Received October 21, 2019, accepted November 1, 2019, date of publication November 8, 2019, date of current version November 21, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2952313

Emotional State Prediction From Online Handwriting and Signature Biometrics

YASEMIN BAY AYZEREN [1], MERYEM ERBILEK [2], AND ERBUĞ ÇELEBI [1]
[1]Management Information Systems Department, Cyprus International University, 99258 Nicosia, Cyprus
[2]Computer Engineering Department, Cyprus International University, 99258 Nicosia, Cyprus

Corresponding author: Yasemin Bay Ayzeren (yaseminbay@ciu.edu.tr)

ABSTRACT Handwriting and signature biometrics have a long history in the literature, especially in terms of identity recognition and/or verification; nevertheless, it reveals more information therefore provides more opportunities for personal characteristics estimation, particularly, emotional state. However, almost all publicly or commercially available databases do not include contributors' demographic labels or emotional status in addition to their identification labels, leading these available databases to be only useful for verification and identification based research studies. For this reason, this paper proposes both offline and online handwriting and signature biometric database with a wide range of ground truths (emotional status labels – happy, sad and stress) in addition to the identity labels and tries to predict ones' emotional state, namely happy, sad and stress, from their online biometric handwriting and signatures. The proposed database comprised total of 134 participants with 804 handwriting and 8040 signature biometric samples. The database presented also includes individuals' demographic information such as age, gender, handedness, education level and nationality. Subsequently, there are several experiments have been conducted, with different thresholds, which present the usability of the proposed database and preliminary results of emotional state prediction from both signature and handwriting biometrics. The experiments achieved remarkable success especially on stress prediction for handwriting. On the other hand, considering the results from signature biometrics, it is observed that happy and then sad and finally stress class forms the most part of the prediction accuracy.

INDEX TERMS Soft-biometrics, handwriting biometrics, signature biometrics, biometric database, emotion prediction.

I. INTRODUCTION

Biometrics is a generic term to reliably identify and/or verify an individual by recognizing the unique physiological (such as face, iris, fingerprint, ear and so on) and/or behavioral (such as handwriting, signature, gait, voice, keystroke and so on) traits that cannot be lost, forgotten, or used by others, as it is often the case with traditional ID (Identity) and password authentication systems [1]. Biometrics is helping reshape industries by relieving businesses' concerns about the security and privacy of their confidential information. This field is therefore gaining stable growth of attention from researchers bringing new features into the identification and authentication processes. Biometric identification has therefore been adopted in security applications (such as airborne, home automation, health care etc.) due to its benefits, such as robustness, uniqueness, collectability, availability and permanence [2].

Soft-biometrics on the other hand identify and/or verify an individual's non-unique characteristics (such as age, gender, ethnicity, and so on) – rather than identity – by using physiological (such as face, iris, fingerprint and so on) and/or behavioral (such as handwriting, signature, gait, voice, keystroke and so on) traits [3]. Increasing importance for all types of soft biometric prediction as a practical applications of technologies become more diverse and widespread, is the issue of the providing important information about the individual for the forensic scenarios – when specific identity cannot be provided [4], elimination of candidate individuals from consideration in an identification scenario to decrease the search space [1], in health care applications playing a role in early diagnosis of illnesses [3] and so on.

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang.

TABLE 1. Biometric databases in soft-biometrics.

Biometric Modality	Database Name	Year	Ref.	No. of subjects	Label	Availability
Face	FRGC	2005	[16]	688	race, age, gender	publicly
	BOSPHORUS	2008	[17]	105	facial expression, gender	publicly
	UTKFace	2017	[18]		age, gender, ethnicity	publicly
	Cohn-Kanade	2010	[19]	100	facial expression	publicly
	FERET	2000	[20]	1199	age, gender, ethnicity	publicly
	BioSecure	2010	[21]	600	age, gender, handedness, visual aids	publicly
Iris	CASIA	2009	[22]	50	age	publicly
	UBIRIS.v2	2010	[23]	261	gender, age	publicly
Fingerprint	CASIA	2009	[22]	49	age	publicly
	-	2006	[24]	2200	gender	not available
	-	2013	[25]	547	gender	not available
Gait	CASIA	2009	[22]	124	age, gender	publicly
	USF Human ID	2005	[26]	74	gender	publicly
Voice	N-Best	2009	[27]	555	age, gender	publicly
	TIMIT	1993	[28]	76	gender	commercially
	-	2004	[29]	10	emotion	not available
	FARSDAT	1994	[30]	304	age	publicly
Hand	-	2008	[31]	50	gender	not available

Furthermore, soft-biometrics provide predictive capabilities to higher levels of abstraction such as mental and emotional state of individuals (i.e. happy or sad, under stress or relaxed, and so on) to support applications mentioned above or analyzing individuals' activities [5]. Emotional state prediction enlightens many areas of the literature, from career counselling [6], [7], to measuring user interaction through entertainment technology [8], from early stress recognition for office environment [6] to early prediction of neurodegenerative diseases such as Alzheimer, mild cognitive impairment and dementia [9], as well as emotion prediction from keystroke dynamics [5] and many more. In addition, predicting mental and emotional state of individuals plays a particularly important role in security and defense systems, especially given the previous airborne terrorist attacks: this is clear evidence of the failure of early detection systems in air transportation. Although human emotions are apparently invisible mental states, behavioral biometrics aims to scan and reveal intentions, therefore there are many research studies examining how technology can capitalize based on the known behavioral responses that a person with malicious intent will exhibit for instantaneous recognition. Unfortunately, there are no automatic and inconspicuous early emotion detection systems that could improve the decision making process in surveillance operations, nevertheless there are several organizations that employ automated border crossing systems, such as Frontex (technical and operational) [10], International Civil Aviation Organization [11], the International Air Transport Association [12], and the European Committee for Standardization [13], who all use biometrics to authenticate travelers. EU first introduced Shengen Information System (SIS) in 1995 and its improved version is in action since 2018 which is called as Central Schengen Information System (CS-SIS) Automatic Fingerprint Identification System (AFIS). New investigations are in progress for the readiness and implementation of Automatic Biometric Identification System (ABIS) fingermark and palm mark technologies for their introduction in the CS-SIS [64]. Although, these implementations and studies are all based on the identification of the visa applicant, however, during visa application evaluation those biometric signatures and palmprints can be used to extract emotional state of the applicant to detect whether they are happy to travel or not.

In order to analyze, develop and improve science of biometrics and soft-biometrics, a fundamental requirement for the research studies is the database availability with biometric data and corresponding data labels (i.e. relevant ground truth information). Although, the availability of biometric database opportunity decreases for the soft-biometric area, still there are considerable amount of publicly and commercially available databases for many physiological (i.e. face, iris, fingerprint, etc.) and behavioral (i.e. signature, handwriting, gait, voice etc.) biometric modalities as presented in Table 1. However, considering the higher prediction level of the soft biometric research, which is the prediction of mental and emotional state of individuals, availability of databases decreases even more considerably. For instance, such a database is publicly and commercially available for face, iris, fingerprint, gait and voice modalities which can be seen in Table 1, and only one private English database (available for the use of established institution only) is existing for handwriting [14] and keystroke [5] modalities.

In the past, many researchers and institutions have joined their effort to establish biometric and multi-biometric databases [15]–[19] regarding to the unimodal and multimodal research and applications. However, a database with a biometric/multi-biometric data and a corresponding several data labels (i.e. relevant ground truth information) is still missing. Hence, nowadays lack of this kind of database is the main barrier regarding the improvement and possible developments of the biometrics (including unimodal and multimodal) and soft-biometrics research area, and also analysis of the relationship between them. As can be noticed from the above discussion, main absence regarding to the availability of databases including biometric data and a corresponding several data labels (i.e. relevant ground truth information) is based on handwritten and signature data.

The main reason behind this could be handwritten data is not frequently used for communication nowadays. However, it is still very important since both handwriting and signature is for instance widely used in digital form for determination of identity and other characteristics in forensic investigations [35]–[37] and healthcare scenarios (for instance, management of Parkinson's diseases, assessment of children with developmental coordination disorder) [5], [9], [18]. Furthermore, predicting one's emotion from handwriting and signature can even support the officers during the visa application process to detect whether the applicant is happy, sad or stressed, and a decision can be made by also taking in consideration their current emotions.

Handwriting and signature have been recognized as behavioral biometric modalities by the research community, are related to the motor skills, and reveal some characteristic information about the individual. Although every child in primary school learns how to write in a specific pattern and it becomes a motor skill as they mature but they can control what to write but not how to write. Therefore, analysis of one's handwriting and signature can reveal some information about the writer, such as behavioral traits, character or personality [38]. The handwriting and signature of individual can be measured and analyzed, therefore its measurability gives rise to opportunities for researchers to identify the identity of an individual as well as to predict their soft-biometric treats such as emotions [35], gender [39], age group [40], and assess the effects of alcohol [41].

Hence, this paper presents the recently acquired Handwritten Database with Multi-Labels – which includes both handwriting and signature data and all necessary soft-biometric data labels – which was collected at CIU (Cyprus International University) in 2017. In total, 134 individuals have been acquired in an indoor environment over one acquisition session for both offline and online handwriting and signature biometric data. The main novelties of this paper can be summarized as follows;

- Detailed overview of handwriting and signature biometric databases for soft-biometric research.
- Detailed description of the first online and offline handwriting and signature biometric data collected under different writing scenarios (fixed and variable task), physiological scenarios (before and after watching positive and negative videos) and with a corresponding several data labels (i.e. relevant ground truth information including identity, age, gender, ethnicity and emotion).
- Showing that online handwriting and signature biometric data includes information (i.e soft-biometrics) in addition to its owners' identity information.
- Showing that higher-level characteristics, particularly, emotional state estimation is possible to predict through ones' handwriting and signature.

The rest of the paper is organized as follows; Section II, will present extensive review of the available databases on handwriting and signature biometrics as well as their relation and differences to the proposed database. Subsequently, Section III will discuss the details of the data collection procedure of the proposed database. Finally, Section IV will focus on the experimental analysis and present first preliminary results which is followed by Section V with conclusion.

II. BACKGROUND

In this section, other reported works related to this study – the handwriting and signature biometric databases with corresponding emotional state of individuals labels – will be described in detailed. This documented survey will allows us to recognize the importance of the proposed database for the future development of the research studies in a wider context. Before all, it is useful to understand the background information (such as representation and text-content) of handwriting and signature biometric data.

Handwriting and signature biometrics can be represented in two ways: off-line and on-line. These differ in the type of information they employ. The input of the off-line systems is an image of a written text; the image is pre-processed by grey scaling, removing noise and segmenting characters and words. The input is treated as pixel patterns or textures, so that different features can be extracted for processing. On-line systems, on the other hand, take the input through an electronic pen used during the handwriting process. The input of the systems is a variety of dynamic features typical of the handwriting, such as X and Y coordinates, pen pressure, pen orientation, timestamp etc. Due to gathering input data during the writing process, on-line systems generally perform better than their off-line counterparts [42]. In our proposed database, we acquired data from participants in both dynamic and static representations.

When the focus of the attention is the handwriting biometrics then the text content becomes an important issue, where another decision that needs to be made in regards to processing is whether to use text-dependent (fixed) or text-independent (variable) content; these can be applied individually or integrally, depending on the available text content to perform the recognition. The text-dependent method requires all the participants to write a predefined text. Contrarily, in the text-independent method, participants can write any text regardless of the content. In our proposed database, we acquired data from participants in both fixed and variable tasks.

The detailed survey based on handwriting and signature biometric databases with corresponding data labels is shown

TABLE 2. Databases related to handwriting and signature biometrics.

Biometric Trait	Database	Year	Ref.	Language	No. of subjects	Label	Content	Online/ Offline	Availability
Handwriting	EMOTHAW	2017	[32]	Italian	129	Id, emotion	Text-dependent	Online	publicly
	-	2015	[15]	English	100	Id, gender, age, emotion	Both	Both	Not available
	CASIA	2015	[22]	Chinese, English	187, 134	Id	Text-independent	Both	publicly
	AltecOnDB	2014	[46]	Arabic	1001	Id, name, gender, age, handedness	Text-independent	Online	publicly
	ADAB	2009	[47]	Arabic	173	Id	Text-dependent	Online	publicly
	BiosecurID	2007	[36]	Spanish	400	Id, gender, age, handedness	Text-dependent	Both	publicly
	MyIDEA	2005	[35]	English	104	Id	Text-dependent	Both	Commercially
	IfN/ENIT	2002	[48]	Arabic	411	Id	Text-dependent	Both	publicly
	IRONOFF	1999	[49]	French, English		Id	Text-dependent	Both	publicly
	UNIPEN	1994	[33]	English	215	Id	Text-independent	Online	publicly
	CEDAR	1994	[50]	English	200	Id	Text-dependent	Both	publicly
Signature	SCUT-MMSIG	2017	[51]	-	150	Id	-	Online	Not available
	SUSIG	2009	[52]	-	100	Id	-	Online	publicly
	BioSecure	2008	[21]	-	713	Id	-	Both	publicly
	MyIDEA	2005	[35]	-	104	Id	-	Both	Commercially
	SVC	2004	[53]	-	100	Id	-	Online	publicly
	MCYT	2003	[54]	-	330	Id	-	Both	publicly
	BIOMET	2003	[34]	-	327	Id	-	Online	publicly

in Table 2. Important points based on the survey presented in Table 2 can be summarised as follows;

- Considering emotion label point of view, there is only one publicly available database based on handwriting biometric data (where the language is Italian) with corresponding emotion labels, and even more interestingly there is not any available database related to the signature biometric with corresponding emotion labels.
- Considering other labels point of view, there are only two publicly available databases based on handwriting biometric data with corresponding soft-biometric labels. However, none of these databases language is based on English Language. Similarly, there is not any available database related to the signature biometric with corresponding soft-biometric labels.

Hence, in order to overcome these problems and allow the both biometric and soft-biometric research community to shed more light on the signature and handwriting biometric investigations and applications, in this paper a database both for signature and handwriting data is proposed which includes the following features;

- All necessary biometric data labels including ID, age, gender, handedness, emotion, nationality and education level.
- Biometric data representation both in online and offline versions.
- Biometric data representation both in fixed and variable content versions.
- Biometric data representation in English Language.
- Biometric data representation and labels from sufficient amount of subjects with sufficient amount of samples per each subject.
- A biometric database with all features mentioned above which is publicly available by the research community and provide base for comparison of proposed techniques and therefore improvement of biometric area.

Furthermore, this study will focus on emotion prediction from handwriting and signature biometric data to perform preliminary experimental analysis using our proposed database, since as will be seen there is very limited work in this area. Hence, first it is also important to review the literature and also several other relevant studies related to handwriting and signature emotion prediction analysis more generally.

One of the preliminary studies demonstrated a viable possibility of emotional state prediction using handwriting [35], in which 100 participants performed 4 tasks to shape the database. Their initial aim was to classify participants as happy/unhappy and stressed/not stressed. Based on 12 features, common in the literature of handwritten signature, prediction evaluations are performed using SVM (Support Vector Machine), K-NN (K-Nearest Neighbor) and Jrip (Optimized Incremental Reduced Error Pruning). Highest accuracies are achieved with the SVM classifier for emotional states, happy (80%) and stressed (70%). Although this study is taken as a baseline for our proposed database, to be able to have penetrative change on the emotional state of the participants, between each task we applied a stronger method: media. The visual and auditory elements of the video increased the effectiveness of each task and the participants in the desired state of feeling were consequently ensured with a stronger grip. This study is later extended and represented by [14].

In the work reported in [18], a new database for emotional state recognition called EMOTHAW (Emotion Recognition from Handwriting and Drawing) for emotional state recognition is presented; it consists of 129 participants' handwritings in Italian language and drawing focusing on the detection of emotional disorders and negative emotions such as anxiety, depression and stress. Their empirical study shows that drawing reveals more cues on depression recognition whereas anxiety and stress is better recognized through both writing and drawing cues. The data gathered during the 7 tasks is: pentagons and house drawings, words copied in handprint, circles, clock drawings and one sentence copied in cursive writing using a digitizing tablet. They also concluded that both features, in-air and on-paper, are useful and are listed in the top 3 after applying the feature selection.

A method has been proposed by [53] to predict a person's personality by exploring various offline handwriting features. They focused on computer assisted handwriting analysis and considered 9 different features such as the size and slant of the words, space between words, breaks in writing, pressure, margin, baseline, loop of 'e' and distance between the title (dot) and the stem of 'i's. The experimentation is conducted on around 80 samples and the results are compared with the manual analysis obtained by a graphologist. They achieved 93.77% estimation accuracy. For each feature, they also estimated the behavioral accuracy, where 93.84% is reached for the baseline feature, with its corresponding behavioral explanation being negative.

Another study [38], is performed to identify people who are emotionally disturbed or depressed, and need psychological help. For this empirical study, offline handwriting features such as baseline, slant, pen-pressure, size, margin and zone is extracted from the scanned handwriting data. It has been stated that the proposed computerized system is more accurate than a human graphologist.

An expert system to detect student's exam anxiety from their offline handwriting samples has developed and presented in [65]. Authors collected 210 handwriting samples in two sessions; one month before the exam and during the exam day. The features which are used in this study are character's height, width and line slant. 66,6% classification accuracy is achieved and it has been concluded that an online system would be more accurate compared to offline.

One of the most recent works conducted by [66] is also aiming to extract emotion from handwriting. Hence, no attempt have been made for signature. Their work achieved 66% accuracy with user independent Support Vector Classifier for 1 in 4 classification (1. High Valence, High Arousal, 2. High Valence, Low Arousal, 3. Low Valence High Arousal and 4. Low Valence Low Arousal). The classification accuracy increased up to 70% with user dependent classifier. The proposed work illustrate how a system can recognize one's emotional state from their handwritten text however with a very limited (12) participants when compared to the data presented in this work. To be able to achieve adequate and qualified classification performance and extract useful information 134 subjects' handwriting and signatures have been collected in this study.

In the work reported in [54] an Analyzer, a real-time handwriting analysis system which is writer independent, has been developed to predict the personality of an individual by his/her writing samples. The experiment shows that a female writer writes longer sentences in comparison to a male. The slope of words is however similar. The slant of male handwriting is straight whereas female writing leans towards the right. The pen pressure of a male is scale whereas a female apply more pressure. With all these features they concluded that a male writer has a technical mind and use it time by time. However, the female writer has an artistic aptitude and uses their time.

On an empirical study demonstrated by [55], they tried to expose the relationship between signature and personality traits using business managers' signatures who sign very often in their busy schedules. They tried to explain the correlation between the fractal dimension of the signatures and the results of the personality tests. They concluded that with a result of 0.05 significance level there is no relationship exists between signature and personality characteristics using graphology method with offline features. The literature based on handwriting and signature emotion analysis summarized in Table 3.

When the focus of attention is the signature biometrics, it becomes even more difficult to find research studies in the literature. For instance, to our knowledge, the only experimental study is [67] which depicts the personality trait prediction based on the features such as underscores below signature, appearance of dot on the letter, curved start, ending stroke, and streaks disconnected and achieved accuracy rates 100%, 95%, 94%, 96% and 92% respectively.

As can be noticed from the above brief survey, there are very limited work related to the emotion prediction from

TABLE 3. Surveys and studies in handwriting and signature analysis.

Field	References	Uses	Features
Biometrics	[32], [10], [41], [8], [7],	Verification, authentication and identification of one's behavioural traits using their handwriting, monitoring health conditions for diagnosis, gender detection, career counselling etc.	Online: X and Y coordinates, pen pressure, pen orientation, timestamp etc. Offline: size, slant, space between words, breaks in writing, pressure, margin, baseline, loop of characters etc.
Forensic Studies	[37], [38], [39], [66],	Police investigations, security agencies and detectives by lawyers in court as evidence material.	Online: X and Y coordinates, pen pressure, pen orientation, timestamp etc. Offline: size, slant, space between words, breaks in writing, pressure, margin, baseline, loop of characters etc.
Graphology	[38], [55], [56], [65], [66], [67]	job applications, recruitment, marriage compatibility, career guidance, motivate workers, and student's exam anxiety, child development.	Offline: baseline alignment, slant, stroke, margin width, spacing between letters/words, text size, height of letters etc.
Palaeography	[58]	assign a scribe, a geographical location or a time period to a piece of handwritten artefact.	Offline: habitual pen grip; slant, curvature, roundness, and personalized set of letter shapes; ink blob shapes, allographs, that a writer has learned to use under educational, cultural and memetic influences.

online handwriting and signature biometric data. This paper will also present preliminary experimental analysis related to the emotion prediction based on both handwriting and signature biometric data to shed more light to this area.

III. THE CIU HANDWRITTEN DATABASE

In this section, the data acquisition system, participants and their demographic information, and the data collection methodology of the CIU Handwritten Database which has been conducted by the CIU in North Cyprus will be documented in detail step by step.

A. DATA ACQUISITION SYSTEM

The data for both handwriting and signature biometric acquisition has been conducted in a homogeneous environment using state-of-the-art pressure sensitive tablet called Wacom Intuos Pro – large (19''×12.5'') and a special writing device called Intuos Inkpen (Wacom Art Pen - KP701E2) [57]. A4 paper overlay is also used to provide visual feedback to contributors since visual and blind corpuses reveal some recognition accuracy differences [58] and makes contributors feel more comfortable while writing or signing.

Other than online samples stored in a text file, offline samples also exists and are saved in jpg format to give researchers more flexibility to work on the biometric data, which is the one of the main strengths of the collected database. More detailed description regarding to these will be given in Subsection C.

B. PARTICIPANTS AND DEMOGRAPHIC INFORMATION

Before collecting any data from participants, informed consent was obtained from all participants to ensure that accepted principles of ethical and professional codes of conduct are implemented.

The database is constructed by 134 volunteers, all from CUI in North Cyprus. Each participant generated 6 handwritten texts and 60 signatures. Coarsely, the database consists of 804 handwritten texts and 8040 signature samples in total. Detailed explanation and description of context of the collected handwriting and signature biometric data will be presented in Subsection C.

The other demographic information each contributor provided are; gender, age (each fall into three different age groups: 53 people under 25, 77 between 25-45 and 4 46 or older), handedness, education level and nationality (contributors from 22 different nations). Participants' demographic information distributions presented Fig. 1.

It is important to mention that the age groups defined above is only for the ease of presentation for the reader. Exact date of birth for each individual is available and stored in the database.

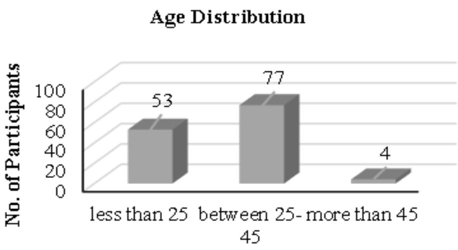
C. DATA COLLECTION METHODOLOGY

The handwriting biometric data has been collected in total of four different tasks using the defined data collection system in section IV; two text-dependent (i.e. fixed Task 1 and Task 4) and two text-independent (variable Task 2 and Task 3).

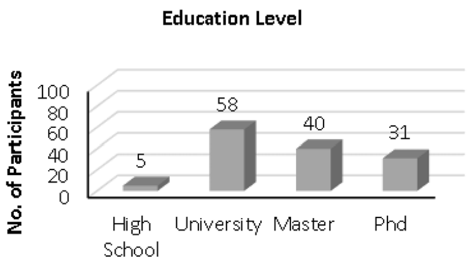
First task was aiming contributors to exhibit their natural handwriting and signature, and the remaining three tasks on the other hand, aiming to change their emotional state. The tasks each participant needs to carry out after providing their demographic information and signing the consent form, are summarised as:

Task 1: Participants were asked to copy a predefined text consisting of 43 words and 243 characters which provide an approximate distribution of the most common digraphs and trigraphs in English language. The predefined text chosen which provides a rich handwritten data referenced from [5].

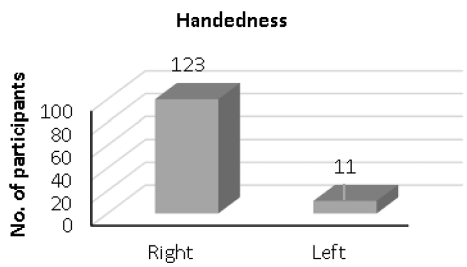
Task 2: Participants were asked to watch a media (video) and write some text with their own cues. This visual stimulus



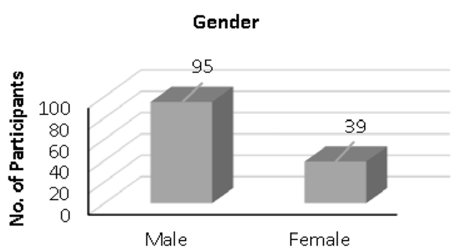
(a)



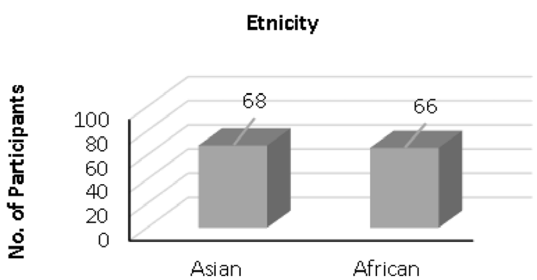
(b)



(c)



(d)



(e)

FIGURE 1. **Participants' demographic information distribution: (a) age, (b) education level, (c) handedness, (d) gender and (e) ethnicity distribution of participants.**

consists of components designed to convey a positive and often 'happy' message.

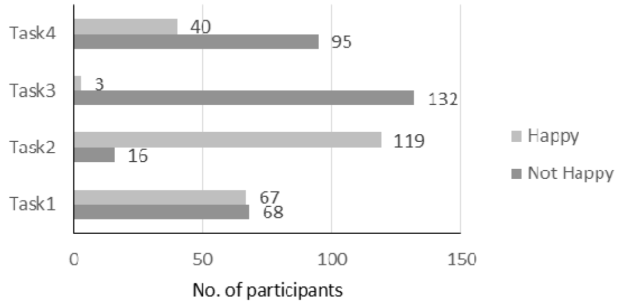


FIGURE 2. **The score distribution of happy emotion recorded in each task.**

Task 3: Same process as in Task2 is performed, except the video aiming to convey a 'sad' message.

Task 4: The previous tasks had no time limit. Contrary to this, in task 4 the participants were asked to copy a predefined text (10 words and 50 characters) within a maximum of 10 seconds. During this time, the countdown timer was a remarkable factor. The predefined text used is chosen to cover all the letters in the English alphabet is referenced from the same work as in Task 1.

The order of the tasks is another factor considered seriously in the data collection process. Assuming the majority of the participants started the data collection process as natural (not very happy, sad or stressed), we aimed to make them happier in the second task. With the media we use in Task 3, we aimed to make a dramatic change on their feelings (as opposed to Task 2) and tried to make them feel sad or upset. After this caused slight change on their stress levels, finally in Task 4 the time limit boost their stress levels to higher.

Furthermore, other than handwritings, at the end of each task, 15 signatures were acquired from participants to be able to measure emotional state reflections on their signatures as well. To obtain the ground truth for the possible three higher-level states, specifically in this case the question of whether or not a subject is 'happy' or otherwise, whether or not a subject is 'sad' or otherwise, and whether or not a subject is experiencing 'stress' or otherwise, participants were asked to express their feelings by filling in the score sheet at the end of each task, scoring their emotional states namely happy, sad and stressed at the time of carrying out the task by assigning a score from 1 (least) to 10 (most) as in [35].

Fig. 2, Fig. 3, and Fig. 4 illustrates the score distribution of happy, sad and stress emotion recorded in each task respectively. It can clearly be observed that the effect of the visual media that has been used during data collection was strong enough to ensure the desired emotion of the participants thus proves the validity of our approach. Having acquired appropriate ground truth, classification can be done more accurately.

In order to make sure that our proposed approach – letting the participants to watch a media (video) and write some text with their own cues, signing their signature and then to express their feelings by filling in the score sheet at the end of each task – would provide us reliable ground truth

TABLE 4. Emotion manipulation success rates.

	Task1 Happy No. of Participants	Task2 Happy No. of Participants	Difference
Female	25 (64.1%)	35 (89.74%)	10 (25.64%)
Male	41 (43.16%)	83 (87.37%)	42 (44.21%)
	Task1 Sad No. of Participants	Task3 Sad No. of Participants	Difference
Female	1 (2.56%)	29 (74.36%)	28 (71.8%)
Male	1 (1.05%)	56 (58.95%)	55 (57.9%)
	Task1 Stress No. of Participants	Task4 Stress No. of Participants	Difference
Female	7 (17.95%)	22 (56.41%)	15 (38.46%)
Male	11 (11.58%)	35 (36.84%)	24 (25.26%)

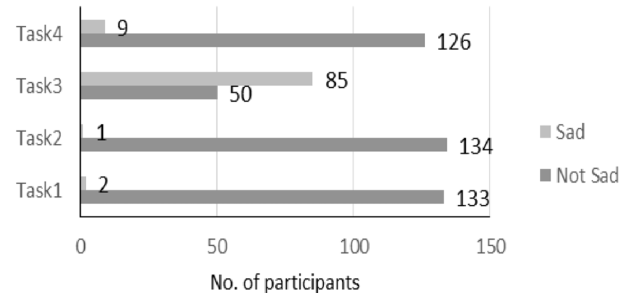


FIGURE 3. The score distribution of sad emotion recorded in each task.

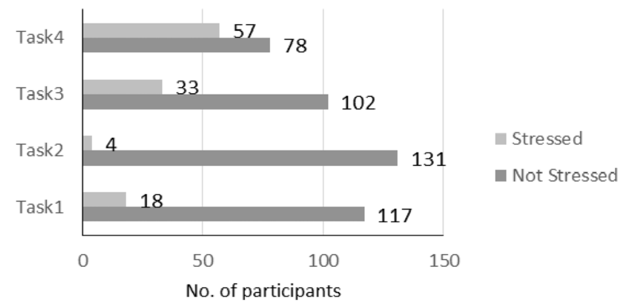


FIGURE 4. The score distribution of stress emotion recorded in each task.

labels of emotion, an experiment is performed based on the obtained labels. The difference between before and after emotion manipulation is evaluated between Task 1 (which is the natural task) and other each Task (i.e. Task2, Task3 and Task4).

Let T_F Total no. of female participants with T_{FHT1} = no. of female participants feeling happy in Task1 and, T_{FHTN} is the no. of female participants feeling happy in TaskN (where N here represents the Task number), then the ratio of happy females between Task1 and TaskN can be defined as follows:

$$RH_F = (T_{FHTN}/T_F) - (T_{FHT1}/T_F) \qquad (1)$$

The approach applied has revealed sound results as can be noticed from the Table 4 with the equation (1) applied. It can clearly be seen that emotion manipulation method was successful for both male and female participants but more successful on female participants except Task 2 (happy task). The biggest difference before and after emotion manipulation was observed is in Task 3 (sad task) with 72% difference on females.

D. ORGANIZATION OF THE DATABASE

The proposed CIU Handwritten Database has been organised in the following format;

- 134 directories are contained under the main directory as shown in Fig. 5.
- Documents: Consist of test subjects' demographic information (134 text file), extracted features and graphical representation of 6 handwritings and 60 signatures.
- File Format: The handwritings and signatures as well as subjects' demographic information are stored in text files (.txt) and graphics files are stored as an image (.jpg) files.

Test subjects are numbered from E001 to E140 sequentially.

However, there are some missing ID numbers which are from subjects who registered for the trial but failed to donate any handwriting and signature samples. Files are named using HW for handwriting and SG for signature followed by the subject ID. For example, the first handwriting donated by subject 1 is contained in HW1.txt (the time stamped list of pen coordinates that comprise the handwriting) and SG1.jpg (the graphic representation of the signature).

The text files themselves, one for each overlay, contain information about the subject (taken from the relevant subject.txt file at the time of capture), the tablet information (technical data retrieved from the capture device) and finally the data packets themselves, consisting of these tab separated fields:

The JPG files are generated at a resolution of 100 pixels per inch. Each handwriting and signature biometric samples is represented as black on a white background. A white border of 20 pixels is included in each image.

E. LEGAL ISSUES AND DATABASE DISRIBUTION

Given the concerns about security, confidentiality of proprietary information, and fraud in general, the collection, storage, and use of handwriting and signature biometrics raise a number of legal, ethical and social concerns. Therefore, before the data collection process, participants have been informed about the purpose of the study, his/her rights, data confidentiality, and protection which can be summarised as follows;

Consent of participants: achieved by reading the information sheet and signing the associated consent form with the opportunity to discuss this with a member of the research team.

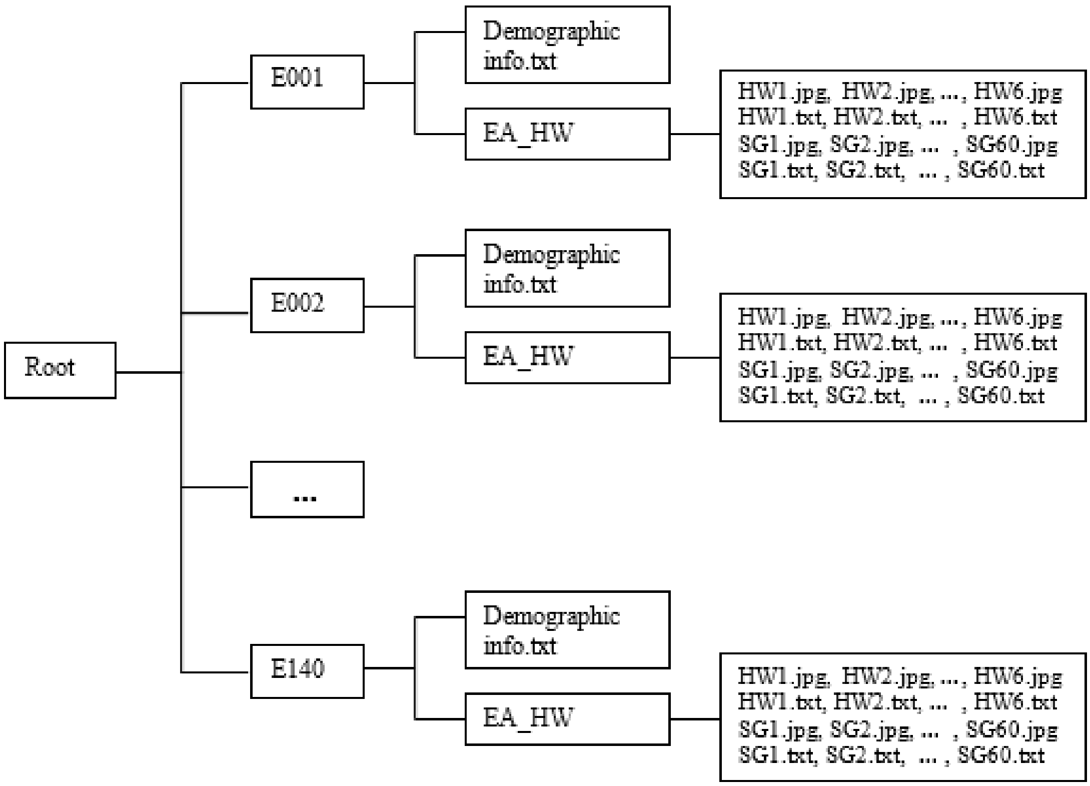


FIGURE 5. Directory structure of the proposed database.

Also participants have been told of the use to which data will be put (e.g. research publication)

Withdrawal from the data collection: participants have been told explicitly (both in the participant information sheet and associated consent form) that they are free to leave the study at any time without jeopardy.

Confidentiality of a participant: have been guaranteed by linking the acquired signature and handwriting samples to an identity number and only this identity number to be used in data analysis.

Protection of participants: the participants are not at risk of physical or psychological harm greater than encountered in ordinary life. This is also approved by the head of Psychology Department of CIU.

IV. EXPERIMENTAL SETTINGS AND RESULTS

In this study, several different types of experiments have been performed to investigate and analyse individuals' emotional state through their handwriting and signature biometric data which can be seen in Fig. 6.

A. FEATURE EXTRACTION

Unlike offline signature and handwriting biometrics that focus on the shape and structural characteristics of an image, this study focuses on the online features collecting path and time dependent features, since due to gathering input data during the writing process, on-line systems generally perform better than their off-line counterparts [42].

By using an electronic tablet and pen to capture online features of signatures and handwriting, it is possible to extract all points of every character. However, this does not mean every point we capture is useful. For example, in Task 1 all participants are asked to write a fixed text but in Task 2 and Task 3 they are free to write what they want (variable text). Therefore, using "the total time taken to execute the text" as a feature is not valuable and can even manipulate the classification results for handwriting biometrics. Table 5 lists the online features, extracted for both signature and handwriting, which are commonly used in signature and handwriting processing in the literature [14], [40], [59]–[61].

B. PREPROCESSING

Since the signature and handwriting data generated in different sizes are heterogeneous, normalization is required to transform these signature and handwriting data into a common domain before analyzing them. The pre-processing method used in this study is called z-score which is one of the most commonly used score normalization techniques in biometrics [62]. The z-score normalization formula can be found in equation (2):

$$Z_k^i = \frac{x_k^i - \mu_k}{\sigma_k} \tag{2}$$

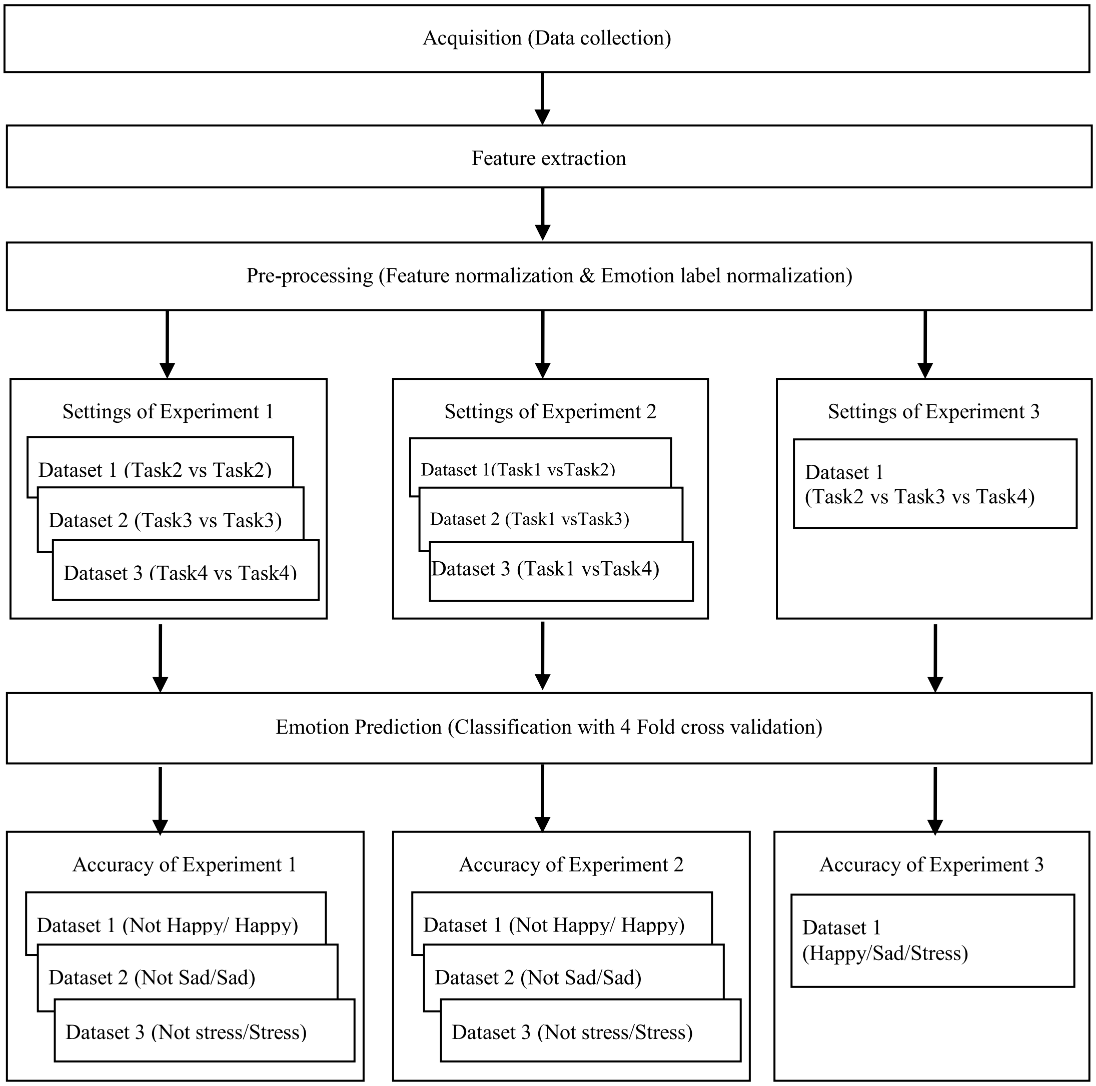


FIGURE 6. System architecture of emotion prediction from online handwriting and signature biometrics.

TABLE 5. Extracted features.

Feature No	Feature Description
F1	Average pen velocity in x
F2	Average pen velocity in y
F3	Maximum pen velocity in x – Average pen velocity in x
F4	Maximum pen velocity in y - Average pen velocity in y
F5	Maximum pen velocity in x - Minimum pen velocity in y
F6	Average pen acceleration in x
F7	Average pen acceleration in y
F8	Azimuth
F9	Altitude
F10	Pressure
F11	The number of times pen passes though the midline of the signature/handwriting

where $i = 1, 2, \ldots n$, and n is the number of samples, $k = 1, 2, \ldots m$, and m is the number of features, μ_k is the mean and σ_k is the standard deviation of all samples in the kth feature, x_k^i is the kth feature of the ith sample, z_k^i is the mean and variance normalized kth feature of the ith sample. Before proceeding to the emotion prediction process, emotion labels (scaled between 1 and 10) which are ten classes also normalized into two class label, by taking into consideration the scores assigned by each subject and simple thresholding technique. The idea behind normalization here is to eliminate any incongruity between the perceived and experienced emotion and meticulously analyse the data for correct prediction. Theoretically, it is difficult to accurately examine handwriting for emotion prediction due to the fact that there is no exact threshold value available [36] to divide emotions into two class problem.

Therefore, we considered two different threshold values in order to compensate with this issue. At first, we considered the threshold value for happy/sad/stressed to be 8, which means if the subject assigned 8 and above for the emotional feeling then it is considered as positive and

TABLE 6. Prediction from handwriting using same task.

Threshold	No. Of Samples Natural	Positive	Prediction	KNN	Jrip	Random Forest
8	68	66	Task2 vs Task2 (NOT HAPPY/ HAPPY)	50,75	52,99	52,24
	132	2	Task3 vs Task3 (NOT SAD/SAD)	97,76	97,76	98,50
	116	17	Task4 vs Task4 (NOT STRESS/STRESS)	75,19	87,22	84,96
7 with 8 ignored	68	34	Task2 vs Task2 (NOT HAPPY/HAPPY)	57,84	66,67	64,71
	132	1	Task3 vs Task3 (NOT SAD/SAD)	99,25	99,25	99,25
	116	12	Task4 vs Task4 (NOT STRESS/STRESS)	82,81	90,63	90,63

negative otherwise. Another threshold value defined in the experimental work is 7, where threshold value 8 is ignored this time to extend the boundaries between the positive and negative feelings.

C. EMOTION PREDICTION

Initially, all features are extracted from the handwriting and signature biometric data as described in Subsection A. Subsequently, extracted features and emotion labels are normalized as described in Subsection B. Finally, several different experiments are performed to predict the emotional state, which will be explained and presented in this Subsection.

For all performed experiments, in order to predict emotion from both handwriting and signature biometric data, all three classification methods namely K-Nearest Neighbor (k-NN), JRIP and Random Forest have been applied using WEKA software [63] with default parameters. KNN is one of the most popularly applied classification method in pattern recognition and, in certain cases, its accuracy is greater than state-of the-art classifiers especially when the value of k is small (in all experiments the value of k is 1) and it takes less time during classification [68]. On the other hand, Random Forest algorithm have an impressive prediction performance since it can handle high dimensional spaces as well as large number of training samples [69] which was the case for signature samples (8040 in total) in the current study. JRIP is another robust algorithm in terms of complexity of tree structure. Therefore, these classification algorithms have been chosen in terms of high accuracy with large number of training samples (RF, JRIP) and time complexity (KNN).

Four fold cross validation technique is used for the prediction performance evaluation. In other words, the whole dataset is randomly partitioned into four subsets of which one of them is used for testing data, and the remaining three subsets are used for training data. Although, three samples per subject were collected in Task 1, only the first samples of subjects are used in classification due to the problem of imbalanced data.

Several experiments are performed to analyze and demonstrate the emotion prediction accuracy using two different threshold values (for the emotion score normalization) and three classifiers.

D. RESULTS

Subsequently, three different types of experiments have been conducted for classification to predict the emotion from both signature and handwriting samples. Experiment 1 comprise different participants' samples from same tasks, Experiment 2 comprise samples from Task 1 with other tasks and Experiment 3 comprise those participants' samples who scored happy or sad or stressed in each task except Task 1.

Experiment 1: In order to be consistent and to be able to compare our findings with the literature, first experimental study is performed using the methodology applied in the study [35] which is considered as a preliminary study to the emotion prediction from handwriting biometrics. In this proposed emotion prediction method, prediction is performed for each task individually. In each task, participants' biometric data labelled as "natural" or "positive" (i.e. not happy vs happy, not sad vs sad and not stress vs stress) based on the defined threshold for the emotion scores. Hence, for this experiment, three sets of dataset are formed as follows;

Dataset 1: contains subjects' biometric samples from Task 2 whose score is labelled as "natural" (not happy) or "positive" (happy).

Dataset 2: contains subjects' biometric samples from Task 3 whose score is labelled as natural (not sad) or "positive" (sad).

Dataset 3: contains subjects' biometric samples from Task 4 whose score is labelled as "natural" (not stressed) or "positive" (stressed).

Exactly the same experiment with the aforementioned study [35] is performed to predict the emotion using both handwriting and signature biometrics. Results are presented in Table 6 and Table 7 for the handwriting and signature biometrics respectively.

The experimental results both for the handwriting and signature biometrics could not be considered as reliable for the Task 3 (sad) and Task 4 (stress) prediction, due to the imbalanced number of samples between "natural" and "positive" emotions (for instance; there are 116 "natural" (not stress) and 17 "positive" (stress) samples for handwriting, and 1740 "natural" (not stress) and 180 "positive" (stress) samples for signature prediction) as can be noticed from Table 6 and Table 7 for the handwriting and signature emotion prediction respectively. Therefore, comparing our prediction results with study [35] and making any conclusions is almost

TABLE 7. Prediction from signature using same task.

Threshold	No. Of Samples Natural	Positive	Prediction	KNN	Jrip	Random Forest
8	1020	990	Task2 vs Task2 (NOT HAPPY/HAPPY)	77,51	63,48	74,18
	1980	30	Task3 vs Task3 (NOT SAD/SAD)	98,56	97,91	98,61
	1740	270	Task4 vs Task4 (NOT STRESS/STRESS)	90,15	87,26	88,91
7 with 8 ignored	1020	510	Task2 vs Task2 (NOT HAPPY/HAPPY)	79,93	69,90	76,93
	1980	15	Task3 vs Task3 (NOT SAD/SAD)	99,29	99,19	99,40
	1740	180	Task4 vs Task4 (NOT STRESS/STRESS)	92,60	90,63	99,19

TABLE 8. Prediction from handwriting using task1 vs other tasks.

Threshold	No. Of Samples Natural	Positive	Prediction	KNN	Jrip	Random Forest
8	68	118	Task1 vs Task2 (NOT HAPPY vs HAPPY)	76,34	87,63	87,01
	132	85	Task1 vs Task3 NOT SAD vs (SAD)	53,46	67,28	62,67
	79	57	Task1 vs Task4 (NOT STRESS vs STRESS)	85,29	84,56	87,50
7 with 8 ignored	68	96	Task1 vs Task2 (NOT HAPPY vs HAPPY)	51,22	57,93	56,10
	132	59	Task1 vs Task3 NOT SAD vs (SAD)	56,55	73,30	69,63
	79	24	Task1 vs Task4 (NOT STRESS vs STRESS)	82,52	89,32	92,23

impossible for the sad and stress prediction. The main reason for unequal number of samples in these tasks in our proposed database is due to the robustness of our data collection methodology used to change the emotions of the subjects during each task.

On the other hand prediction results based on Task 2 (happy) yields balanced (i.e. acceptable) number of samples between "natural" and "positive" emotions when threshold is selected as 8 (for instance; there are 68 "natural" (not happy) and 66 "positive" (happy) samples for handwriting, and 1020 "natural" (not happy) and 990 "positive" (happy) samples for signature prediction) as can be noticed from Table 6 and Table 7 for the handwriting and signature emotion prediction respectively.

Hence, comparing our prediction results based on happy task with the study in [35] based on handwriting biometrics, it can be observed that both findings exhibit similar prediction accuracy which is around 50-53 %. In addition, as first experimental result in the literature, our results presents that it is possible to predict "happy" emotion from the signature biometrics up to approximately 75% accuracy as shown in Table 7.

Since the methodology applied in [35] is providing unreliable results due to the imbalanced number of samples between "natural" and "positive" emotions, second experiment is performed by using our new proposed methodology. Recalling the represented experimental analysis in Table 4 (in Section III - C); it can be noticed that the majority of the participants started to the data collection without having any extreme emotions (i.e. not very happy, not very sad or not very stressed). Hence, the participants mostly scored their emotions below the defined boundaries (i.e. below 7 or 8 and above 5) in Task1. Therefore, in this paper, as a new methodology; it is reliably assumed that participant's emotions in Task1 represent the "natural" emotion. On the other hand, it is reliably assumed that participant's emotions in Task2, Task3 and Task4 represent the "happy", "sad" and "stress" emotions respectively.

Experiment 2: Hence, as a second experimental study, three datasets are formed for both signature and handwriting biometric data;

Dataset 1: Contains biometric samples in Task1 which are labelled as "natural" (not happy) and biometric samples in Task2 which are labelled as "positive" (happy).

Dataset 2: Contains biometric samples in Task1 which are labelled as "natural" (not sad) and biometric samples in Task3 which are labelled as "positive" (sad).

Dataset 3: Contains biometric samples in Task1 which are labelled as "natural" (not stressed) and biometric samples in Task4 which are labelled as "positive" (stress).

Using the three formed datasets the obtained results are shown in Table 8 and Table 9 for the handwriting and signature emotion prediction respectively.

Considering handwriting and signature biometrics, most consistent prediction performance between the emotions (i.e. prediction of happy, sad and stress) and most reliability (balanced number of samples in the "natural" and "positive" classes) is achieved using threshold 8 and a Random Forest classifier with accuracies approximately between 60% - 88% for the handwriting and 65% - 70% for the signature biometrics. Classification results were even better when threshold was 7 (8 ignored), up to 92% for handwriting and 85% for signature with the same classifier (see Table 8 and Table 9). Robustness of Random Forest classifier is related with the strength against high dimensional data and large data sets which is aforementioned in Subsection C. Another remarkable point elicited here that is worth of attention was; best prediction results obtained in this experiment was stress both for handwriting and signature features.

TABLE 9. Prediction from signature using task1 vs other tasks.

Threshold	No. Of Samples Natural	Positive	Prediction	KNN	Jrip	Random Forest
8	1020	1770	Task1 vs Task2 (NOT HAPPY vs HAPPY)	68,17	62,15	67,53
	1980	1275	Task1 vs Task3 NOT SAD vs (SAD)	67,10	60,58	65,50
	1740	855	Task1 vs Task4 (NOT STRESS vs STRESS)	72,25	65,40	69,98
7 with 8 ignored	1020	1440	Task1 vs Task2 (NOT HAPPY vs HAPPY)	68,05	59,47	66,79
	1980	885	Task1 vs Task3 NOT SAD vs (SAD)	73,19	68,24	71,66
	1740	360	Task1 vs Task4 (NOT STRESS vs STRESS)	82,81	82,29	85,29

TABLE 10. Prediction from handwriting using happy, sad and stress samples from task 1, task 2 and task 3 in terms of balanced and imbalanced sample sizes.

Threshold	Class distribution	No. of Samples Happy	Sad	Stress	KNN	Jrip	Random Forest
8	Imbalanced	118	85	57	47,31	61,92	55,00
	Balanced	57	57	57	54.39	56.73	57.90
7 with 8 ignored	Imbalanced	96	59	24	50,28	58,66	58,66
	Balanced	24	24	24	52.78	55.56	55.56

TABLE 11. Prediction from signature using happy, sad and stress samples from task 1, task 2 and task 3 in terms of balanced and imbalanced sample sizes.

Threshold	Class distribution	No. of Samples Happy	Sad	Stress	KNN	Jrip	Random Forest
8	Imbalanced	1770	1275	855	48,44	44,82	47,54
	Balanced	855	855	855	50,49	36,34	45,30
7 with 8 ignored	Imbalanced	1440	885	360	57,02	53,45	55,27
	Balanced	360	360	360	55,37	41,85	50,74

In a wider context, it can be noted that these results obtained in handwriting and signature biometrics are encouraging with slightly high recognition rates and proves that handwriting and signature contain valuable information about one's emotional states.

Experiment 3: The third experiment is carried out using three class classification (happy or sad or stressed). Opposite to the previously performed experiments, which predicts emotion using data from two tasks (Task1 that comprise "natural" and other tasks that comprise "positive" labels), the aim of this experiment is to identify those participants' emotional state information on their handwriting and signature samples from three tasks (Task2, Task3 and Task4). The prediction performed in this experiment is between three tasks, which can be called as inter-task classification.

To carry out this experiment, participants' biometric data are tested within those who have a positive emotion, in other words, contains all subjects' biometric samples whose score in Task 2 is labelled as Happy, all subjects' biometric samples whose score in Task 3 is labelled as Sad and all subjects' biometric samples whose score in Task 4 is labelled as Stressed. Hence, all samples from Task 1 are ignored. Therefore there is only one dataset used in this experiment for classification with three different labels. However, since the number of samples belonging to each class is not balanced and this could affect the obtained results. Exactly same experiment is performed by forming a balanced dataset, with including first 57 and 24 samples (for threshold 8 and threshold 7 with 8 ignored respectively) for each class for the handwriting biometrics and, and first 885 and 360 samples (for threshold 8 and threshold 7 with 8 ignored respectively) for each class for the for signature biometrics.

The obtained results are shown in Table 10 and Table 11 for the handwriting and signature emotion prediction respectively.

Considering handwriting and signature biometrics, the best prediction accuracy is achieved using a Random Forest classifier with accuracies approximately between 55% - 58% for the handwriting and 45% - 50% for the signature biometrics (see Table 10 and Table 11).

V. CONCLUSION

In this paper, initially an extensive review of a broad range of unimodal and multimodal biometric databases have been reported – especially on signature and handwriting biometrics – in terms of their availability (i.e. publicly, commercially or privately), labels (i.e. identity and/or soft-biometric label), content (i.e. fixed and/or variable task for the handwriting and offline and/or online for the signature biometrics) and the number of samples. The review and discussion presented in itself is a very valuable study since it clearly identifies the strengths and weaknesses of the currently available biometric databases, and illustrates the characteristics likely to be needed for the future research in the biometrics and soft-biometrics fields.

Secondly, the new database – which will be publicly available for research studies – proposed containing signature and handwriting with online and offline biometric samples, and with soft-biometric labels including emotional state has been introduced based on the light shed from the initial review and discussion. Proposed and presented database is the main contribution and novelty of this study since it will be the first publicly available database in the literature which includes all soft-biometric labels – specially the emotional state – for both signature and handwriting modality. Hence, the resulting database is flexible and adaptive enough to be a gateway to a deeper exploration as well as act as a benchmark to many researches. For instance, by providing the demographic information of the participants, this database will be opening new investigation approaches in handwriting and signature biometric identification, verification, soft biometric (such as gender, handedness, age, education level and nationality) and emotion prediction. It can also be beneficial for scientific community for further testing for offline and offline data since the database comes with both online (dynamic features recorded in txt format) and offline (static features recorded in jpg format).

Finally, we have specifically carried out several experiments to show that it is possible to reliably predict the emotional state from both handwriting and signature biometrics using our new recently proposed database. Although the experimental results presented here are extending the findings presented in the first study to predict emotion from the handwriting biometrics [35] by using different prediction methodology, and predicting emotion from signature biometrics as a first time, still the experimental results presented here are preliminary. However, findings are extremely encouraging for further work, on which we are currently engaged. Hence, conclusions based on the experiments – on emotion prediction – are thus expressed more in terms of overall indications of potential, rather than definitive statements.

Consequently, the proposed database hopefully may shed light on future research efforts on emotional state prediction from online and offline handwriting and signature biometrics and open new investigation approaches for both identity and soft-biometric prediction.

This study, by presenting a comprehensive handwriting and signature biometrics database, raises a number of opportunities for future research such as predicting the emotional state based on the offline features (since it provides all the collected documents in jpg format) as well as hybrid (combination of dynamic and static features) method.

REFERENCES

[1] I. Kavati, M. V. N. K. Prasad, and C. Bhagvati, "Search space reduction in biometric databases: A review," in *Computer Vision: Concepts, Methodologies, Tools, and Applications*. Harrisburg, PA, USA: IGI Global, 2018, pp. 1600–1626.

[2] S. Bhatia, "Systematic review of biometric advancement and challenges," *Int. J. Electron. Eng.*, vol. 11, no. 1, pp. 812–821, 2019.

[3] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? A survey on soft biometrics," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 3, pp. 441–467, Mar. 2016.

[4] P. Tome, J. Fierrez, R. Vera-Rodriguez, and M. S. Nixon, "Soft biometrics and their application in person recognition at a distance," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 3, pp. 464–475, Mar. 2014.

[5] M. Fairhurst, C. Li, and M. Erbilek, "Exploiting biometric measurements for prediction of emotional state: A preliminary study for healthcare applications using keystroke analysis," in *Proc. IEEE Workshop Biometric Meas. Syst. Secur. Med. Appl. (BIOMS)*, Oct. 2014, pp. 74–79.

[6] A. Alberdi, A. Aztiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *J. Biomed. Inform.*, vol. 59, pp. 49–75, Feb. 2016.

[7] A. Kathait and A. Singh, "Automated prediction of human behavior system for career counsellingof an ndividual through handwriting analysis/graphology," *HCTL Open Int. J. Technol. Innov. Res.*, vol. 12, 2014.

[8] I. Carrillo, V. Meza-Kubo, A. L. Morán, G. Galindo, and E. García-Canseco, "Emotions identification to measure user experience using brain biometric signals," in *Human Aspects of IT for the Aged Population. Design for Aging*, vol. 9193, J. Zhou and G. Salvendy, Eds. Cham, Switzerland: Springer, 2015, pp. 15–25.

[9] S. Van der Mussele, K. Bekelaar, N. Le Bastard, Y. Vermeiren, J. Saerens, N. Somers, P. Mariën, J. Goeman, P. P. De Deyn, and S. Engelborghs, "Prevalence and associated behavioral symptoms of depression in mild cognitive impairment and dementia due to Alzheimer's disease," *Int. J. Geriatric Psychiatry*, vol. 28, no. 9, pp. 947–958, 2013.

[10] Frontex, in *Proc. 2nd Global Conf. Exhib. Future Develop. Automated Border Control*, Immigration and Border Service, Portugal, Warsaw, 2013.

[11] ICAO. *ICAO (International Civil Organization)*. Accessed: Aug. 20, 2017. [Online]. Available: http://www.icao.int

[12] IATA. *IATA, IATA (International Air Transport Association)*. Accessed: Aug. 20, 2017. [Online]. Available: http://www.iata.org

[13] *European Committee for Standardization*. Accessed: Aug. 20, 2017. [Online]. Available: http://www.cen.eu

[14] M. Fairhurst, M. Erbilek, and L. Cheng, "Study of automatic prediction of emotion from handwriting samples," *IET Biometrics*, vol. 4, no. 2, pp. 90–97, 2015.

[15] *Center for Biometrics and Security Research*. [Online]. Available: http://www.cbsr.ia.ac.cn/english/Databases.asp

[16] S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J. L. Les Jardins, J. Lunter, Y. Ni, and D. Petrovska-Delacrétaz, "BIOMET: A multimodal person authentication database including face, voice, fingerprint, hand and signature modalities," in *Proc. Int. Conf. Audio-Video-Based Biometric Person Authentication*, Berlin, Germany, 2003, pp. 845–853.

[17] B. Dumas, C. Pugin, J. Hennebert, D. Petrovska-Delacrétaz, A. Humm, F. Evéquoz, R. Ingold, and D. Von Rotz, "MyIDea-multimodal biometrics database, description of acquisition protocols," *Proc. 3rd COST*, vol. 275, 2005, pp. 59–62.

[18] L. Likforman-Sulem, A. Esposito, M. Faundez-Zanuy, S. Clemencon, and G. Cordasco, "EMOTHAW: A novel database for emotional state recognition from handwriting and drawing," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 2, pp. 273–284, Apr. 2017.

[19] J. Fierrez, J. Galbally, J. Ortega-Garcia, M. R. Freire, F. Alonso-Fernandez, D. Ramos, D. T. Toledano, J. Gonzalez-Rodriguez, J. A. Siguenza, J. Garrido-Salas, and E. Anguiano, "BiosecurID: A multimodal biometric database," *Pattern Anal. Appl.*, vol. 13, no. 2, pp. 235–246, May 2010.

[20] (Jun. 2014). *Face Recognition Grand Challenge*. [Online]. Available: http://www.nist.gov/itl/iad/ig/frgc.cfm

[21] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in *Proc. Eur. Workshop Biometrics Identity Manage.*, Berlin, Germany, 2008, pp. 47–56.

[22] *UTKFace Large Scale Face Dataset*. [Online]. Available: https://susanqq.github.io/UTKFace/

[23] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2010, pp. 94–101.

[24] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

[25] J. Ortega-Garcia *et al.*, "The multiscenario multienvironment biosecure multimodal database (BMDB)," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 6, pp. 1097–1111, Jun. 2010.

[26] H. Proença, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, "The UBIRIS.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1529–1535, Aug. 2010.

[27] A. M. Badawi, M. Mahfouz, R. Tadross, and R. Jantz, "Fingerprint-based gender classification," in *Proc. IPCV*, vol. 6, 2006, pp. 41–46.
[28] R. J. Tom and T. Arulkumaran, "Fingerprint based gender classification using 2D discrete wavelet transforms and principal component analysis," *Int. J. Eng. Trends Technol.*, vol. 4, no. 2, pp. 199–203, 2013.
[29] S. Sarkar, J. P. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanid gait challenge problem: Data sets, performance, and analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 162–177, Feb. 2005.
[30] D. A. Van Leeuwen, J. Kessens, E. P. Sanders, and H. van den Heuvel, "Results of the n-best 2008 dutch speech recognition evaluation," Tech. Rep., 2009, pp. 2571–2574.
[31] *TIMIT Acoustic-Phonetic Continuous Speech Corpus*. [Online]. Available: https://catalog.ldc.upenn.edu/LDC93S1
[32] A. Batliner, C. Hacker, S. Steidl, E. Nöth, S. D'Arcy, M. J. Russell, and M. Wong, "'You stupid tin box'-children interacting with the AIBO robot: A cross-linguistic emotional speech corpus," in *Proc. LREC*, 2004.
[33] M. Bijankhan, J. Sheikhzadegan, and M. R. Roohani, "FARSDAT—The speech database of Farsi spoken language," in *Proc. Austral. Conf. Speech Sci. Technol.*, 1994, pp. 826–831.
[34] G. Amayeh, G. Bebis, and M. Nicolescu, "Gender classification from hand shape," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2008, pp. 1–7.
[35] M. Fairhurst, M. Erbilek, and C. Li, "Enhancing the forensic value of handwriting using emotion prediction," in *Proc. 2nd Int. Workshop Biometrics Forensics (IWBF)*, 2014, pp. 1–6.
[36] J. Fisher, A. Maredia, A. Nixon, N. Williams, and J. Leet, "Identifying personality traits, and especially traits resulting in violent behavior through automatic handwriting analysis," in *Proc. Student-Fac. Res. Day*. New York, NY, USA: Pace Univ., 2012.
[37] T. Chatzisterkotis, "An examination of quantitative methods for forensic signature analysis and the admissibility of signature verification system as legal evidence," M.S. thesis, Univ. Kent, Canterbury, U.K., 2015.
[38] S. Kedar, V. Nair, and S. Kulkarni, "Personality identification through handwriting analysis: A review," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 5, no. 1, pp. 548–556, 2015.
[39] M. Erbilek, M. Fairhurst, and C. Li, "Exploring gender prediction from digital handwriting," in *Proc. 24th Signal Process. Commun. Appl. Conf. (SIU)*, Zonguldak, Zonguldak, May 2016, pp. 789–792.
[40] M. Erbilek and M. Fairhurst, "Framework for managing ageing effects in signature biometrics," *IET Biometrics*, vol. 1, no. 2, pp. 136–147, Jun. 2012.
[41] F. Asicioglu and N. Turan, "Handwriting changes under the effect of alcohol," *Forensic Sci. Int.*, vol. 132, no. 3, pp. 201–210, 2003.
[42] R. Plamondon and S. N. Srihari, "Online and off-line handwriting recognition: A comprehensive survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 63–84, Jan. 2000.
[43] I. Abdelaziz and S. Abdou, "AltecOnDB: A large-vocabulary arabic online handwriting recognition database," 2014, *arXiv:1412.7626*. [Online]. Available: https://arxiv.org/abs/1412.7626
[44] H. El Abed, V. Märgner, M. Kherallah, and A. M. Alim, "ICDAR 2009 arabic handwriting recognition competition," in *Proc. 10th Int. Conf. Document Anal. Recognit. (ICDAR)*, 2009, pp. 1379–1383.
[45] M. Pechwitz, S. S. Maddouri, V. Märgner, N. Ellouze, and H. Amiri, "IFN/ENIT-database of handwritten Arabic words," in *Proc. CIFED*, vol. 2, 2002, pp. 127–136.
[46] C. Viard-Gaudin, P. M. Lallican, S. Knerr, and P. Binter, "The IRESTE On/Off (IRONOFF) dual handwriting database," in *Proc. 5th Int. Conf. Document Anal. Recognit. (ICDAR)*, 1999, pp. 455–458.
[47] I. Guyon, L. Schomaker, R. Plamondon, S. Janet, and R. Liberman, "UNIPEN project of on-line data exchange and recognizer benchmarks," in *Proc. Int. Conf. Pattern Recognit.*, 1994, pp. 29–33.
[48] J. J. Hull, "A database for handwritten text recognition research," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 550–554, May 1994.
[49] X. Lu, Y. Fang, W. Kang, Z. Wang, and D. D. Feng, "SCUT-MMSIG: A multimodal online signature database," in *Proc. Chin. Conf. Biometric Recognit.*, 2017, pp. 729–738.
[50] A. Kholmatov and B. Yanikoglu, "SUSIG: An on-line signature database, associated protocols and benchmark results," *Pattern Anal. Appl.*, vol. 12, no. 3, pp. 227–236, 2009.
[51] D.-Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll, "SVC2004: First international signature verification competition," in *Proc. ICBA*, in Lecture Notes in Computer Science, vol. 3072. Berlin, Germany: Springer, 2004, pp. 16–22.
[52] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero, and Q.-I. Moro, "MCYT baseline corpus: A bimodal biometric database," *IEE Proc.-Vis., Image Signal Process.*, vol. 150, no. 6, pp. 395–401, 2003.
[53] S. Manimala, G. Meghasree, G. G. Poornima, and C. Sindhu, "Automated handwriting analysis for human behavior prediction," *Int. J. Comput. Sci. Eng.*, vol. 5, nos. 7–16, p. 5, 2016.
[54] F. Rizvi and M. S. H. Khiyal, "Personality prediction through offline handwriting analysis," *J. Multidisciplinary Eng. Sci. Stud.*, vol. 3, no. 1, pp. 1282–1288, 2017.
[55] K. Kaymaz, N. Gürsakal, and U. Eroğlu, "Analyzing the relationship between signature and personal traits: A research on managers," *J. Ind. Relations Hum. Resour.*, vol. 12, no. 3, pp. 29–40, 2010.
[56] Y. Liang, M. C. Fairhurst, R. M. Guest, and M. Erbilek, "Automatic handwriting feature extraction, analysis and visualization in the context of digital palaeography," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 30, no. 4, 2016, Art. no. 1653001.
[57] Wacom. [Online]. Available: http://www.wacom-europe.com
[58] Y. Bay, M. Erbilek, A. F. G. Cyprus, and E. Celebi, "The impact of visual and blind signing on signature biometrics," in *Proc. Int. Conf. Comput. Intell. Commun. Netw. (CICN)*, Kyrenia, Northern Cyprus, 2017, pp. 161–164.
[59] S. Rashidi, A. Fallah, and F. Towhidkhah, "Authentication based on signature verification using position, velocity, acceleration and Jerk signals," in *Proc. 9th Int. ISC Conf. Inf. Secur. Cryptol. (ISCISC)*, 2012, pp. 26–31.
[60] M. Fayyaz, M. Hajizadeh-Saffar, M. Sabokrou, and M. Fathy, "Feature representation for online signature verification," 2015, *arXiv:1505.08153*. [Online]. Available: https://arxiv.org/abs/1505.08153
[61] D. Impedovo and G. Pirlo, "Automatic signature verification: The state of the art," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 38, no. 5, pp. 609–635, Sep. 2008.
[62] F. Alsaade, N. Zaman, M. Z. Dawood, and S. H. A. Musavi, "Effectiveness of score normalisation in multimodal biometric fusion," *J. Inf. Commun. Technol.*, vol. 3, no. 1, pp. 29–35, 2009.
[63] *Weka 3.8*. [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/
[64] R. Haraksim, J. Galbally, and L. Beslay, "Study on fingermark and palmmark identification technologies for their implementation in the schengen information system," Joint Res. Centre, Luxembourg City, Luxembourg, Tech. Rep., 2019.
[65] B. Uğurlu, R. Kandemir, A. Carus, and E. Abay, "An expert system for determining the emotional change on a critical event using handwriting features," *TEM J.*, vol. 5, no. 4, pp. 480–486, May 2016.
[66] J. Han, G. hernyshov, D. Zheng, P. Gao, T. Narumi, K. Wolf, and K. Kunze, "Sentiment pen: Recognizing emotional context based on handwriting features," in *Proc. 10th Augmented Hum. Int. Conf.*, Mar. 2019, Art. no. 24.
[67] V. R. Lokhande and W. Bharti Gawali, "Analysis of signature for the prediction of personality traits," in *Proc. IEEE 1st Int. Conf. Intell. Syst. Inf. Manage. (ICISIM)*, Oct. 2017, pp. 44–49.
[68] A. Singh, M. N. Halgamuge, and R. Lakshmighantan, "Impact of different data types on classifier performance of random forest, Naive Bayes, and k-nearest neighbors algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 12, pp. 1–11, 2017.
[69] A. Sage, "Random forest robustness, variable importance, and tree aggregation," Tech. Rep., 2018.

YASEMIN BAY AYZEREN received the B.Sc. degree in computer science and information technology and the M.Sc. degree in computer science from Eastern Mediterranean University, Cyprus, in 1996. She is currently pursuing the Ph.D. degree in management information systems with Cyprus International University. Her research interests include information retrieval, machine learning, and handwriting and signature biometrics.

MERYEM ERBILEK received the Ph.D. degree in electronic engineering from the University of Kent, Canterbury, U.K., in 2013. She is currently an Assistant Professor with the Department of Computer Engineering, Cyprus International University. Her research interests include pattern recognition, image analysis, and handwritten signature analysis, especially, security and biometrics.

ERBUĞ ÇELEBI received the B.Sc. degree in computer science and applied mathematics from Ege University, Turkey, in 1996, and the M.Sc. and Ph.D. degrees in computer engineering from Dokuz Eylul University, İzmir, Turkey, in 1999 and 2006, respectively. He is currently a Professor with the Department of Management Information Systems, Cyprus International University. His interests include information retrieval, machine learning, and information systems.