 85% for all five temperament types, the highest being obtained for sanguine (92.11%), followed by melancholic (91.17%), and supine (90.64%) temperament types. The correlations identified in this case are equivalent to those obtained through unimodal analyses, showing that the proposed architecture is robust. Using these correlations, the accuracy increases with up to 5%, all temperament types being predicted with over 90% accuracy. Compared with other methods in the

literature, the proposed architecture offers an accuracy 9% higher than the best state-of-the-art method for assessing the personality type based on the multimodal analysis.

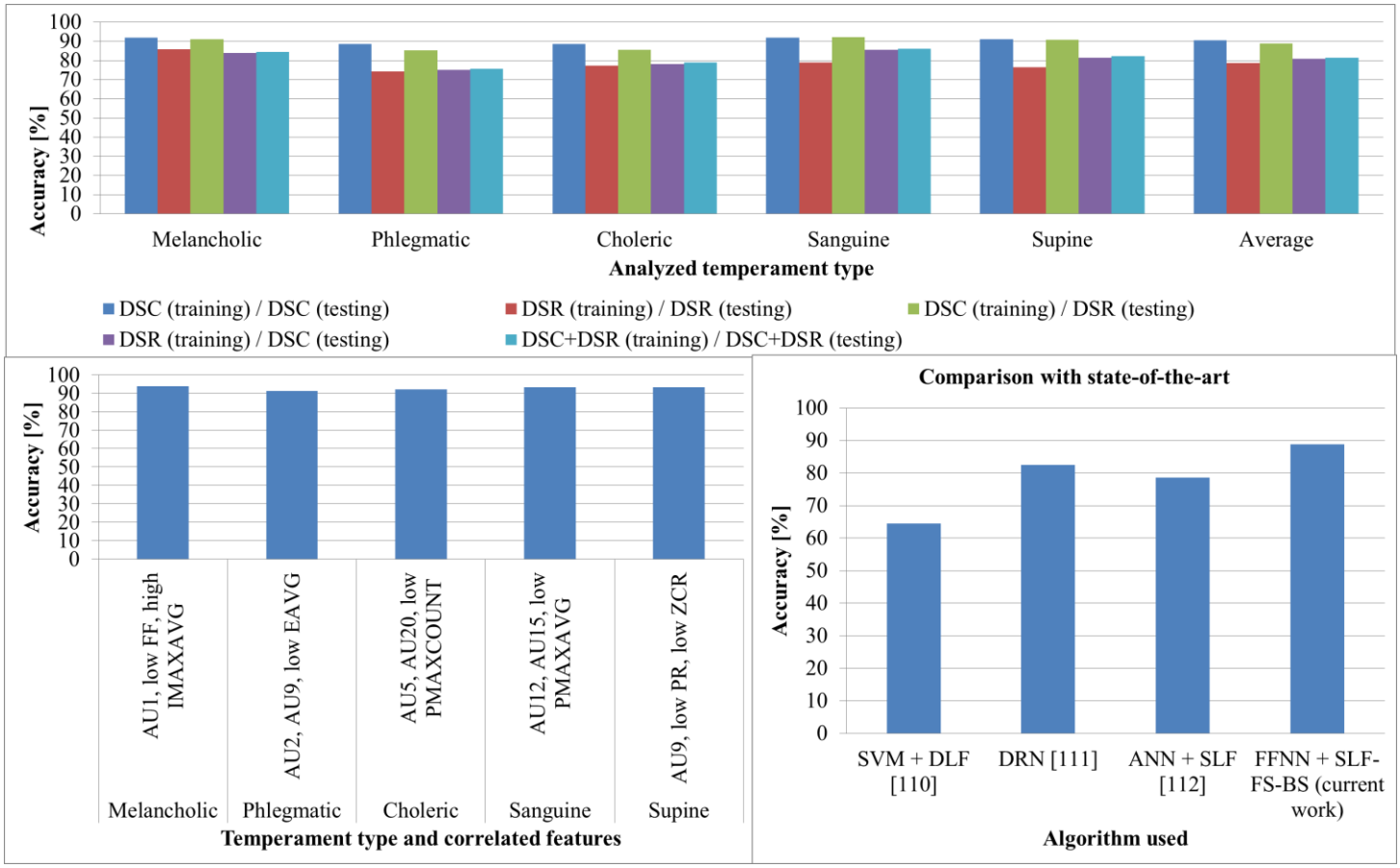

Fig. 7.3 *Temperament type prediction based on multimodal analysis*

For **predicting the emotional states based on multimodal analysis**, the optimal fusion method is DLF-FS. The handwriting features do not contain additional information compared to that provided by the other two modalities (face and speech), an optimal structure being one that analyzes only the face and speech features, without taking into account handwriting features. The highest accuracy is obtained when DSC is used for both training and testing phases: 91.58% (depression), 90.2% (anxiety), and 90.48% (stress). If DSC is maintained for training and DSC is used for testing, the accuracy decreases with less than 1.5%. All DASS levels are predicted with an accuracy of over 89%, which shows that fusing the two features solves the problem encountered in the case of unimodal analyses where predicting intermediary levels was done with a significantly lower accuracy compared to extreme levels. By analyzing the classification results from FM and SM when the three emotional states are predicted with high accuracy, we obtain correlations similar to those obtained through unimodal analyses, showing that the proposed architecture is robust. Using these correlations between speech and face features and the three emotional states, the accuracy is improved with up to 4%, all three emotional states being predicted with over 93% accuracy. We compare the proposed architecture for depression prediction with the method based on FLF through Linear Regression Classifiers (LRC) [113] and that based on Space-Temporal Interesting Point (STIP) [114] on MENTAL and AVEC2014 [27] databases. Results show that our proposed architecture offers the highest accuracy in literature for depression prediction through multimodal analysis.

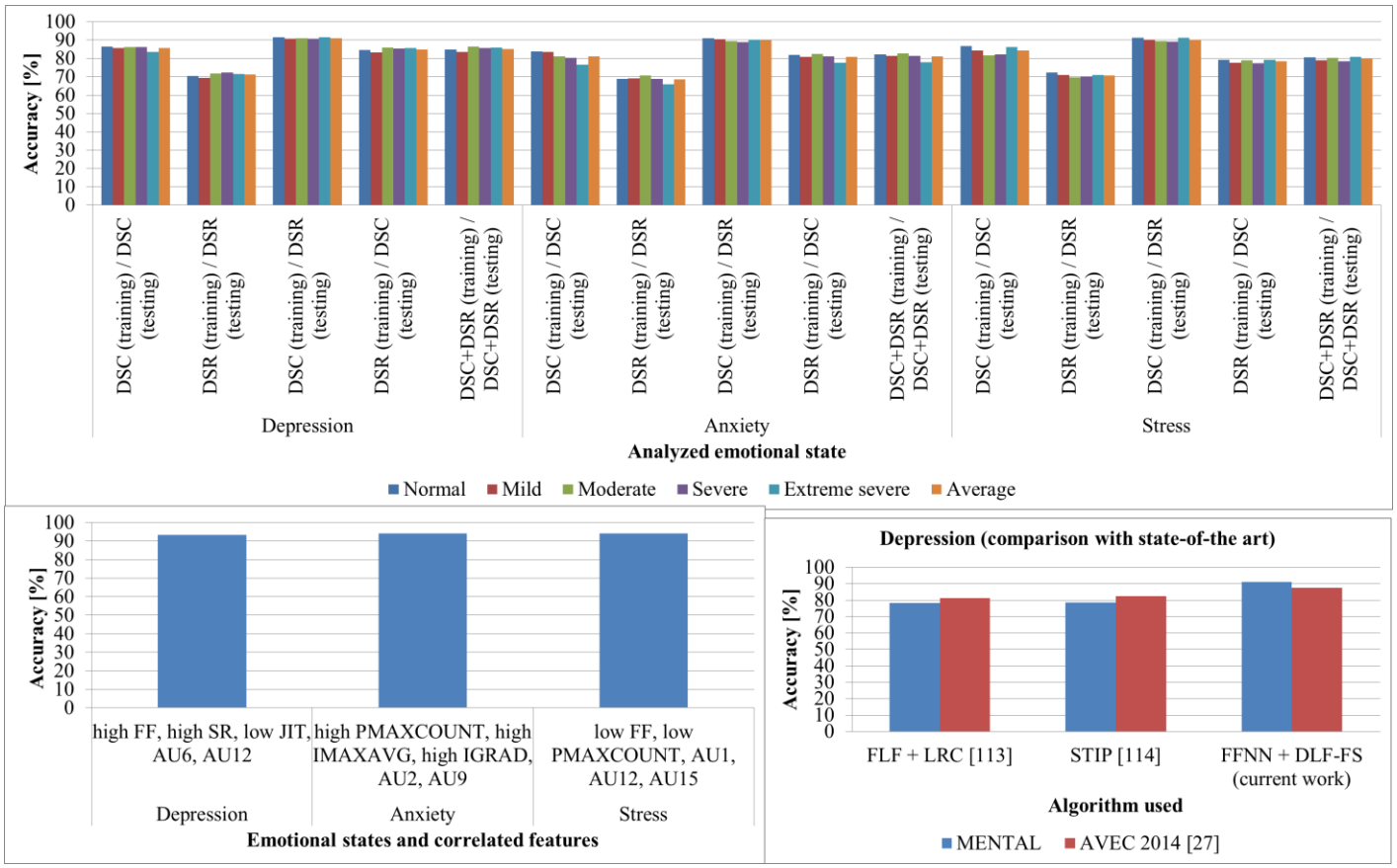

Fig. 7.4 *Emotional state prediction based on multimodal analysis*

CONCLUSIONS

In this thesis, we conduct a series of studies in which we analyze the possibility of predicting emotions, personality traits, temperament types, and emotional states through unimodal architectures (face, speech or handwriting) as well as by fusing the three modalities in a multimodal architecture, determining the optimal ways in which these modalities can be analyzed to ensure high accuracy, sensibility, and specificity. All proposed architectures are unique in literature and are based on different configurations of FFNNs, one of the purposes of this research being to determine the ideal neural network configuration and the optimal fusion method, alongside the main purpose of building a system capable of recognizing with high accuracy and in a timely manner human affective processes, using exclusively non-invasive methods. Each of these architectures is tested using intra-subject and inter-subject methodologies and in both cases we observe that the dataset containing emotion stimulated samples or predefined handwritten text are vital only for training the architecture, while, if it is used for testing, the accuracy is not significantly different than the one obtained using a random dataset, which shows that the proposed architectures are practical, as the analyzed subject doesn't have to watch emotion-inducing videos or write predefined texts for the affective process to be predicted, the system only needs to be pre-trained on such samples. We also propose two unique fusion models for the face and speech modalities based on Behavioral Stamp generated using the handwriting features which are unique in literature. In the multimodal analysis, we build an architecture for predicting emotions using a unique feature-level fusion method for face and speech by infusing the handwriting modality through BS, offering an accuracy of 90.46% (3% higher compared to other state-of-the-art methods) and a processing time of 94 seconds which makes this architecture suitable for replacing the DEQ questionnaire. We also build an architecture for predicting FFM dimensions and temperament type based on an unique score-level fusion for face and speech

modalities by infusing the handwriting features through BS, obtaining 84.71% accuracy for predicting FFM dimensions (4% higher than other state-of-the-art methods) with a processing time of 115 seconds and 88.85% accuracy for predicting temperament type (6% higher than other state-of-the-art methods) with a processing time of 117 seconds. These architectures can successfully replace the FFM and FIRO-B questionnaires. For predicting DASS levels, we show that decision-level fusion for face and speech modalities is optimal, obtaining over 90% accuracy for all three emotional states and all their levels, with a processing time of 112 seconds, also being suitable for replacing SAQ. This thesis also contains the first study in literature that determines the relation between the analyzed features and the evaluated human affective processes, obtaining correlations that are used to improve the accuracy of the proposed architectures, leading to an increase of up to 2% for emotion prediction, 5% for FFM dimensions and temperament prediction, and over 3% for DASS level prediction.

The proposed architectures also have a large spectrum of applications, besides replacing the standard questionnaires. The architectures for predicting emotions can be successfully integrated into any human-computer interface with face or speech interaction, as well as in applications for professional or personal development or virtual psychologist to monitor the subjects' emotional patterns and offer personalized advice [115]. Also, the architectures can be used in computer-mediated education software where the tutor can evaluate the emotional state of his/her students and can adapt the presentation to improve the learning rate [116]. The car's or airplane's computerized system can also use such architectures by analyzing the facial expressions or speech of the driver or pilot and take preventive measures to ensure their and the passengers' safety [117]. Knowing the fact that the emotional patterns of a subject can be associated with different mental disorders [118] or physical diseases [119], the proposed architecture can be used to analyze and monitor them in real-time. Also, the architecture could help people suffering from autism to integrate in society by helping them recognize emotions in others [120]. Of course, the proposed architectures can be useful in forensic departments to evaluate suspects' emotions and determine the perpetrator [59]. They can also be used to personalize web sites or for online advertising based on subjects' emotions [121]. Last but not least, facial expressions can be used to make the face recognition systems immune to spoofing attacks, as we showed in [117]. The architectures proposed for predicting FFM dimensions can be used to determine people who are suffering from substance use [13] or those that are suffering from chronic physical diseases, such as heart disease, cancer, diabetes or respiratory diseases which are correlated with different FFM dimensions patterns. Given the correlations between FFM dimensions and learning types, the architectures can also be used to adapt the course material to the student's learning type. The architectures proposed for predicting temperament can be used for efficient team building, as well as for career counseling and professional development [15]. The architectures proposed for determining the levels of anxiety, depression or stress can be used for psychological profiling of employees [16], adaptive learning [17], as well as diagnosing physical diseases that have as symptoms different patterns of these emotional states [18].

We observe that some of the described architectures offer a high accuracy which makes them suitable for being used in the applications mentioned previously, others require drastic improvement of their prediction ability before being released for practical use. In future studies, other AUs alongside the 31 that are analyzed in this thesis can be included, as well as the body position or hand gestures that can carry relevant information to

allow for an increased prediction accuracy (as we showed in [50] and [51]). Similarly, we can include more speech or handwriting features. The context in which the samples are collected can be changed to improve the accuracy of some of the analyzed affective processes, such as *Agreeableness*, which manifests in social contexts which are not considered in the proposed database. Other methods based on other types of neural networks (CNN or RNN) or other algorithms (SVM or HMM) can be evaluated separately or fused with the methods proposed in this thesis to increase prediction accuracy and reduce processing time.

LIST OF ORIGINAL PAPERS

1. M. Gavrilescu, *Proposed architecture of a fully integrated modular neural network-based automatic facial emotion recognition system based on Facial Action Coding System*, **International Conference on Communications (COMM)**, pp. 1-6, Bucharest, 29-31 May 2014, WOS: 000345844600098.
2. M. Gavrilescu, *Improved automatic speech recognition system using sparse decomposition by basis pursuit with deep rectifier neural networks and compressed sensing recomposition of speech signals*, **International Conference on Communications (COMM)**, pp. 1-6, Bucharest, 29-31 May 2014, WOS: 000370971100021.
3. M. Gavrilescu, *Noise robust Automatic speech Recognition system by integrating Robust Principal Component Analysis (RPCA) and Exemplar-based Sparse Representation*, **7th International Conference on Electronics, Computers and Artificial Intelligence**, pp. S-29-S-34, Bucharest, 25-27 June 2015, WOS: 000370971100022.
4. M. Gavrilescu, *Improved Automatic Speech Recognition System by using compressed sensing signal reconstruction based on L0 and L1 estimation algorithms*, **7th International Conference on Electronics, computers and Artifical Intelligence**, pp. S-23-S-28, Bucharest, 25-27 June 2015, WOS: 000345844600055.
5. M. Gavrilescu, *Study on determining the Big-Five personality traits of an individual based on facial expressions*, **E-Health and Bioengineering Conference**, pp. 1-6, Iași, 19-21 November 2015, WOS: 000380397900257; ***The paper was awarded „Young Researcher" prize.***
6. M. Gavrilescu, *Study on determining the Myers-Briggs personality type based on individual's handwriting*, **E-Health and Bioengineering Conference**, pp. 1-6, Iași, 19-21 November 2015, WOS: 000380397900256.
7. M. Gavrilescu, *Recognizing emotions from videos by studying facial expressions, body postures and hand gestures*, **23rd Telecommunications Forum Telfor**, pp. 720-723, Belgrade, 24-26 November 2015, WOS: 000380397000162.
8. M. Gavrilescu, *Study on using individual differences in facial expressions for a face recognition system immune to spoofing attacks*, **IET Biometrics**, Vol. 5, Issue 3, pp. 236-242, September 2016, IF2016 = 1.382, Q3, WOS: 000382809100010.
9. M. Gavrilescu, *Recognizing human gestures in videos by modeling the mutual context of body position and hands movement*, **Multimedia Systems**, Vol. 23, Issue 3, pp. 381-393, June 2017, IF2017 = 1.703, Q2, WOS: 000438134600003.
10. M. Gavrilescu, N. Vizireanu, *Predicting the Sixteen Personality Factors (16PF) of an individual by analyzing facial features*, **EURASIP Journal on Image and Video Processing**, Vol. 59, August 2017, IF2017 = 1.737, Q2, WOS: 000438134600003.
11. M. Gavrilescu, N. Vizireanu, *Using Off-line handwriting to predict blood pressure level: a neural network-based approach*, **International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures (FABULOUS)**, pp. 124-133, Bucharest, 12-14 October 2017, WOS:000481658200019.
12. M. Gavrilescu, N. Vizireanu, *Neural Network Based Architecture for Fatigue Detection Based on the Facial Action Coding System***, International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures (FABULOUS)**, pp. 113-123, Bucharest, 12-14 October 2017, WOS:000481658200018.
13. M. Gavrilescu, *3-Layer Architecture for determining the personality type from handwriting analysis by combining neural networks and support vector machines*, **UPB Scientific Bulletin**, Series C, Vol. 79, Issue 4, pp. 135-150, October 2017, WOS:000424339100011, ISI.
14. M. Gavrilescu, N. Vizireanu, *Predicting the Big Five personality traits from handwriting*, **EURASIP Journal on Image and Video Processing**, Vol. 57, December 2018, IF2018 = 1.534, Q3, WOS: 000438134600003.

15. M. Gavrilescu, N. Vizireanu, *Feedforward Neural Network-Based Architecture for Predicting Emotions from Speech*, **Data,** Vol. 4, Issue 3, pp. 101, July 2019, ISI.
16. M. Gavrilescu, N. Vizireanu, *Predicting Depression, Anxiety, and Stress Levels from Videos Using the Facial Action Coding System*, **Sensors**, Vol. 19, issue 17, August 2019, IF2017 = 3.031, Q1.

The 16 mentioned papers were published in 7 ISI journals (1 Q1, 2 Q2, 2 Q3) and 9 ISI conferences. The papers received a total number of 45 citations.

REFERENCES

[1] S. Petrovica, A. Anohina-Naumeca, H. Kemal Ekenel, *Emotion Recognition in affective Tutoring Systems: Collection of Ground-truth Data*, **Procedia Computer Science**, Vol. 104, pp. 437-444, 2017.

[2] G.H. Graham, J. Unruh, P. Jennings, *The Impact of Nonverbal Communication in Organizations: A Survey of Perceptions*, **International Journal of Business Communication**, 1991.

[3] M. Szwoch, W. Szwoch, *Emotion Recognition for Affect Aware Video Games*, **Image Processing & Communications Challenges**, Vol. 6, pp. 227-236, 2015.

[4] A. Fernandez-Caballero, A. Martinez-Rodrigo, J.M. Pastor, J.C. Castillo, E. Lozano-Monasor, M. T. Lopez, R. Zangroniz, J. M. Latorre, A. Fernandez-Sotos, *Smart environment architecture for emotion detection and regulation*, **Journal of Biomedical Informatics**, Vol. 64, pp. 55-73, December 2016.

[5] J.E. Steiner, *Human Facial Expressions in Response to Taste and Smell Simulation*, **Advances in Child Development and Behavior**, Vol. 13, pp. 257-295, 1979.

[6] M. Pantic, A. Pentland, A. Nijholt, T.S. Huang, *Human Computing and Machine Understanding of Human Behavior: A survey*, **Artificial Intelligence for Human Computing**, pp. 47-71, 2007.

[7] R.W. Picard, *Affective Computing*, **MIT Press**, 2000.

[8] M. Clynes, *Sentics: The Touch of the Emotions*, **Anchor Press**, 1978.

[9] P. Ekman, W.V. Friesen, *Facial Action Coding System: Investigator's Guide*, **Consulting Psychologists Press**, 1978.

[10] P. Ekman, W.V. Friesen, J.C. Hager, *Facial Action Coding System. Manual and Investigator's Guide*, **Salt Lake City, UT: Research Nexus**, 2002.

[11] P.T.Jr. Costa, R.R. McCrae, *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual*, **Psychological Assessment Resources**, Odessa, Florida, 1992.

[12] N. Zilberman, G. Yadid, Y. Efrati, Y. Neumark, Y. Rassovsky, *Personality profiles of substance and behavioral addictions*, **Addictive Behaviors**, Vol. 82, pp. 174-181, July 2018.

[13] S.J. Karau, R.R. Schmeck, A.A. Avdic, *The big five personality traits, learning styles, and academic achievement*, **Journal on Personality and Individual Differences**, Vol. 51, Issue 4, pp. 472-477, September 2011.

[14] W.C. Schutz, FIRO: *A Three Dimensional Theory of Interpersonal Behavior*, **Holt, Rinehart, & Winston**, New York, 1958.

[15] E. Schnell, A. Hammer, *FIRO-B: Technical Guide*, **Mountain View**, CPP, 2000.

[16] S.H. Lovibond, P.F. Lovibond, *Manual for the Depression Anxiety Stress Scales*, **Psychology Foundation**, Sydney, 1995.

[17] P. Jafari, F. Nozari, F. Ahrari, Z. Bagheri, *Measurement invariance of the Depression Anxiety Stress Scales-21 across medical student genders***, International Journal of Medical Education**, Vol. 8, pp. 116-122, 2017.

[18] M.A.M. Gomaa, M.H.A. Elmagd, M.M. Elbadry, R.M.A. Kader, *Depression, Anxiety and Stress Scale in patients with tinnitus and hearing loss*, **European Archives of Oto-Rhino-Laryngology**, Vol. 271, Issue 8, pp. 2177-2184, August 2014.

[19] Y. Tian, T. Kanade, J. F. Cohn, *Recognizing Action Units for Facial Expression Analysis*, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol. 23, Issue 2, pp. 97-115, February 2001.

[20] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, *The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression*, **IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops**, June 2010.

[21] M.J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, *Coding Facial Expressions with Gabor Wavelets*, **Proceedings of the third IEEE International Conference on Automatic Face and Gesture Recognition**, April 1998.

[22] M. Pantic, M. Valstar, R. Rademaker, L. Maat, *Web-based database for facial expression analysis*, **IEEE International Conference on Multimedia and Expo**, July 2005.

[23] A. Dhall, r. Goecke, S. Lucey, T. Gedeon, *Collecting Large, Richly Annotated Facial-Expression Databases from Movies*, **IEEE MultiMedia**, Vol. 19, Issue 3, pp. 34-41, July 2012.

[24] I. Sneddon, M. McRorie, G. McKeown, J. Hanratty, *The Belfast Induced Natural Emotion Database*, **IEEE Transactions on Affective Computing**, Vol. 3, Issue 1, pp. 32-41, March 2012.

[25] J. Kossaifi, G. Tzimiropoulos, S. Todorovic, M. Pantic, *AFEW-VA database for valence and arousal estimation in-the-wild*, **Image and Vision Computing**, Vol. 65, pp. 23-36, September 2017.

[26] N. Sharma, A. Dhall, T. Gedeon, R. Goecke, *Thermal spatiotemporal data for stress recognition*, **EURASIP Journal on Image and Video Processing**, Volume 28, May 2014.

[27] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, M. Pantic, *AVEC 2014: 3D Dimensional Affect and Depression Recognition Challenge*, **Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge**, pp. 3-10, November 2014.

[28] B. Schuller, B. Vlasenko, F. Eyben, M. Wollmer, A. Stuhlzatz, A. Wendemuth, G. Rigoll, *Cross-Corpus Acoustic Emotion Recognition: Variances and Strategies*, **IEEE Transactions on Affective Computing**, Vol. 1, Issue 2, pp. 119-131, December 2010.

[29] O. Martin, I. Kotsia, B. Macq, I. Pitas, *The eNTERFACE' 05 Audio-Visual Emotion Database*, **22nd International Conference on Data Engineering Workshops**, April 2006.

[30] T. Balomenos, A. Raouzaiou, K. Karpouzis, S. Kollias, R. Cowie, *An Introduction to Emotionally Rich Man-Machine Intelligent Systems*, **Eunite**, 2003.

[31] S. Kim, M. Fillippone, F. Valente, A. Vinciarelli, *Predicting Continuous Conflict Perception with Bayesian Gaussian Processes*, **IEEE Transactions on Affective Computing**, Vol. 5, Issue 2, pp. 187-200, May 2014.
[32] J.H.L. Hansen, *Analysis and Compensation of Speech under Stress & Noise for Environmental Robustness in Speech Recognition*, **ESCA/NATO Tutorial and Research Workshop on speech Under Stress**, September 1995.
[33] L. Likforman-Sulem, A. Esposito, M. Faundez-Zanuy, S. Clemencon, G. Cordasco, *EMOTHAW: A Novel Database for Emotional State Recognition from Handwriting and Drawing*, **IEEE Transactions on Human-Machine Systems**, Vol. 47, Issue 2, pp. 273-284, April 2017.
[34] D. France, R. Shiavi, S. Silverman, M. Silverman, D. Wilkes, *Acoustical properties of speech as indicators of depression and suicidal risk*, **IEEE Transactions of Biomedical Engineering**, Vol. 47, Issue 7, pp. 829-837, 2000.
[35] A. Pampuchidou, K. Marias, M. Tsiknakis, P. Simos, F. Yang, G. Lemaitre, F. Meriaudeau, *Video-based depression detection using local curvelet binary patterns in pairwise orthogonal planes*, **Proceedings of annual International Conference of the IEEE Engineering in Medicine and Biology Society**, pp. 3835-3838, August 2016.
[36] G.B. Huang, *Learning capability and storage capacity of two-hidden-layer feedforward networks*, **IEEE Transactions on Neural Networks**, Vol. 14, Issue 2, pp. 274-281, March 2003.
[37] J. Li, J.H. Cheng, F. Huang, *Brief Introduction of Back Propagation BP) Neural Network Algorithm and Its Improvement*, **Advances In Computer Science and Information Engineering**, pp. 553-558, 2012.
[38] S. Masood, M.N. Doja, P. Chandra, *Analysis of weight initialization techniques for gradient descent*, **Annual IEEE India Conference**, December 2015.
[39] S. Kherchaoui, A. Houacine, *Face detection based on a model of the skin color with constraints and template matching*, **International Conference on Machine and Web Intelligence**, October 2010.
[40] I. Budiman, Herlianto, D. Suhartono, F. Purnomo, M. Shodiq, *The effective noise removal techniques and illumination effect in face recognition using Gabor and Non-Negative Matrix Factorization*, **International Conference on Informatics and Computing**, October 2016.
[41] P. Salembier, *Structuring element adaptation for morphological filters*, **Journal of Visual Communication and Image Representation**, Vol. 3, Issue 2, pp. 115-136, June 1992.
[42] A.M. Burton, K. Bindemann, *The role of view in human face detection*, **Vision Research**, Vol. 49, Issue 15, pp. 2026-2036, July 2009.
[43] G.J. Edwards, C.J. Taylor, T.F. Cootes, *Interpreting face images using active appearance models*, **Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition**, April 1998.
[44] G. Tzimiroupoulos, M. Pantic, *Fast Algorithms for Fitting Active Appearance Modes to Unconstrained Images*, **International Journal of Computer Vision**, Vol. 122, Issue 1, pp. 17-33, March 2017.
[45] F. Dornaika, J. Ahlberg, *Efficient active appearance model for real-time head and facial feature tracking*, **IEEE International SOI Conference**, October 2003.
[46] V.P. Kshirsagar, M.R. Baviskar, M.E. Gaikwad, *Face recognition using Eigenfaces*, **3rd International Conference on Computer Research and Development**, March 2011.
[47] C.T. Liao, Y.K. Wu, S.H. Lai, *Locating facial feature points using support vector machines*, **9th International Workshop on Cellular Neural Networks and Their Applications**, May 2005.
[48] M. Gavrilescu, *Proposed architecture of a fully integrated modular neural network-based automatic facial emotion recognition system based on Facial Action Coding System*, **International Conference on Communications (COMM)**, pp. 1-6, Bucharest, 29-31 May 2014.
[49] M. Gavrilescu, N. Vizireanu, *Neural Network-Based Architecture for Fatigue Detection Based on the Facial Action Coding System*, **International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures**, pp. 113-123, Bucharest, 12-14 October 2017.
[50] M. Gavrilescu, *Recognizing emotions from videos by studying facial expressions, body postures and hand gestures*, **23rd Telecommunications Forum Telfor**, pp. 720-723, Belgrade, 24-26 November 2015.
[51] M. Gavrilescu, *Recognizing human gestures in videos by modeling the mutual context of body position and hands movement*, **Multimedia Systems**, Vol. 23, Issue 3, pp. 381-393, June 2017.
[52] S. Chin, C.Y. Lee, J. Lee, *An automatic method for motion capture-based exaggeration of facial expressions with personality types*, **Virtual Reality**, Vol. 17, Issue 3, pp. 219-237, September 2013.
[53] A. D. Setyadi, T. Harsono, S. Wasista, *Human character recognition application based on facial feature using face detection*, **International Electronics Symposium**, September 2015.
[54] L. Teijeiro-Mosquera, J.I. Biel, J.L. Alba-Castro, D. Gatica-Perez, *What your face Vlogs about: Expressions of Emotions and Big-Five Traits Impressions in Youtube*, **IEEE Transactions on Affective Computing**, Vol. 6, Issue 2, pp. 193-205, December 2014.
[55] M. Gavrilescu, *Study on determining the Big-Five personality traits of an individual based on facial expressions*, **E-Health and Bioengineering Conference**, pp. 1-6, Iași, 19-21 November 2015.
[56] M. Gavrilescu, N. Vizireanu, *Predicting the Sixteen Personality Factors (16PF) of an individual by analyzing facial features*, **EURASIP Journal on Image and Video Processing**, Vol. 59, August 2017.
[57] J. Joshi, A. Dhall, R. Goecke, M. Breakspear, G. Parker, *Neural-net classification for spatiotemporal descriptor based depression analysis*, **Proceedings of the 21st International Conference on Pattern Recognition**, November 2012.
[58] M. Gavrilescu, N. Vizireanu, *Predicting Depression, Anxiety, and Stress Levels from Videos Using the Facial Action Coding System*, **Sensors**, Vol. 19, issue 17, pp. 3693, August 2019.
[59] W.J. Yoon, K.S. Park, *A study of Emotion Recognition and Its Applications***, International Conference on Modeling Decisions for Artificial Intelligence**, pp. 455-462, 2007.
[60] N. Upadhyay, A. Karmakar, *Speech Enhancement using Spectral Subtraction-type Algorithms: A comparison and simulation study*, **Procedia Computer Science**, Vol. 54, pp. 574-588, 2015.
[61] N.V. Prasad, S. Umesh, I*mproved cepstral mean and variance normalization using Bayesian framework*, **IEEE Workshop on Automatic Speech Recognition and Understanding**, December 2013.

[62] F. Nolan, *Intonational equivalence: An experimental evaluation of pitch scales*, **Proceedings of the 15th International Congress of Phonetic Sciences**, January 2003.
[63] M. Gavrilescu, *Improved automatic speech recognition system using sparse decomposition by basis pursuit with deep rectifier neural networks and compressed sensing recomposition of speech signals*, **International Conference on Communications (COMM)**, pp. 1-6, Bucharest, 25-27 June 2015.
[64] M. Gavrilescu, *Noise robust Automatic speech Recognition system by integrating Robust Principal Component Analysis (RPCA) and Exemplar-based Sparse Representation*, **7th International Conference on Electronics, Computers and Artificial Intelligence**, pp. S-29-S34, Bucharest, 25-27 June 2015.
[65] M. Gavrilescu, *Improved Automatic Speech Recognition System by using compressed sensing signal reconstruction based on L0 and L1 estimation algorithms*, **7th International Conference on Electronics, computers and Artifical Intelligence**, pp. S-23-S28, Bucharest, 25-27 June 2015.
[66] I. Guoth, M. Chmulik, J. Polacky, M. Kuba, *Two-dimensional cepstrum analysis approach in emotion recognition from speech*, **39th International Conference on Telecommunications and Signal Processing**, December 2016.
[67] H.K. Vydana, P. Vikash, T. Vamsi, K.P. Kumar, A.K. Vuppala, *Detection of emotionally significant regions of speech for emotion recognition*, **Annual IEEE India Conference**, December 2015.
[68] S.G. Koolagudi, K.S. Rao, *Real life emotion classification using VOP and pitch based spectral features*, **Annual IEEE India Conference**, February 2011.
[69] L. Fu, C. Wang, Y. Zhang, *Classifier fusion for speech emotion recognition*, **IEEE International Conference on Intelligent Computing and Intelligent Systems**, December 2010.
[70] M.T. Shami, M.S. Kamel, *Segment-based approach to the recognition of emotions in speech*, **IEEE International Conference on Multimedia and Expo**, October 2005.
[71] N. Ding, N. Ye, H. Huang, R. Wang, R. Malekian, *Speech emotion features selection based on BBO-SVM*, **10th International Conference on Advanced Computational Intelligence**, June 2018.
[72] Y. Kan, M. Xu, Z. Wu, L. Cai, *Automatic Emotion Variation Detection in continuous speech*, **Signal and Information Processing Association Annual Summit and Conference**, February 2015.
[73] M. Gavrilescu, N. Vizireanu, *Feedforward Neural Network-Based Architecture for Predicting Emotions from Speech*, **Data,** Vol. 4, Issue 3, pp. 101, July 2019.
[74] M. Fallahnezhad, M. Vali, M. Khalili, *Automatic Personality Recognition from reading text speech*, **Iranian Conference on Electrical Engineering**, July 2017.
[75] M.H. Su, C.H. Wu, K.Y. Huang, Q.B. Hong, H.M. Wang, *Personality trait perception from speech signals using multiresolution analysis and convolutional neural networks*, **Asia-Pacific Signal and Information Processing Association Annual Summit and Conference**, February 2018.
[76] S. Jothilakshmi, R. Brindha, *Speaker trait prediction for automatic personality perception using frequency domain linear prediction features*, **International Conference on Wireless Communications, Signal Processing and Networking**, September 2016.
[77] V. Mitra, A. Tsiartas, E. Shirberg, *Noise and reverberation effects on depression detection from speech*, **IEEE International Conference on Acoustics, Speech and Signal Processing**, May 2016.
[78] L. He, C. Cao, *Automatic depression analysis using convolutional neural networks from speech*, **Journal of Biomedical Informatics**, Vol 83, pp. 103-111, 2018.
[79] L. Guarnera, G.M. Farinella, A. Furnari, A. Salici, C. Ciampini, V. Mantraga, S. Battiato, *GRAPHJ: A forensic tool for handwriting analysis*, **19th International Conference on Image Analysis and Processing**, September 2017.
[80] M.M. Prunty, A.L. Barnett, K. Wilmut, M.S. Plumb, *An examination of writing pauses in the handwriting of children with developmental coordination disorder*, **Research in Developmental Disabilities**, Vol. 35, Issue 11, pp. 2894-2905, November 2014.
[81] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, A.S. di Freca, *Handwriting analysis to support neurodegenerative diseases diagnosis: A review*, **Pattern Recognition Letters**, May 2018.
[82] R.N. Morris, *Forensic Handwriting Identification: Fundamental Concepts and Principles*, 2000.
[83] W.L. Lee, K.C. Fan, *Document Image preprocessing based on optimal Boolean filters*, **Signal Processing**, Vol. 80, Issue 1, pp. 45-50, 2000.
[84] S.C.F. Lin, C.Y. Wong, G. Jiang, M.A. Rahman, T.R. Ren, N. Kwok, H. Shi, y.H. Yu, T. Wu, *Intensity and edge-based adaptive unsharp masking filter for color image enhancement*, **Optik - International Journal for Light and Electron Optics**, Vol. 127, Issue 1, pp. 407-414, January 2016.
[85] R. Legault, C.Y. Suen, *Optimal local weighted averaging methods in contour smoothing*, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol. 18, pp. 690-706, July 1997.
[86] Y. Solihin, C.G. Leedham, *Integral ratio: a new class of global thresholding techniques for handwriting images*, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol. 21, pp. 761-768, August 1999.
[87] K. Chen, F. Yin, C.L. Liu, *Hybrid page segmentation with efficient whitespace rectangles extraction and grouping*, **12th International Conference on Document Analysis and Recognition**, pp. 958-962, 2013.
[88] V. Papavassiliou, T. Stafylakis, V. Katsouro, G. Carayannis, *Handwritten document image segmentation into text lines and words*, **Pattern Recognition**, Vol. 43, pp. 369-377, 2010.
[89] M.B. Menhaj, F. Razazi, *A new fuzzy character segmentation algorithm for Persian/Arabic typed texts*, **International Conference on Computational Intelligence**, pp. 151-158, 1999.
[90] R. Coll, A. Fornes, J. Llados, *Graphological analysis of handwritten text documents for human resources recruitment*, **12th International Conference on Document Analysis and Recognition**, pp. 1081-1085, July 2009.
[91] L. Deng, *The MNIST database of handwritten digit images for machine learning research*, **IEEE Signal Processing Magazine**, Vol. 29, Issue 6, pp. 141-142, November 2012.
[92] O. Golubitsky, S.M. Watt, *Distance-based classification of handwritten symbols*, **International Journal on Document Analysis and Recognition**, Vol. 13, Issue 2, pp. 133-146, June 2010.

[93] M. Gavrilescu, N. Vizireanu, *Using Off-line handwriting to predict blood pressure level: a neural network-based approach*, **International Conference on Future Access Enablers of Ubiquitous and Intelligent Infrastructures**, pp. 124-133, Bucharest, 12-14 October 2017.
[94] Z. Chen, T. Lin, *Automatic personality identification using writing behaviors: an exploratory study*, **Behavior and Information Technology**, Vol. 36, Issue 8, pp. 839-845, 2017.
[95] M. Gavrilescu, N. Vizireanu, *Predicting the Big Five personality traits from handwriting*, **EURASIP Journal on Image and Video Processing,** Vol. 57, December 2018.
[96] M. Gavrilescu, *Study on determining the Myers-Briggs personality type based on individual's handwriting*, **E-Health and Bioengineering Conference**, pp. 1-6, Iași, 19-21 November 2015.
[97] M. Gavrilescu, *3-Layer Architecture for determining the personality type from handwriting analysis by combining neural networks and support vector machines*, **UPB Scientific Bulletin**, Series C, Vol. 79, Issue 4, pp. 135-150, 2017.
[98] E.S. Friedman, D.B. Clark, S. Gershon, *Stress, anxiety, and depression: Review of biological, diagnostic, and nosologic issues*, **Journal of Anxiety Disorders**, Vol. 6, Issue 4, pp. 337-363, December 1992.
[99] A.S. Patwardhan, *Multimodal mixed emotion detection*, **2nd International Conference on Communication and Electronics Systems (ICCES)**, October 2017.
[100] S. Zhalehpour, Z. Akhtar, C.E. Erdem, *Multimodal emotion recognition with automatic peak frame selection*, **IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA) Proceedings**, June 2014.
[101] J. Xue, Z. Luo, K. Eguchi, T. Takiguchi, T. Omoto, *A Bayesian nonparametric multimodal data modeling framework for video emotion recognition*, **IEEE International Conference on Multimedia and Expo**, July 2017.
[102] P. Tzirakis, G. Trigeorgis, M.A. Nicolau, B.W. Schuller, S. Zafeiriou, *End-to-End multimodal emotion recognition using deep neural networks*, **IEEE Journal of Selected Topics in Signal Processing**, Vol. 11, Issue 8, pp. 1301-1309, October 2017.
[103] R. Gajsek, V. Struc, F. Mihelic, *Multimodal emotion recognition using canonical correlations and acoustic features*, **20th International Conference on Pattern Recognition**, August 2010.
[104] Z. Zhang, F. Ringeval, B. Dong, E. Coutinho, E. Marchi, B. Schuller, *Enhanced semi-supervised learning for multimodal emotion recognition*, **IEEE International Conference on Acoustics, Speech and Signal Processing**, May 2016.
[105] Z. Xie, Y. Tie, L. Guan, *A new audiovisual emotion recognition system using entropy-estimation-based multimodal information fusion*, **IEEE International Symposium on Circuits and Systems**, May 2015.
[106] A. Agrawal, N.K. Mishra, *Fusion Based Emotion Recognition System*, **International Conference on Computational Science and Computational Intelligence**, December 2016.
[107] D. Nguyen, K. Nguyen, S. Sridharan, D. Dean, C. Fookes, *Deep spatiotemporal fusion with compact bilinear pooling for multimodal emotion recognition*, **Computer Vision and Image Understanding**, July 2018.
[108] S. Poria, H. Peng, A. Hussain, N. Howard, E. Cambria, *Ensemble Application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis*, **Neurocomputing**, Vol. 261, pp. 217-230, October 2017.
[109] G. Cariadakis, K. Karpouzis, M. Wallace, L. Kessous, N. Amir, *Multimodal user's affective state analysis in naturalistic interaction*, **Journal on Multimodal User Interfaces**, Vol. 3, Issue 1-2, pp. 49-66, March 2010.
[110] L. Batrinca, N. Mana, B. Lepri, F. Pianesi, *Multimodal Personality Recognition in Collaborative Goal-Oriented Tasks*, **IEEE Transactions on Multimedia**, Vol. 18, Issue 4, pp. 659-673.
[111] Y. Guckuturk, U. Gucku, M. Perez, H.J. Escalante, X. Baro, C. Andujar, I. Guyon, J.J. Junior, M. Madadi, S. Escalera, M.A.J. van Gerven, R. van Lier, *Visualizing apparent personality analysis with deep residual networks*, **IEEE International Conference on Computer Vision Workshops**, October 2017.
[112] J. Gorbova, I. Lusi, A. Litvin, G. Anbarjafari, *Automated Screening of Job Candidate Based on Multimodal Video Processing*, **IEEE Conference on Computer Vision and Pattern Recognition Workshops**, July 2017.
[113] H. Dibeklioglu, Z. Hammal, J.F. Cohn, *Dynamic Multimodal Measurement of Depression Severity using deep autoencoding*, **IEEE Journal of Biomedical and Health Informatics**, Vol. 22, Issue 2, pp. 525-536, March 2018.
[114] L. He, D. Jiang, H. Sahli, *Multimodal depression recognition with dynamic visual and audio cues*, **International Conference on Affective Computing and Intelligent Interaction**, December 2015.
[115] G. Cariadakis, K. Karpouzis, M. Wallace, L. Kessous, N. Amir, *Multimodal user's affective state analysis in naturalistic interaction*, **Journal on Multimodal User Interfaces**, Vol. 3, Issue 1-2, pp. 49-66, March 2010.
[116] A. Landowska, *Affective Computing and affective learning - methods, tools and prospects*, **EduAkcja. Magazyn edukacji elektronicznej**, Vol. 1, Issue 5, pp. 16-31, 2013.
[117] M. Gavrilescu, *Study on using individual differences in facial expressions for a face recognition system immune to spoofing attacks*, **IET Biometrics**, Vol. 5, Issue 3, pp. 236-242, September 2016.
[118] J.M. Girard, J. F. Cohn, M. H. Mahoor, S. Mavadati, D. P. Rosenwald, *Social Risk and Depression: Evidence from Manual and Automatic Facial Expression analysis*, **Proceedings of International Conference on Automatic Face and Gesture Recognition**, pp. 1-8, 2013.
[119] E. L. Rosenberg, P. Ekman, W. Jiang, M. Babyak, R. Coleman, M. Hanson, C. O'Connor, R. Waugh, J.A. Blumenthal, *Linkages between facial expressions of anger and transient myocardial ischemia in men with coronary artery disease*, **Emotion**, Vol. 1, Issue 2, pp.07-115, June 2001.
[120] T. Guha, Z. Yang, A. Ramakrishna, R. B. Grossman, H. Darren, S. Lee, S.S. Narayanan, *On quantifying facial expression-related atypicality of children with Autism Spectrum Disorder*, **Proceedings of IEEE International Conference on Acoustics and Speech Signal Processing**, pp. 803-807, 2015.
[121] A. Kolakowska, a. Landowska, M. Szwoch, M. Wrobel, *Emotion Recognition and Its Applications*, **Advances in Intelligent Systems and Computing**, Vol. 300, pp. 51-62, July 2014.