Extending the Predictive Capabilities of Hand-oriented Behavioural Biometric Systems

A Thesis Submitted to The University of Kent
For The Degree of Doctor of Philosophy (Ph.D.)
In Electronic Engineering

By

Cheng Li

September 2016

Abstract

The discipline of biometrics may be broadly defined as the study of using metrics related to human characteristics as a basis for individual identification and authentication, and many approaches have been implemented in recent years for many different scenarios. A sub-section of biometrics, specifically known as soft biometrics, has also been developing rapidly, which focuses on the additional use of information which is characteristic of a user but not unique to one person, examples including subject age or gender. Other than its established value in identification and authentication tasks, such useful user information can also be predicted within soft biometrics modalities. Furthermore, some most recent investigations have demonstrated a demand for utilising these biometric modalities to extract even higher-level user information, such as a subject's mental or emotional state. The study reported in this thesis will focus on investigating two soft biometrics modalities, namely keystroke dynamics and handwriting biometrics (both examples of hand-based biometrics, but with differing characteristics). The study primarily investigates the extent to which these modalities can be used to predict human emotions. A rigorously designed data capture protocol is described and a large and entirely new database is thereby collected, significantly expanding the scale of the databases available for this type of study compared to those reported in the literature. A systematic study of the predictive performance achievable using the data acquired is presented. The core analysis of this study, which is to further explore of the predictive capability of both handwriting and keystroke data, confirm that both modalities have the capability for predicting higher level mental states of individuals. This study also presents the implementation of detailed experiments to investigate in detail some key issues (such as amount of data available, availability of different feature types, and the way ground truth labelling is established) which can enhance the robustness of this higher level state prediction technique.

Acknowledgements

I would like to thank many people, in many different places now, who have help me throughout my research career.

I would like to, firstly, express my sincere gratitude and appreciation to my supervisor Professor Michael Fairhurst for the enormous support and guidance of my research journey. His demonstration of exemplary professional work ethic for the many years I worked alongside him are inspiring and will be benefiting me for my future work and life. Professor Michael Fairhurst have been an inspirational supervisor, a life teacher, a music history mentor and most importantly a great friend.

My deepest love and appreciation go to my parents Li Xi Ming and Cheng Qin Ming. The unconditional love and support I have been given by both of them are the most powerful motivation throughout my academic career. The value they taught me have make the person I am today and for that I am eternally grateful.

I would like to express my thanks to Dr Meryem Erbilek and Dr Marjory Da Costa-Abreu for their support and contribution to my work. I would also like to thank Dr Shivam Guness and Dr Shu Yang for the countless lunch adventures and their friendship and especially their positivity.

I want to thank my close friends, Dr QianQian Wu, Dr Xuan Jian Wang, Dr Yang Song, Dr Lu Zhao and ManLing Zhang for always being there for me and for the joyful memories they shared with me.

My thanks go to the EDA staff for their help and support throughout the years.

Special mention goes to my friends Dr Oscar Miguel Hurtado, Iulia Motoc and Ryan Husbands, for their friendship and support.

Finally, Dr David Rodrigues and Dr Richard Bonner are two special friends to me. They are close to my heart, for being supportive friends and most importantly for being a brother figure to support and guide my way through lifes obstacles.

Contents

List of Figures

List of Tables

Publications arising from this work:

- M. C. Fairhurst, M. Erbilek, C. Li, "Enhancing the forensic value of handwriting using emotion prediction" - *2nd International Workshop on Biometrics and Forensics (IWBF)*, pp. 1-6, 2014
- M. C. Fairhurst, C. Li, M. Erbilek,"Exploiting biometric measurements for prediction of emotional state: A preliminary study for healthcare applications using keystroke analysis" - *2014 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings*, pp. 74-79, 2014
- M. C. Fairhurst, M. Erbilek, C. Li,"Study of automatic prediction of emotion from handwriting samples" - *IET Biometrics*, IET (Volume: 4, Issue: 2), pp. 90-97, 2015
- M. C. Fairhurst, C. Li , M Abreu, "Exploring emotion prediction from biometric-based keystroke dynamics data using multiagent systems" - *6th International Conference on Imaging for Crime Prevention and Detection (ICDP-15)*, pp. 1-6, 2015
- M. Erbilek, M. C. Fairhurst, C. Li,"Exploring gender prediction from digital handwriting" - *24th Signal Processing and Communication Application Conference (SIU)*, pp. 789-792, 2016
- M. C. Fairhurst, C. Li , M Abreu,"Predictive biometrics: a review and analysis of predicting personal characteristics from biometric data" - submitted to *IET Biometrics*, 2016

CHAPTER 1

Biometrics and soft-biometrics information prediction

Biometrics is widely regarded as the study of using metrics related to human characteristics as a reliable method for individual identification and authentication. The study to be presented in this thesis concerns the utilisation of the established field of biometrics to determine characteristics of individuals beyond the determination of individual identity. The development of what might be termed *predictive biometrics* is not new, but is nevertheless an area which has developed relatively recently, and which has not generated as large a body of work as that of conventional biometrics. In particular, this study will address perhaps the newest aspect of predictive biometrics, which is the prediction of what have been described as *higher-level* properties of individuals, which relate to emotions and mental state rather than more obvious physiological characteristics.

The field of biometrics has been rapidly developing for the past two decades and more. Many currently established areas within the field have evolved from theoretical research topics to real-world applications and practical solutions, and the deployment of biometrics-based technologies is established worldwide. In the last decade, the development of wireless internet connection options and readily-available hand-held devices, such as smart phones and digital tablets, has led to an extraordinary explosion of human and computer interaction. Almost every minute of our daily life finds most individuals interacting either directly or indirectly with other people or with other devices by means of our digital devices through the internet. The growth of commercial biometric applications is now benefiting almost every industry, with areas of application which include data security, human computer interaction, border control, secure payment transactions and very many others. The rate of growth is not showing any signs of slowing down, and with the development of innovative wireless technology and more powerful and handy digital devices, the opportunities for biometric applications are facing even more demands. There is a great amount of impressive and diverse innovative work being carried out across the world in research laboratories and in the commercial and industrial sectors, which are allowing this field of study to continue to expand and penetrate more and more aspects of our lives.

Traditionally, the principal focus of studies of biometrics has been centred around the individual identification and verification capability of biometric data. This is a worldwide research topic, which has been thoroughly explored within various biometric modalities, including the recognition of fingerprints [1–4], facial appearance [5–8], iris patterning [9–11], voice characteristics [12] among the most commonly encountered. Biometrics-based processing allows individuals to be accurately identified or a claim of identity to be authenticated based on the measurement of their physiological or behavioural characteristics. Indeed, identification or verification applications remain as the main target and motivation for the majority of the current studies in the field, although new areas of research constantly emerge as deployment becomes longer-term and more routine (for example, the issues associated with individual ageing and the implications for biometric data have recently become much more high profile than previously reported [13, 14]).

Let us take the modality generally referred to as "keystroke dynamics" as an illustrative example for more detailed consideration, due to the semi-autonomous nature of the behaviour associated with this modality and the consistency and robustness of typing at a keyboard as a frequently and commonly used means of communication and interaction (and its particular relevance to a principal focus of the study to be reported in the thesis). This modality uses the fact that people tend to develop autonomous and repeatable patterns of typing which are individually characteristic, and which therefore make the typing behaviour a useful biometric, as it is complex to imitate and possesses an inherently "stable" pattern. Bhatt and Santhanam [15] reported a survey for using keystroke dynamics for authentication, which provided a detailed study on the history of keystroke dynamics development as a tool of authentication. The studies reported in [16–24] present various different methodologies and approaches to identify and authenticate users based on the behaviour pattern the users display while typing.

Keystroke data have been used in a commercial environment as the basis of security systems. Pisani and Lorena [25] report a systematic review on the use of keystroke data for the user recognition task. The review presents a rigorous method of review, by following explicit protocols for obtaining information of relative studies in the literature. The protocol defines all the keywords used to find all relevant studies through five major databases (ACM, IEEE, Science Direct, Web of Science, Scopus). The review thoroughly compares the advantages and disadvantages of keystroke dynamics. A major advantage of using keystroke dynamics is said to be that it is much more economical in comparison with many other biometric technologies, given that there is a minimum additional cost associated with hardware. However, a significant disadvantage is that keystroke dynamics usually have a high false alarm rate, where a legitimate subject is classified as an imposter.

Banerjee and Woodard [18] provide a review on the background of the psychological basis of keystroke dynamics. This study discusses different data capture methods and compares the performances achieved by researchers when experimenting using standard computer keyboards. The survey discusses the data acquisition methods and approaches, and the authors suggest that the use and acceptance of keystroke dynamics can increase by standardising databases and common data formats, and establishment of evaluation methods.

The authors in [23, 26–30] present studies which utilise keystroke data for password hardening in a typical user-name and password access control environment. The password hardening is achieved by combining the password with the way the password is entered, especially the keystroke dynamic data. Therefore, this enhances the level of security when using password-based access control system.

Tsimperidis and Katos [31], from a slightly different perspective, explore the capability for determining the user's device characteristics (desktop PC or laptop), which can be helpful in providing additional information when considering a forensic scenario, for example, where the device characteristics can help determine the location of the user.

Handwriting can be considered as another useful illustrative example, similar to the keystroke modality in some ways, and also especially relevant to the present study. Handwriting is a robust behavioural trait that has a significantly longer history in society, and is commonly used and developed throughout the entire life of most people. The authors in [32–37] show that the handwriting behavioural trait is personal to each individual and, as is required by an appropriate biometric modality, characteristic of their individual identity. The appearance of handwritten characters and the style of their formation and integration are visually different [38]. The handwriting identification task is one which has been addressed for many years by means of human expertise but which for some time has also has been developing as a target for automated processing, with image processing and pattern recognition techniques at the forefront of the tools which are generally adopted [39–42]. Authors in [43, 44] provided a review on using handwriting sample for identification and authentication. As shown in [32–34, 45, 46], writer identification is the process of identifying the specific writer of a piece of handwritten text, while in a related problem area, writer verification [47–49] aims to verify whether a handwritten text fragment is written by a named writer, generally requiring the direct comparison of two handwriting samples. Abdl et al [37] present a review on using handwriting data for user identification. The study identifies one of the main issues in handwriting identification, which is to extract sufficient features to reflect individual variations in handwriting behavioural patterns. The authors describe the difference between on-line and off-line features and suggest that feature selection based handwriting identification is needed. Other relevant work can also find in [9, 50].

The areas in which biometrics-based studies generally are expanding cover many different disciplines, and the field remains highly dynamic. For example, [51–53] present implemented biometrics applications that can be used for user identification, signature verification and payment systems on mobile devices. The authors demonstrate the real life mobile applications that can benefit by using a biometrics-based approach. The studies reported in [54–57] discuss some practical solutions for dealing with "biometrics at a distance", where the biometric data capture environments are less restricted than hitherto and offer more flexibility for the user in relation to the capture of biometric data. The reported studies use biometric data from body images, iris images and facial images captured from a distance to improve the identification and verification accuracy, this technique can benefit real scenarios such as surveillance and security, access control, e-passport and watch-list surveillance. There has also been a specific cross-disciplinary development which focuses on the interface between biometrics and forensic analysis, and [58–63] report a variety of recent studies on ideas and techniques that have proved to be mutually beneficial to both disciplines.

Recently, then, there have been a number of studies which investigate the predictive capability of biometric data, in order to determine users' characteristics, specifically beyond the general applications of biometrics which, as presented earlier, mostly relate only to the absolute identity of the individual providing the data. The study and application of so-called "soft-biometrics" has been rapidly developing. Soft-biometrics is the term which refers to characteristics which are specific to an individual, but not in themselves unique. The age or gender of an individual are good examples, since an age and gender can generally be assigned to an individual, but others with the same characteristics can readily be found.

The studies reported in [64–74] provide good demonstrations of how these soft-biometrics can improve identification processes as an assisting method to help determine information such as age and gender. The work reported in [75–77] predict handedness from handwriting and the results demonstrate that a significant level of predictive accuracy can be achieved for determining whether the handwriting samples are from a left-handed writer or a right-handed writer. This additional information can be valuable in areas such as handwriting forensics, and handwriting authentication or verification, as this additional information can assist identification, or can narrow down the sample range of the investigation.

The increasing interest in the use of soft-biometrics has allowed some excellent results to be obtained when *predicting* some specific characteristics of human subjects. This has potential implications in a variety of predictive applications.

More importantly in the context of the present study, a most recent topic of interest within the spectrum of biometrics-based prediction, is the idea of predicting other and more immediately personal characteristics of individuals, such as their emotional or mental "state", also often referred to as "higher-level" characteristics [78, 79]. Later chapters will return to a discussion of this type of prediction, but for now it is important to note that there are a limited number of studies reported in this area in the literature, which is much less extensive than either that relating to identity prediction or even that concerned with the prediction of conventional soft-biometrics such as age and so on. The facial image biometric modality is one of the most commonly explored modality considered in relation to the prediction of these higher-level states. Furthermore, this often involves a process for the recognition particularly of facial expression, then indirectly extracting the higher-level characteristics of the subjects (for example, [80]), and is thus a rather different approach for the conventional predictive biometrics approach as we shall refer to the approach adopted here.

However, facial expression is not the main method of how humans interact with computers. In today's world, the majority of the interactions between humans which are mediated via technology-based platforms are performed with either a keyboard or a handwriting tablet. We may usefully refer to these in this thesis as hand-oriented methods of interaction.

Handwriting is still applicable and, obviously, widely used in most people's daily communication and interaction activities. However, typing at a keyboard, in the last two decades has developed into another of the most common ways of communication. As already been extensively studied with facial expression, “hand-oriented behaviour” can potentially also carry additional “higher-level” information about individuals. In this thesis, our aim is extend the predictive capability of biometric data into the area of hand-oriented behavioural biometrics.

In this chapter, an overview of the state of art of the current field of soft-biometrics information prediction will be presented. The overview will include both conventional prediction of soft-biometrics characteristics, as well as the prediction of the “higher-level” soft-biometrics characteristics noted above. In this chapter the field as a whole will be reviewed in terms of a general survey of the principal topics and developments of relevance, but the subsequent experimental chapters will supplement this with more in-depth reviews relating to specific topics as they arise.

This chapter will therefore provide a base for the further studies described in the thesis, highlighting the challenges and main questions to be addressed. The study which is the focus of the thesis will suggest solutions for some of these challenges and questions in later chapters.

Section 1.1 will review the prediction of “lower-level” soft-biometrics information. Section 1.2 will review the literature of “higher-level” soft-biometrics information prediction. Section 1.3 will further review hand-oriented behavioural biometrics. Section 1.4 will discuss the research motivation in more detail and Section 1.5 will conclude this chapter and present an outline for the remainder of the thesis.

1.1 Review of prediction of “lower-level” soft-biometrics information

Individual identification/authentication is still the primary focus of biometric data processing, as presented in the brief review in the previous paragraphs. Recently, it has become more increasingly observable that the predictive capability of biometric data can provide information relating to important characteristics, thus the “lower-level” soft-biometrics information.

The “lower-level” soft-biometrics information referred to here is defined in contrast with the “higher-level” information which is the particular focus of this study. The lower-level soft-biometrics information in this case relates to information about individuals such as their age, gender, height, ethnicity, handedness and similar characteristics which fall short of identifying the individual but are nevertheless properties which are not shared by everyone. The ability to predict such varied information about an individual can offer many areas of application. The following sections will report two of the main studies that have investigated the prediction of one or multiple types of this lower-level soft-biometrics information.

1.1.1 Age estimation

Age estimation has been studied by many researchers. The authors of [81, 82], for example, present a survey of age estimation by using facial images. The survey summarizes and systematically discusses the existing models and facial image processing methods, age estimation performances, publicly available facial ageing databases, performance evaluation protocols, technical challenges and promising future directions. Although age estimation has mostly been studied in the discipline of facial image processing, as presented in three points of interests, which include: studies of different features used for age estimation or features that can be affected by ageing [83–93] , studies experiment with different algorithms that can handle the ageing effect in facial [65, 94–103] and studies that focus on handle the age changes in the facial images through the fusion of information or by using ensemble of classifiers [104–107].

Although the face modality has been most commonly investigated for age prediction, other biometric modalities have also been used to explore the nature of the relationship between subject age and their biometric data.

The authors in [108, 109] present results for age prediction based on the analysis of handwritten signatures. Fairhurst and Da Costa-Abreu [109] present a study where a multiclassifier system (MCS) was used to predict age from the writers that were categorized into three age bands: *<25*, *25-60* and *>60* years of age. The results reported indicated some general trends of the MCS which can improve the predictive accuracy when applied. The authors also show how a processing system configuration can be improved (potentially optimised in the future) in age prediction studies. Erbilek et al [108] present a novel approach of combining two biometric modalities (iris and handwritten signature) for age prediction. This study also uses the same three age bands: *<25*, *25-60* and *>60* and adopts a multiclassifier system and a novel multiagent system. The results indicate that predictive performance is likely to improve when a more extensive basic feature pool can be accessed or when optimisation of the classifier system adopted can be achieved.

Gait data also been used by the authors in [67, 110] for age prediction, and this study demonstrates the age predictive capability of using human gait appearance. Makihara et al [110] overcome the issue of the size of databases reported in the literature which are relatively small, by creating a significantly larger database. More recently, iris data have also been used for the prediction of age, as reported in [69, 111], where different sets of geometric and texture features are used. Both studies have shown preliminary indications of the age estimation capability of using iris data. However, studies using iris biometrics for age prediction are very limited in number.

The authors in [68, 112, 113] explore the age predictive capability of voice data. Fujisawa et al [68] report a study that compares the age estimation with facial images and shows that the subjective age estimated from speech signals are generally older than the age estimation achieved by using facial images. This study identifies a key factor to consider when considering two different modalities to predict age from the same subjects.

Dobry et al [112] adopt a novel dimension reduction technique, weighted-pairwise principal components analysis (WPPCA) and apply this technique to two types of systems, an age-group classifier and a precise age estimator by regression. The proposed method achieves higher predictive accuracy than the baseline system where no dimensionality reduction technique is applied and also higher accuracy than with a standard dimension reduction technique such as principal component analysis (PCA). The authors introduce a novel dimension reduction technique that can be used for further improvement of the predictive performance.

Gautam and Singh [113] designed a system that recognizes age group from unknown speech samples. The authors show that the best predictive performance achieved by this type of study is by using a neuro-fuzzy classifier. The authors focus on using this technique on children (under 9 years old), in order to develop a possible solution to help predict and diagnosis neuro-development or language delay disorder at an early stage. The fingerprint modality has also been used for age prediction by the Merkel et al [114], the study utilising a recently introduced image manipulation technique StirTrace for age estimation. The results show that decreasing image size has a greater impact on age estimation performance than a decrease in resolution of the fingerprint images. This study illustrates the key impact factor when using the StirTrace technique for age estimation with fingerprint biometrics.

1.1.2 Gender prediction

Similar to age prediction, the main modality that has been used for gender prediction is the analysis of facial images. Studies presented in [70, 115–122] have highlighted the different techniques adopted to predict gender from facial images. A relatively limited amount of work on gender prediction is reported with other biometric modalities. The authors in [123, 124] use iris data for gender prediction. Thomas et al [124] establishe one of the earliest studies for gender prediction, and reports a predictive accuracy close to 80% when using a decision tree classifier. Tapia et al [123] analyse iris texture to predict the gender of the sample donor. The study was the first to explore uniform Local Binary Patterns and has achieved very encouraging predictive accuracy, at around 91%.

Hassaine et al [125] present the details of a competition of gender prediction from handwriting. The authors present the dataset used for the competition and participating methods and predictive results. This contest provided a large publicly available dataset which allows benchmarking predictive results of gender predictions with handwriting.

Bouadjenek et al [126–128] use handwriting data for gender prediction. The authors adopt Local descriptors in order to improve the gender classification based on offline handwritten text. The proposed prediction system is based on the use of the Histogram of Oriented Gradients (HOG) feature, which aims to extract gradient directions from the handwritten text and the so-called gradient local binary patterns feature, which is an improved gradient feature that incorporates the local binary pattern neighbourhood in the gradient calculation. The author demonstrates that the features show a 7% improvement in comparison to pixel density, which is a commonly used grid feature.

In recent years (specifically over the past four years or so, from 2013 to 2016), gender prediction has also been investigated by using data collected on the internet, such as online chat and social media data, as presented in [31, 129–131].

Merler et al [129] propose a method to investigate pictures posted by users in social media, and use this information to determine the gender of the user. The authors extract *semantics* features from the images uploaded by recognizing objects, scenes and activities. The study has shown that semantic content of the images users upload can be used as an indicator to predict their gender.

Crawford and Zhu [130] first analyse the random chat network, where users are talking to randomly selected users with little information about the peers. The authors propose an approach of using masked words (an encoded list of words used between two users in a chat session) feature and network topology statistics to predict gender. The results suggest that significantly better gender prediction performance can be achieved when using network topology statistics feature. This study investigates a very specific area, random chat networks, but has shown an ability to predict gender.

Gulsen et al [131] study user (Turkish users) web behaviour to predict gender. The authors use three types of feature extracted from web log data, which include the URLs, textual content and DMOZ hierarchies of the visited pages. The results suggest that when analysis this type of data is undertaken, a Logistic Regression classifier used with a selected URL feature produces the best gender prediction performance.

This has been a valuable contribution in addressing the issues associated with the lack of appropriate databases for this type of study.

There are also studies reported using voice for age and gender prediction [73, 132]. Metze et al [132] present preliminary information on predicting age and gender on telephone speech and Bahari and Hamme [73] present a hybrid architecture of Weighted Supervised Non-Negative Matrix Factorization (WSNMF) and General Regression Neural Network (GRNN) to predict gender and age. Gait biometrics have also been used for age, gender and height prediction [133–136].

As can be seen, there have been a variety of interesting studies in the area of predicting lower level soft-biometrics information. These have many potential applications, especially with the advent of the “big data” era, where huge amounts of biometric data will become available to help techniques to develop further, and therefore more efficiently assist and benefit the general public.

Although the area of predicting lower level soft-biometrics information has shown promising progress, this raises a further significant and important question of particular relevance to the present study. This question is “What else can be predicted from biometric data?” Or, in the context of this thesis, a more specific formulation would be “What higher-level soft-biometrics information can be predicted from biometric data?”

The next section will present and briefly survey the current state of art in studies aimed at predicting higher-level biometric information.

1.2 Review of prediction of “higher-level” soft-biometrics information

As we noted previously, the term “higher-level” state in the present context refers to individual characteristics that reflect a subject's “mental” or “emotional” state. In the discussions in this thesis, these terms will generally be used interchangeably throughout, although it is recognised that different studies may adopt somewhat different terminology, and that considerable discussion is possible about subtle differences in what might be meant by these terms. There have been a limited number of reported studies which have attempted to use different biometric modalities to explore the possibilities of predicting these higher-level human mental states. To be more specific, the target is to use soft-biometrics measurement to capture the pattern representing different states, and possibly recognise what state each pattern is reflecting. Examples of what sort of “state” might be the target of such studies include the question of how happy, anxious, stressed, or relaxed a subject is feeling. This type of information can be extremely valuable in the context of systems or applications that aim to evaluate or predict human behaviour.

Human emotional state is a complex and continuously changing state. The accurate determination of the emotional state is, in itself, a challenging task. The authors in [193] conducted a survey of methods and approaches to assessing emotional state, (which is also important in establishing ground truth labelling for experimentation purposes) in emotion-dependent data. The survey also raises a number of important issues with respect to the available methods, which include whether any inconsistencies exist between perceived and experienced emotion. Ground truth labelling is an important task and inaccuracies in this task can potentially lead to data mislabelling. In order to address these issues, the authors investigate the main factors that have an impact on the labelling process, which include demographical details (age, cultural and social groups) of participants, emotion labelling methods, induction context and modalities.

To capture the soft-biometrics pattern for different emotional states, emotion induction and assessment is also a very important factor.

The studies described in [138, 139] report the influence that music audio introduces to a subject's emotional state and also explain how music audio can be utilised as a tool for emotion induction. Livingstone et al [138] aim to formalise the relationship between musical elements and their perceived emotion. The authors develop an architecture for influencing the perceived emotion of music with intent, where this can be used in a gaming scenario, and a real-time background music can render according to the development of the game, instead of a static pre-defined background music. Nakahara et al [139] use heart rate and heart rate variability as indicators to investigated the differences with music-induced emotion while playing music on the piano and listening to the same piece of music recorded. The results suggest that there is significantly greater amount of emotional induction when playing the musical instrument. Cardoso et al [140] propose a discrete approach to emotion assessment, the Circumplex Affect Assessment Tool (CAAT), which is an approach designed to be an alternative of a discrete emotion selection approach (selecting from a list of emotions), where the emotions are categories as shown in Figure 1.1 ,therefore providing a sense of level emotion's intensity. This approach has shown a viable alternative method for emotion assessment.

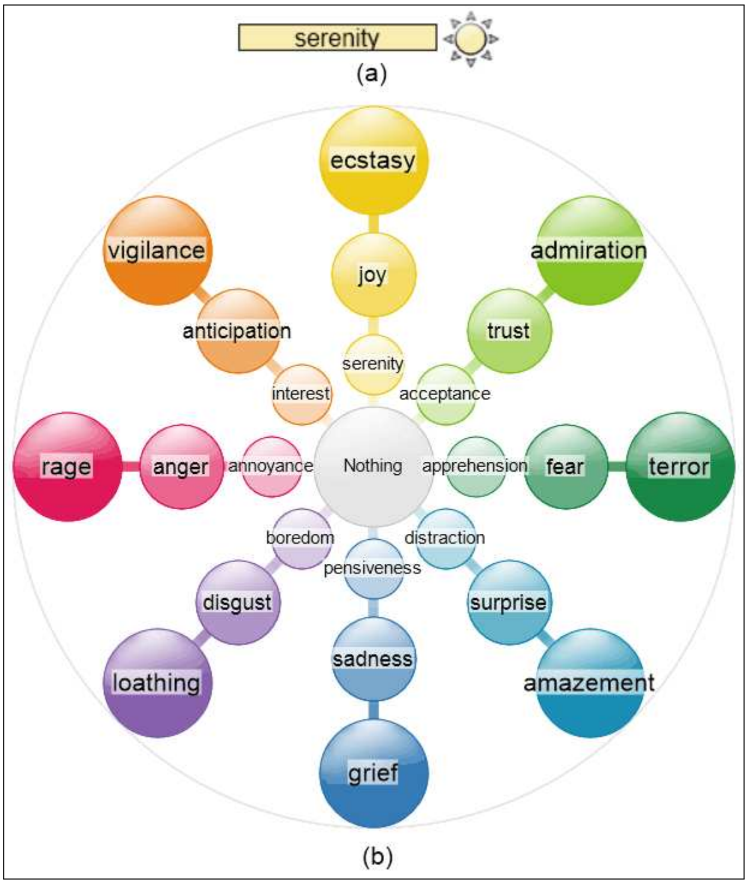

FIGURE 1.1: The CAAT tool in (a) closed state and (b) open state, featuring reversed intensity ordering on the axes, taken from [140]

Mower [141] presents a framework to predict emotion using a paradigm based on “emotion profiles”. This approach, instead of assigning a more conventional single hard emotion label, interprets human emotion expression by providing multiple probabilistic class labels.

It is clear that the number of existing studies reported for higher-level state prediction are considerably fewer than is the case for the work reported for the prediction of lower level characteristics. Nevertheless, the reported studies have provided a concrete foundation for emotion prediction with soft-biometrics. There are a number of soft-biometrics modalities which have been studied, and some typical examples include:

- Face: Similar to the case for the conventional identification task, face is also one of the most studied modalities for emotion recognition, although here it is actually the computation of facial expression which is used as the principal determinant of recognising an individual's emotional state.

  Adolphs [142] found that recognising facial emotional characteristics draws on multiple strategies, for example: recognition as part of perception which recognises the difference between two emotion stimuli and recognition via the generation of associated knowledge where emotions are recognized with past experience. The authors demonstrate that “emotion recognition is not monolithic but consists of a diverse array of strategies and processes”.

  Li et al [143] attempt to determine facial expressions with a “single-image-based” approach for face recognition, where the system is given an expressioned testing face and finds the given face in a face gallery with 15 individuals providing 90 expressioned face images. The authors recognize that the performance of their proposed method can improve if a more specific database can be captured.

  Tsihrintzis et al [144] attempt to improve the emotion recognition accuracy by including an additional modality (keystroke information). The results presented suggest that two modalities can achieve different recognition accuracies depending on the emotion in question, but the two different modalities can complement each other.

  Yan et al [145] present an approach which uses an adaptive discriminative metric instead of the more conventional simple Euclidean distance metric when recognising facial expressions, in order to increase the effectiveness of characterising the similarity/dissimilarity of facial images.

- Voice: Voice is a modality through which individuals commonly, and involuntarily, often express emotions, and the authors of [146, 147] focus on this behavioural trait. Schuller et al [146] use speech features including pitch and energy and semantic and intention based features include wording, degree of verbosity, temporal intention and word rate, with the history of user utterances. The system shown a prediction accuracy of more than 80% with data captured from 15 participants.

  Ciota [147] shows some promising emotion recognition results. However, the main difficulty that the author reports is the acquisition of a proper base of voice samples representing different emotions. This is also one of the main challenges for any emotion prediction study, which will be considered further below.

- Keystroke dynamics: The main method of how individuals interact with any computer system, which includes PCs, laptops and hand-held mobile devices, is still through typing activity. For the past two decades, typing activity has become a significant part of everyday life for most individuals. There are a number of studies which have attempted to investigate this type of behaviour to explore the possibility of determining human mental state and, in particular, human emotions.

  For example, Zimmermann et al [148] measure mood based on monitoring mouse and keyboard activity. The authors discuss different emotion capture methods, which include *Physiological signals, Behavioural methods* and *Self-reports*. Physiological signals require many sensors directly attached to the user's body, which is not suitable in many situations, and it can also require relatively expensive equipment and technical expertise. Behavioural methods normally are tested on “produced” emotional data, such as facial samples from actors. The author of this paper recognize that self-reports are still the primary emotion assessment method. The reported study uses six film clips to induce different emotions and uses a graphical self-assessment Manikin questionnaire (as shown in Figure 1.2) for users to rate their emotional state.

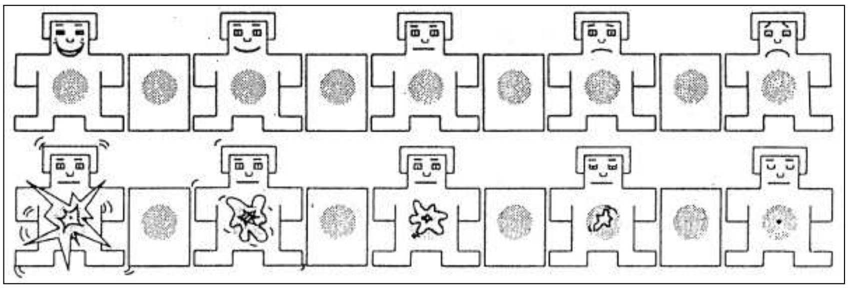

Figure 1.2: The scales valence (top) and arousal (bottom) of the Self-Assessment-Manikin (SAM), taken from [148]

Alepis et al [149] present a system predicting student emotions with data captured from keyboard and microphone. The authors categorise students' input by considering two types of actions: keyboard actions, including three typing speed, same(normal), higher or lower than the usual typing speed of the user, backspace key frequency, unrelated keys usage or no keyboard usage, and microphone-captured actions including, the user speaking strong language, user using exclamations, user speaking at a higher or lower volume than the average recorded level, the user speaking a pre-categorised word which indicates an emotion, and no speech. The combination of action features from both the keyboard behaviour and speech are analysed using the multi-criteria method of Simple Additive Weighting(SAW). They use the likelihood formula to calculate the likelihood of an emotion.

Lv et al [150] present a novel approach to recognising emotion by adopting a modified keyboard that can also collect pressure data for each key press, and the authors [154][156] focus on stress detection from keystroke data. These are the most relevant studies to the core investigation of this thesis, and thus will be presented in detail in the next section.

Santos et al [151] use mathematical exercises during data collection by asking the participants to perform exercises and type what emotion they were feeling during the task. They ask the participants to complete a SAM scale (similar to Figure 1.2), then using the keystroke data to detect emotion. The SAM scale scores are processed by an expert to identify the participants' valence. A preliminary result of around 60% accuracy in predicting emotion was obtained in this study. The authors also discuss their ongoing work of combining multiple data sources from video, facial action, heart and breath parameters, keystroke and mouse movement, and eye tracking. The end result of their ongoing work might be able, they claim, to provide a clearer picture of what the predictive performance will be when combining multiple data sources.

Kolakowska [152] reports a survey based on keystroke data and mouse movement data, to recognize the user's emotional state. The author summarizes the key studies and produce a Table containing the methodology and results of these studies. The review shows the importance of the data collection process and suggests that the data collection should try to avoid users becoming accustomed to the same input environment. The review also suggests that intentionally induced emotions may cause a different reaction than natural circumstances do. These valuable points will be considered in our study, as will be discussed in more detail in later chapters which describe a new data collection process and analytical approach.

Bixler and DMello [153] focus on boredom prediction and engagement during interaction via keyboard use, and their study asks participants (students) to write essays on three different topics as part of their course credit. The three topics are "*(a) academic topics which were obtained from the writing portion of the American College Testing (ACT) test, (b) socially charged issues like abortion and the death penalty, and (c) personal emotional experiences such as recent angry or happy experiences*". The authors suggest that these topics can reflect their expected emotional impact. The whole process of writing the essays are recorded including participants' facial expressions. The recordings are then played back to the participants. At each 15-second interval, the participants were asked to determine their emotional state for the moment by selecting one out of a list of fifteen affective states, which include "*anger, contempt, disgust, fear, happiness, sadness, surprise, boredom, confusion, delight, engagement, frustration, anxiety, curious, and finally, neutral*". Keystroke features used include the commonly used timing feature latency, pausing behaviour features which include the number of pauses for each pause length category, keystroke verbosity features including the number of keys within each interval and number of backspace used within each interval. The study collected data from 44 participants and the authors used the WEKA data-mining tool for supervised classification. The best performance achieved was a predictive accuracy of around 87%. The authors believe that free text, where the content participants are required to type are not pre-defined, can provide a more accurate representation of real world situations.

Tsui et al [154] combine facial image feedback with keystroke data for emotion prediction. The authors investigate the difference of the keystroke data between positive and negative emotional sates. The study discusses two emotion inducing methods. The first method uses picture, film or music and the second method uses emotional behaviour as stimuli. The authors suggest that when a person is forced to smile, that person then becomes happy. Facial feedback is used as an emotion induction method to make sure the keystroke patterns are collected under different emotional states. Facial feedback includes forcing the participant to hold a pen with their lips and teeth, where the authors claim when holding the pen with lips induce a negative emotion and when holding the pen with teeth induce a positive emotion. The studies collected data from 15 subjects and results suggest that there are differences in the typing patterns under positive and negative emotions. It is arguable whether such a small database allows such conclusions to be drawn.

Bakhtiyari and Husain [155] combine keystroke, mouse and touch-screen interactions for emotion prediction. The authors identify three problems usually ignored for emotion prediction studies reported in the literature, which are that human emotions are not constant over a period of time, human beings may have multiple emotions occurring at the same time and psychological studies show human emotion can change differently based on different cultural and language backgrounds. The study adopts a model that classifies the emotion into five levels (designated 0 to 4), ranging from 0 representing 0% of the emotion while an assigned level of 4 represents 100% of the emotion. 27 different emotions were available for the participants to choose and eight basic emotions were studied. An SVM classifier was used for the emotion prediction task and a best 16.7% false positive rate was achieved.

Predictive results reported in the literature for emotion prediction by using keystroke data vary considerably, with accuracies ranging from around 60% (for predicting valences [151]) to around 90%(for predicting happiness [150]) when a database including key pressure sequence information was used.

- Handwriting: Studies that use the analysis of handwriting and drawing movements in healthy subjects for medical diagnosis have been reported, exemplified by [162] for example. Handwriting has also been shown to be a good indicator of a writer's mental state. Mutalib et al [156] collected 30 scanned handwritten samples and feature are generated based on baseline detection. A set of 33 fuzzy rules are used for emotion control prediction. For example, one of the rules is if baseline is straight then emotion control is medium. Kedar et al [157] review emotion recognition through scanned handwriting samples. Six different features are presented for emotion recognition, which include slant (the angle of the upward or downward strokes) of the handwriting, baseline (the line along which most of the letters are written), pen-pressure (the force applied while writing, detected by the thickness of the handwriting strokes), size (size of the letters), margin (the margin around the handwriting), zone (the area that the handwriting sample covers). The authors in the reported studies believe that these are indicative of different emotions of the writer.

It is apparent that very limited extensive higher-level emotional state prediction studies have been carried out, and this emphasises the potential value of one of the principal aims of the present study.

With the predictive capability of the biometrics modalities present above, it is obvious that the potential applications for predicting human higher-level state can and have already been applied to help solve various real world problems. In fact, soft-biometrics information prediction, in its widest and most general sense, has proved beneficial and successful in many real life applications. The authors in [158–160] report work which uses image processing and body data collected from pregnant women for prediction/estimation of gestational age. Tranter et al [161] report a study examining the changes in emotional processing by using a “face emotion recognition paradigm”, where the patients were asked to respond as quickly as possible to recognize the facial expressions which include facial expressions from six basic emotions: happiness, sadness, fear, surprise, anger and disgust, to determine which treatment works more effectively. For the study the authors recruited 173 depressed patients attending a GP practice in North West Wales, where 88 randomly selected patients were using citalopram (antidepressant drug) and the other 85 were using reboxetine (antidepressant drug).

It is therefore readily apparent that medical practice can benefit considerably from this type of work. Bidet-lldei et al [162] use handwriting segments from patients with Parkinson's disease to evaluate their condition in order to find ways to determine how their handwriting patterns differ from healthy participants. The authors found that patients with Parkinson's disease are less able to write continuously. However, this does not confirm the hypothesis that “Parkinson's disease patients do not anticipate future movements”. O'Reilly and Plamondon [163] explore the biomedical potential for using handwritten signatures to predict medical information that can reflect the effects of stroke.

Human computer interaction, also known as HCI, is an area also of particular interest in the context of this topic. Mandryk et al [164] report studies that test the efficacy of physiological measures to evaluate subjects' experiences. Kapoor et al [165] attempt to predict the degree of frustration experienced by subjects to support learners during the learning process, and Mandryk and Atkins [166] focuse on determining users' emotional state while they play computer games. Hong et al [167] report a study to develop an understanding of a subject's physiological status and uses the information from physiological status to predict the degree of stress being experienced. Vizer et al [168] detect both cognitive and physical stress by monitoring keystroke activities from the user. The authors suggest that the early detection of changes in cognitive and physical states can allow interventions that can treat functional impairments.

Other than collecting data from conventional sensors, Gu et al [169] use a variety of physiological signals to predict emotion. Physiological characteristics of interest in this respect include temperature, muscle electrical activity, respiration, cardiac function, skin conductance and electrical activity of the brain. Liapis and Michalis [170] also collect physiological measurements including heart rate, sweating (skin conductivity), muscle tension, and respiration rates to assess the emotional experience of users.

Following the trend of the rapidly expanding gaming industry, Sykes and Brown [171] focus on detecting the force of button-pressing on a game-pad and use this pressure data to measure and predict the affective state of the players, and Gao et al [172] study the data collected from a mobile phone touch screen while the user plays games on the phone. Captured “screen touching” data which are based on touch behaviour(finger strokes) of the player, which are measured by length, pressure, direction and speed of the strokes. User emotional states were captured at the end of each game level with a self-assessment questionnaire. The author suggest that the result demonstrates this technique can be used to monitor users' emotional state.

Therefore, it is clear that further development of more robust techniques and methods are more important and essential than ever, since this would allow more reliable and accurate prediction to be made for both conventional soft-biometrics information as well as higher-level mental state information. This is especially the case, given that biometric data and information collection are becoming increasingly common in everyones daily life. It is equally clear that within the field of soft-biometrics information prediction, biometric data and information collection is an important area, which is still relative limited, and this suggests that there is significant amount of fundamental work still to be done.

As presented in this section, there are a relatively limited number of detailed studies found on exploring higher-level predictive biometrics , and especially a lack of investigation for emotional state prediction. It seems likely that this is an area which will become more and more important in the near future, for example by helping to fulfil the target of future human computer interaction, which is to create seamless intuitive interaction between user and computer systems. In this sense, to improve techniques which can reliably and correctly determine human mental state may be seen as the foundation of creating future computer systems that have a greater awareness of users physical and mental states and thus the ability to make appropriate dynamic adjustments during interaction.

It is not practical or effective for the study at the focus of this thesis to investigate the predictive capability of all the biometric modalities. We have to be selective in terms of the biometric modalities that have been chosen for study. Therefore, keystroke dynamics and handwriting, which are both less extensively reported modalities of behavioural biometrics have been chosen. Both are hand-oriented behavioural biometric modalities, yet they present rather different activities, and will provide a good basis for comparison. Handwriting is of particular interest as there has not been an exploration of emotion prediction in this modality which differentiates the use the static (related to the visual appearance of the handwriting) and dynamic (related to the execution sequence of the handwriting) features generated from data captured using a digitiser tablet. Therefore, the next section will focus on a more detailed review of methodologies used for emotion prediction using these modalities.

1.3 Using hand-oriented behavioural biometrics for "higher-level" mental state prediction

It is suggested that keyboard-based behaviour and writing-based behaviour are the most dominant current methods of interaction between most humans and computer systems, and most human to human interaction as well, and are therefore good candidates for study. They share a common property in both being hand-based biometrics, yet they are sufficiently different at a functional level to allow the investigation of techniques which cross the boundary between different modalities. However, both methods are still two of the least explored behavioural biometric modalities for higher-level mental state prediction studies, and are also therefore worthy of more detailed attention. Therefore, in this thesis, our target is to better understand and to extend the predictive capabilities of these two hand-oriented behavioural biometric systems.

As explained in the previous section, this section will provide a more detailed review of hand-oriented behavioural biometric for emotional state prediction.

The study reported by Epp et al [173] adopt a 5-point Likert scale [174] for emotion labelling, where users are given a scale of *strongly disagree, disagree, neither agree nor disagree, agree, or strongly agree* to rate themselves for a specific emotion, the emotion labels generated by grouping *strongly disagree with disagree* and *agree with strongly agree* therefore created three classes, agree, neutral, and disagree. The study uses background key event collection software to record the keystroke data. The study developed a database, but one which included only 12 participants, and the data were collected by running the collection in the background during the participants' daily tasks. Therefore, all the data collected were free-text, which means the content of the typing was not predefined. The feature used are all timing features developed from investigating digraphs and trigraphs. The experiment did show encouraging results, suggesting around a 77% predictive accuracy for recognising emotions including confidence, hesitancy, nervousness, relaxation, sadness and tiredness, but the limitations of the database size obviously make strong conclusions difficult.

Vizer et al [175] also employ key event collection software for keystroke data collection, the keystroke timing data was recorded to the resolution of 10 milliseconds, and features generated include: timing features (for example, time per keystroke), keystroke feature (for example, backspace key rate), and word feature (for example, noun rate, verb rate). A 11-point Likert scale was adopted for emotion labelling at the end of the entire experimental session. The experiment was carried out while the participants were under a controlled environment (in a lab with supervision) and collected both fixed (predefined) and free (unconstrained) text. Using a database only slightly larger than that noted previously (here, 24 participants), this study reports an achieved recognition rate of around 75% for predicting cognitive stress.

Taking a rather less standard approach, Lv et al [150] introduce a pressure sensor keyboard and records both the key events and pressure sequences. The data are labelled according to emotional state, by assuming that each emotion of interest is automatically linked to and generated by the specific task content presented to the participants, which may be seen as a significant area of weakness. This study developed a database consisting of keystroke data samples from 50 participants and commonly used timing features (key down time and key up time) with a novel pressure sequence feature generated an error rate of around 14% for “Happiness” and “Sadness” prediction.

Hernandez et al [176] record both pressure sequence and key events and mouse movements data in the database. This database again consists of only 24 participants. Emotional state scores are collected at the end of each task based on a 7-point Likert scale(the scale ranging from “very stressed” to “not stressed”). This study provides three tasks for the participants to complete, which include: Text transcription task (Fixed text task, where a predefined piece of text was provided to be typed), Expressive writing task (Free text task, where the participants were asked to write about a stressful experience and a relaxing experience), and a Mouse clicking task. This study reports that around 83% of participants showed increased physical keyboard pressure when put under stress.

One of the key issues in this area, as can be observed from the literature, is that there are no publicly available databases which provide the range of data required to carry out investigations of the predictive capabilities of biometrics-derived data in relation to the higher-level states of interest. Each study has to design and implement a new database. As organised and displayed in Table 1.1, four of the main databases reported which contain keystroke data and emotion labels show various numbers of participants (Often not adequate for reliable experimentation), labelling settings (different types of Likert scale, or completely

Reference	Keystroke data capture method	Number of subjects	Emotion capture method	Emotion prediction result
[173]	Background key event collection software	12	5-point Likert scale	77.4% - 87.8% for hesitation, nervous, relax, sad or tired
[175]	Key event collection software	24	11-point Likert scale	75.0% for cognitive stress
[150]	Key event and pressure sequence collection software	50	Content determined emotion	86% - 92% for neutral, anger, fear, happiness, sadness, surprise
[176]	Key event, pressure sequence and mouse collection software	24	7-point Likert scale	83% for level of stress

TABLE 1.1: Keystroke databases developed for emotion analysis.

different emotional state assessment methods). The reported studies are also seen to share a common problem, which is the limitation of the size of the databases, the smallest database consisting of data captured from just 12 participants. This limitation presents obvious potential difficulties when attempting to draw any strong conclusions from the experiments carried out using such databases.

There have been many examples of public interest in applying emotion prediction to real life applications. The following applications represent a variety of disciplines. The authors of [177] developed an e-Commerce system that can detect users' emotional state by using a combination of user's speech data, facial image, motor functioning and body gesture. [178] monitors the emotional state of students by using a mouse-based system, while [179] uses a combination of keystroke and mouse dynamics to estimate affective state and communication preference in order to improve employees' efficiency. In [180], the authors use social media data collected from mobile applications to detect user emotion. The study presented in [181] uses keystroke dynamics to detect a student's mental state in the context of automated tutoring, when students are interacting with software-realised virtual "intelligent tutors".

As presented in Section 1.2 and Section 1.3, there are great opportunities for further study of the ability of predicting lower-level and higher-level biometric information with biometric data. Successfully predicting the gender of a writer or a typist by their behaviour pattern can help screen subjects in a database. The gender information can be used in, for example, forensic investigations when a list of suspects need to be narrowed down, or extra layer of security when password or signature are used as part of a user verification system. The emotion information can be used in many disciplines. For example, in education, where students' emotional state can be monitored during their learning process, the emotional information predicted can be used to optimize the teaching methods and students' learning experience. Predicted emotional information can also be used as a user experience monitoring tool, for example, a mobile phone application can use their users' typing behaviour during their interaction with the application to determine how well each user are feeling when using each functionality, and then use the feedback to upgrade their application's usability. There are also possible healthcare applications when mental state diagnoses can be utilized by this technique to more effectively access patients' mental state. Even these few examples show that there are many domains of real life applications which can benefit from successfully predicting higher-level soft biometric information.

1.4 Research problem and motivation

Before undertaking a detailed exploration of the topic, it is important briefly to consider also the following issues of key relevance to soft-biometrics information prediction, which need to be further discussed and analysed.

1.4.1 Databases

For almost any biometrics-based studies, the database adopted for experimental investigations is a key component, and a well-designed and only a well-executed data capture process can ensure the robustness of the results generated and the analysis and evaluation of those results. There are a considerable number of popular publicly available databases which can be found (at least for a small number of modalities), but these databases are mainly implemented to support mainstream identification or authentication tasks, and are therefore generally unsuitable for, or typically only support a small range of predictive opportunities.

Databases for emotion/mental state prediction are both rare and, as the earlier review has shown, usually limited in size. Taking keystroke databases, for example, as shown in Table 1.1, the size and number of databases developed are very limited and no publicly available databases were found to be available. However, the creation of appropriate databases, especially if they are to include a large number of participants, requires a great effort, time and resources (including both financial and human resources) to implement. Perhaps this explains in part why the majority of emotion prediction studies reported for the keystroke modality are based on very small sample sizes, which have the potential risk of undermining the generality of conclusions drawn from the experiments. A principal contribution of the study to be reported here is the creation of a new database twice the size of the largest of those hitherto adopted in studies of this type.

1.4.2 Ground truth determination

It is clear that in order to achieve a robust result, the data on the emotional state of each participant need to be labelled reliably and accurately. Some soft-biometrics information, such as age or gender, are relatively straightforward to determine in most cases. However, such ground truth labelling becomes a significantly more challenging task than when investigating relatively less clear-cut characteristics such as age, gender and so on, other characteristics such as the higher-level mental states that are the focus of the present study, are less easy to label accurately. Consider, say, the emotion "Happy" as an example. Even defining a concept such as "happiness" is itself a difficult task, let alone attempting to quantify an index of "degree of happiness". As reported in Section 1.2 and Section 1.3, various approaches have been attempted to generate accurate ground truth labels in these circumstances.

Again taking keystroke as an example, as shown in Table 1.1, the authors in [150] label users emotions based on making assumptions which are directly linked to the task content, which can potentially create some fundamentally inaccurate labels. This is a challenging task and the potential problems can only be minimized by adopting a more appropriate method of emotion assessment, paired with an effective capture procedure.

1.4.3 Task definition and feature analysis

Many predictive studies involve the testing of individuals participating in some specific activity or task in order to capture their responses. Emotional state prediction is likely to be significantly task-dependent, but within this area of study there are no agreed guidelines about the methodology of defining tasks, yet the predictive performance may well vary considerably depending on the data collection environment.

Therefore, it is essential to understand how to define the tasks as well as possible, with well-designed content and execution instructions in order to capture the data across a range of appropriate behaviours, thereby maximizing the opportunity to choose an "optimal" feature set.

Another important point to consider is that some of the feature sets conventionally used directly in this area of work (or with only slight variations), yet most of these features were initially developed primarily for identification proposes. It is therefore advisable, and perhaps necessary, to explore the possibility of developing new features to observe the potential impact that potentially more appropriate features might have on the task of predicting higher-level states.

1.4.4 Optimizing processing infrastructure

The data processing infrastructure is another key component when the exploring the predictive capabilities of any biometric modality. It is clear that the predictive accuracy achievable can vary depending on the data processing algorithms adopted. It is necessary to explore a range of classifier designs in order to find out which classifier might provide the best processing engine for the purpose of predicting higher-level soft-biometrics information. Taking the studies [182] and [183] as illustrative examples, the idea of adopting a more flexible and "intelligent" data processing system which is based on a multi classifier processing configuration has shown some promising results in improving performance in conventional identification tasks. Hence it is important to determine whether these techniques and this type of processing novel system reported could also bring benefits in tackling soft-biometrics information prediction tasks.

In summary, then, the aim of the work reported in this thesis is to extend the predictive capability of hand-oriented behavioural biometrics, specifically using handwriting data and keystroke data. The investigation of both modalities ideally requires a comprehensive database where the data are collected with a consideration of the issues mentioned above. There is not any publicly available database which can provide the necessary data for extended prediction tasks. Therefore, not only is there a significant potential value in collecting an enhanced database that is specifically designed and implemented for the purpose of extended higher-level soft-biometrics information prediction, but it is absolutely necessary in undertaking the proposed study.

The experimental work and analysis undertaken will be presented in the remaining chapters of this thesis. To summarize the principal objectives of the study to be reported, the core messages which this thesis will address are as follows:

- Preliminary analysis of higher-level prediction studies in order to review the current adopted data collection methods and performance achievable as reported in the current literature.
- An investigation into the development of suitable data collection principles, in order to collect data under appropriate conditions to capture participants' behaviours in different states.
- Study of higher-level human mental state prediction with an enhanced database, which contains samples for both the handwriting and the keystroke modalities. This will allow an investigation of the achievable predictive accuracies for both handwriting and keystroke modalities.
- Address and discuss the limitations and challenges in different processes of prediction system.
- Consider some likely productive directions for future work.

1.5 Chapter conclusion and thesis organisation

This chapter has presented an overall review of the biometrics information prediction studies that have been reported in the literature. The review has illustrated the need for further development of techniques for soft-biometrics information prediction and has also led to a clearer understanding of the field and brought into focus the some of the key issues to be addressed when dealing with soft-biometrics information prediction. Further, more detailed surveys of relevant literature will also be provided where helpful in the remaining chapters.

The literature survey has focused on the prediction of both “lower-level” soft-biometrics information and “higher-level” soft-biometrics information. The review has presented the less extensively studied “higher-level” soft-biometrics information prediction and identified the need for a more extensive investigation of the predictive capability of biometric data from across the available modalities and, in the context of the present study, has concentrated on hand oriented behavioural biometrics, specifically handwriting analysis and keystroke dynamics. The methods and techniques reported in the previous studies have provided an initial starting point for a development of a more efficient and effective methodology when investigating higher-level soft-biometrics information prediction with hand-oriented behavioural biometrics.

The core objectives of the thesis are investigated in the following chapters and an outline of the overall thesis organisation is as follows:

- Chapter 2: Experimental Infrastructure

  This chapter will cover the design of the data capture protocol for the hand oriented behavioural biometrics database for higher-level soft-biometrics information prediction and, most importantly, will explain the reasons for the decisions made about the details of the protocol. This chapter also reports the software and hardware setup for the implementation of the large hand oriented behavioural biometrics database for higher-level soft-biometrics information prediction.

- Chapter 3: Prediction of gender from handwriting data

  This chapter will report the findings of predicting a more conventional soft-biometrics characteristic (gender). These predictive results fill the gap in the literature for gender prediction with handwriting data captured with a digitising tablet, but also provides a useful comparative study as a basis for the higher-level predictive experiments to follow This chapter also addresses two important issues that could have an impact on the gender prediction performance of handwriting data.

- Chapter 4: Prediction of higher-level states from handwriting data

  This chapter describes the higher-level states prediction results obtained from an analysis of handwriting data across various different tasks and under different experimental conditions. This chapter also presents the procedure for handwriting feature extraction and ground truth labelling. Three of the main areas, which include the labelling method, categories of feature sets and different categories of experimental conditions (tasks), which could potentially influence the predictive performance, are also addressed in this chapter.

- Chapter 5: Prediction of higher-level states from keystroke data

  This chapter presents an investigation of higher-level states prediction with keystroke data, which will cover the procedure of keystroke feature extraction and ground truth labelling. This chapter will also investigate the impact of different feature set parameters on the predictive performance of keystroke data.

- Chapter 6: Other issues in predictive biometrics

  As a useful complementary extension of the core messages of this study, this chapter presents a collection of additional experiments to further explore three principal key issues that are drawn out of the previous chapters to help complete a comprehensive investigation of relevant aspects of higher-level state prediction. This chapter examines the predictive capability of some newly developed non-traditional keystroke feature sets as an illustrative example. An investigation of higher-level states prediction with a novel data processing infrastructure, and an investigation of two novel alternative ground truth labelling methods are also described.

- Chapter 7: Conclusions and suggestions for further work

  This chapter concludes the study and discusses the contributions of the study reported in this thesis. This chapter will also briefly describe a further step that was initiated (specifically, a second data capture process) for future work and also discusses the possibility of taking further advantage of the newly acquired database, such as the combination of both modalities to further extend the higher-level state predictive capability of hand oriented behavioural biometrics.

CHAPTER 2

Experimental Infrastructure

In order to investigate the correlation between hand-oriented behavioural biometric data and mental states (as explained in Chapter 1, here the term "mental states" includes emotional states and other similar characteristics of the subject), an appropriate database is required, which fulfils the following criteria. It should contain the captured hand-oriented (typing and writing in this study) behavioural data under different categories of mental status. The data should be correctly labelled with the participants' mental states (i.e. reliable ground truth data should be tagged to the samples). The data capture should be as intuitive as possible. This requires the collection process to create a typing and writing set-up and experimental capture software, which support easy user-system interaction. At the same time, the study also requires an efficient method to collect the participants' mental state. A self-evaluation 10 point Likert Scale questionnaire (explained in detail in Section 2.1) was adopted for this study. After reviewing the literature in the field of study in Chapter 1, it was established that there was not any suitable database or data collection protocol available to conduct the research proposed for our study. Therefore, it was determined that a new data capture initiative was required in order to provide appropriate data to carry out the analysis that was essential to support this study.

This chapter covers the data capture section, which includes the design of the protocol, execution of the experimental data collection exercise, an initial data analysis and the structure of the database, and finally a description of the experimental specifications.

Section 2.1 will present the design of the data capture protocol, which include detailed specifications for both handwriting and keystroke capture protocol implementation. Section 2.2, will present the software and hardware setup for the data capture. Section 2.3 will present the data collected. Section 2.4 will make some concluding observations.

2.1 Data capture protocol implementation

2.1.1 Design of the protocol

The data capture process is a fundamental building block of the study, as the analysis makes use of the data acquired to develop algorithms and methods to explore one of the central questions addressed, namely “To what extent can mental status be predicted based on hand oriented behavioural biometrics?” A data collection protocol was designed and developed to provide a definition of the type of data to be collected, together with a description of the methods and infrastructure necessary for the configuration of the acquisition process. There was no suitable keystroke or handwriting databases made publicly available in the emotion prediction studies reported in the literature, and all the publicly available keystroke and handwriting databases available were missing the key data, specifically the emotion labels corresponding to the mental state of the subjects at the time of donation. Therefore, in order to ensure that the data captured can sufficiently serve the purpose of our study, the task of designing the protocol must be considered very carefully. A thorough study of the available databases and their data capture methods was conducted as described in Chapter 1 and most of the reported studies share the same core idea for their experiments, which is to capture the participants' interaction behaviour within an intuitive environment, while at the same time keeping control of the experimental parameters. The idea of an “intuitive environment” refers to having a set-up where the participants can perform the tasks in the way they normally would type or write. This is key, because it is essential to capture their natural behaviour pattern in order to correctly predict their emotions.

Two different approaches for the overall procedure of the experiment were considered for the protocol.

- Multiple sessions approach: Used in [173], this approach requires the same participant provide multiple data capture sessions at different circumstances (different time/date). The advantage of using this approach is that it can support separate samples from the same user from multiple sessions, which means that the results can be verified by using data captured from different sessions for each participant and therefore cross-examine the conclusions we can draw out of the results. The principal difficulty is to maintain a uniform level of mental status activation in each session, which is one of the essential requirements for our experiment.
- Single session with multiple tasks approach: Used in [148–151], this approach only requires the participants to attend a single data capture session to complete a series of tasks. This approach provides the option to

influence the participants' mental status on a continuous basis. Each task acts as a checkpoint to capture the flow of different aspects of the mental status. Therefore, this approach makes it possible to capture any change in the mental condition and also to finish the experiment in one session and ensure that every participant experiences broadly the same degree and type of mental stimulation.

After carefully considering both methods, the second method seemed to be more suitable for the proposed experiment. It is very difficult to maintain the level of mental status activation the experiment provides in multiple sessions. The second method is also commonly adopted in the relevant studies reported in the literature.

Our data collection process consists of two related strands: keystroke capture and handwriting capture. Both parts were designed based on the core principle of placing the participants in an identical, consistent, intuitive (in the sense that how they would normally write and type) environment during their data capture session.

There are two types of tasks that the experiment contains, fixed typing/writing tasks [148, 150, 173] and free typing/writing tasks [149, 151, 173]. Fixed tasks are those where the participants were given predefined words to copy. Free tasks are those where the participants were asked to type/write unconstrained content without any limit on length (although such tasks are guided in terms of the broad content, in order to give a context for the required response). Both types of tasks are included in our experiment, to ensure the richness of the data captured.

The overall keystroke capture process encompasses five tasks and the handwriting capture process involves four tasks. Each task serves a particular purpose in the experiment with different interactive content for the participants (which potentially, in some cases, "nudges" participants towards a particular emotional status, as will be explained later). Therefore, we can observe the changes in the participants' mental status and influences on their typing/writing pattern. Each task will be described in detail in the following section.

All the data were collected anonymously by assigning each participant a unique ID number at the beginning of data capture session, which links the data capture to their demographic and emotion data. The demographic data were collected by means of a questionnaire which included categories for a subject's Age, Gender and Handedness. The participants were asked to highlight the correct category they are in and a 10 point Likert scale style questionnaire was adopted for emotion data acquisition. The participants were asked to fill in the latter questionnaire in between tasks to reflect their state of mind at the time of commencing each task. The emotions that were assessed include happiness, relaxed, bored, confident and curious. For examples, on the Likert score where 1 on the scale signifies “completely not happy/ not relaxed” and 10 on the scale signifies that the participant is “completely happy/relaxed”. In the study reported in this thesis, the investigation is mainly focused on happy and relaxed emotions, collection of scores for other emotions avoid the participant only anticipating the experiment only focused on two of the emotions, also this provide data for further investigation of a wider range of emotions. The keystroke data capture software was purpose - written for this particular study, and was thus designed, implemented and developed from scratch. The handwriting data acquisition tool was designed internally within the School of Engineering and Digital Arts at University of Kent. Each data capture session, which include both keystroke and handwriting, was supervised while the participants were using the capture software. During the capture session, in order to maintain the consistency with the verbal instruction given by the supervisor, a verbal script was developed for the experiment to ensure the consistency when providing instructions to the participants. For the keystroke data capture tool, the interface was designed to be clear and uncomplicated. All the participants were able to interact with the software with minimum effort. The handwriting capture tool was operated by the supervisor while the participant writing on the digital tablet. The experimental set up will be covered in detail in Section 2.2. As noted above, the core foundation of the experiment is to set the participants in an identical, consistent, intuitive environment.

After successfully completing an initial testing phase, which include testing both capture tools and hardware, it was concluded that the system is working appropriately. The recruitment of the subjects was initiated within University of Kent School of Engineering and Digital Arts. In order to ensure that the experiment could be carried out correctly in the actual data capture session. The three participants on the first day's sessions were used as “evaluation” subjects. After all the data captured from those three participants passed the inspection, where a series of checks were performed on all the capture equipment and software to make sure the system was working as expected and the data was correctly obtained. Then the recruitment carried on within University of Kent School of Engineering and Digital Arts and also went on to the whole university. People from outside the University of Kent were also recruited to avoid the database predominantly(only) containing university students and staff.

For the actual capture of data, the procedure was as follows for each subject. The subject was introduced into the room where the capture equipment was set up. The session began with the keystroke data section, and then at completion the subject was instructed to use the handwriting tablet to carry out the handwriting section. Each capture session, which include keystroke and handwriting sections, required on average around 45 mins with the shortest session took around 25 mins and the longest session lasted for around 60 mins depending on the speed of the participants.

The set-up for both handwriting and keystroke data proved to be completely appropriate to meet the experimental requirements. The data capture for the 100 targeted population of participants was completed successfully within 2 months, this time-scale requiring a significant effort, involving the recruitment of such a large number of volunteers, and managing the collection timetable for the sessions, as well as the actual capture of the data for the whole range of tasks.

2.1.2 Keystroke capture specifications

For keystroke data acquisition, five tasks were delineated which contain various materials. The specifications were chosen to promote a range of messages sent to the subject, conveying “Happy” to “Not Happy”, “Relaxed” to “Not Relaxed” notions. However, the responses from the participants are entirely individual. The objective is to place the participants in different intuitive environments and record their actual behaviour patterns during the typing tasks.

The main difficulty of the design is to put all the participants under a consistent and intuitive typing environment throughout the entire experiment, because the goal is to capture how he/she would normally behave on the keyboard. Therefore, even the smallest detail of the experiment could have an impact on the participants during the capture session. In this section, each task is explained in detail individually, and the description below includes the design of the task and the logic behind why the task was designed in a specific way. Each task required a number of iterations for revising and readjusting to ensure that the experiment achieved our goal as far as possible.

- Task 1: This task contains three sections. In the first section, participants were asked to type the following predetermined list of words.

  "*time , be , good , to , the , actress , header , person , have , new , of , and , another , shotgun, year , do , first , in , that , turn , polarised , way, say , last , for , it , those , comment , day , get , long , on , not, hesitate , talent , thing , make , great , with , he , rent , google , man , face , little , at , as , health , arched, world , know , own , by , you , illusion , hotmail , life , take , other , from , university , snowman , regards , hand , see , old , general , internet , balances , youtube , education , copy , further , rest , need , range , facebook*"

  The words were chosen to cover all the letters of the alphabet. Therefore, it provides a basic template of keystroke features for the whole alphabet for subsequent analysis. Then, in the second section, participants were given the following paragraph to type.

  "*Steve Jobs who started Apple in his parents' garage when he was twenty years old was once fired from Apple at the age of thirty. This rarely mentioned story was found in his 'Stay Hungry. Stay Foolish' speech for Stanford University: I didn't see it then, but it turned out that getting fired from Apple was the best thing that could have ever happened to me. The heaviness of being successful was replaced by the lightness of being a beginner again, less sure about everything. It freed me to enter one of the most creative periods of my life.*

  During the next five years, I started a company named NeXT, another company named Pixar, and fell in love with an amazing woman who would become my wife. Pixar went on to create the world's first computer animated feature film, Toy Story, and is now the most successful animation *studio in the world. In a remarkable turn of events, Apple bought NeXT, I returned to Apple, and the technology we developed at NeXT is at the heart of Apple's current renaissance. And Laurene and I have a wonderful family together.*"

The text was chosen and edited so that it contains all of the most common digraphs in English. Therefore, it provides a comprehensive template for the further analysis of data provided under different emotional states. As the final part of Task 1, participants were given a series of numbers to type, which covered all the numbers from zero to nine in a random order.

The principle behind this task overall is to get the participants familiar with the acquisition software user interface. The content was chosen to be neutral, in order to minimize the possibility of causing any unintentional external influence on their mental status. However, although the content does not convey any strong emotional message, it is acknowledged that participants can have their unique individual responses to the materials presented, although such effects are expected to have minimal influence generally.

- Task 2: Participants were asked to type in a description of an escape route from a maze, which was presented to them visually. The maze was in a grid format with a yellow arrow representing the entrance and a blue arrow representing the exit, as shown in Figure 2.1.

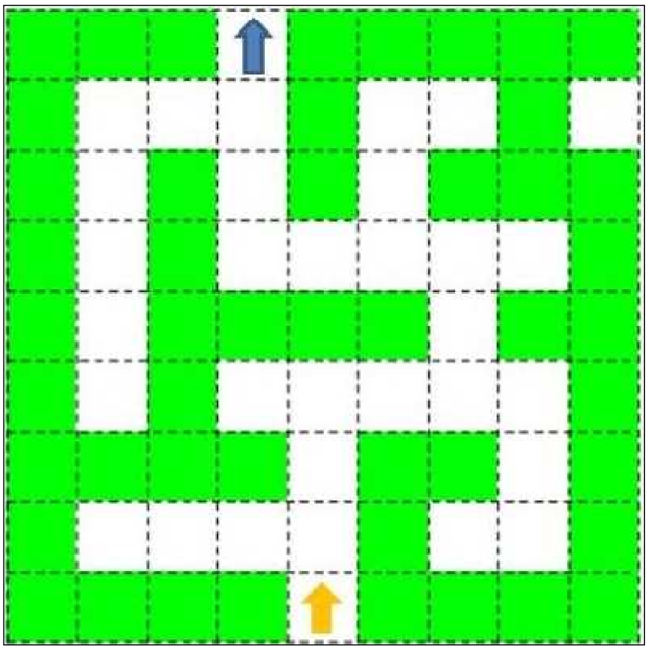

Figure 2.1: Keystroke experiment Task 2 - Maze

This task is the first "free text" task, which means that instead of asking the participants to copy a set of predefined textual symbols, they can choose their own words to describe the escape route they will take and any other text they feel necessary. This task aimed to deliver a different type of intuitive environment to the participants, as what they can type was not limited. The picture was only a guideline for the participant to initiate their thinking process.

This task again was designed to be neutral, giving the participants a chance to connect their thinking with their typing behaviour, and performing naturally. It was not designed to impose or suggest any particular mental status, and thus no predetermined mental/emotional state is assumed.

- Task 3: Participants were presented with a short visual clip of an animated sequence and asked to watch it as many times as they wanted. After watching the clip, they were asked to write a paragraph about what they had seen. This could be a simple description of the clip or anything else they felt like writing after watching. The length and the content of the paragraph were not prescribed in any way and were thus entirely determined by the subjects. The general appearance of the video clip is illustrated in 2.2.

FIGURE 2.2: Keystroke experiment Task 3 - Video Task (A few screen shoots are presented in this figure to demonstrate the content of the video)

For this task, the video was chosen to "nudge" the participants towards a positive mental state as the clip imposes a more positive emotional message in comparison to the previous tasks. The clip appears to convey a "Happy/positive" message. However, we do not make any assumptions about how the participants would respond.

- Task 4: Participants were presented with a series of four related images in the form of a cartoon strip and they were asked to write a description, again for which the content and length was not prescribed, in the same way as for Task 3. In this task, however, the picture presented was chosen to convey a rather more negative (less happy) message.

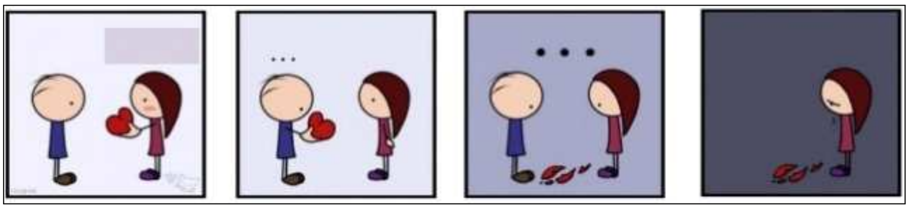

FIGURE 2.3: Keystroke experiment Task 4 - Sequence of pictures

This task was designed to create a contrast with Task 3. The sequence of pictures suggests a less happy story. The participants still have the option to type anything that they please, which says something about what they have seen (as for Task 3). The content is to be related to the picture. This task was designed with the aim of "nudging" the participants to a less positive mental state, using the principle explained above.

- Task 5: Participants were given a predetermined sentence to type, but in this case within a specific time limit (10 seconds). To emphasise this, a counter is displayed, which starts running down as soon as the user types the first letter. In addition, a coloured light changes its colour depending on the time left on the timer. The task was designed so that most participants would struggle to complete the whole sentence within the set time limit. Thus, this task is aims to create a typing environment where the participants were under a certain amount of pressure.

This task introduces additional dimensions to the possible influence on the participants' possible mental status in addition to the previous tasks. Specifically, it enables us to capture the behaviour pattern while the participants were under a variety of pressured situations - time running out, maintaining the correctness of the word they were copying, visual pressure (timer and changing light) and so on.

Thus, the keystroke capture session overall was designed to capture the participants' behaviour patterns during various types of tasks, thereby allowing us to study the pattern of their typing under different circumstances. The five tasks share the same underlying principle but are also individually different, allowing the greatest possibility of investigating a varied range of potential mental states.

2.1.3 Handwriting capture specifications

The handwriting section of the experiment was designed under exactly the same principles as the keystroke section. In this case, there are four tasks, each containing different specific materials than for the keystroke section, therefore avoiding the risk of bias in the natural reflection on the participants' emotional status. However, the materials were again chosen to promote the same range of messages from happy to not happy, relaxed to not relaxed. Again, we were fully aware that the responses from the participants are entirely individual, and the aim of the experiment was to record their handwriting patterns and actual emotional state in each case, rather than making any assumptions based on the task itself.

- Task 1: Participants were asked to copy a list of predefined words. The words were chosen to encompass the execution of all the most common character-to-character-transitions in English, providing a rich data generation environment for subsequent analysis.

"*The communication method: Subroutine call or method invocation will not exit until the next invoked computation has been terminated. Asynchronous message passing, by contrast, can result in a response arriving a significant time after the request message has been sent through the net.*"

This task was designed closely to parallel the typing task (Task 1) which also followed the principle of providing the participants some text to write and helping them to become comfortable when writing in our experimental setting. The setting will be explained in Section 2.2.2.2.

- Task 2: Participants were presented with a picture (Figure 2.4) and asked to write a description of it in their own words, in exactly the same way as for the typing task described above (Task 3). The picture was also chosen to convey a positive and "happy" message.

FIGURE 2.4: Handwriting experiment Task 2 - Picture to be described

- Task 3: This parallels Task 4 of the keystroke capture, as described above. For this task, participants were again presented with a picture (Figure 2.5) and asked to write a description of it in their own words. This task was designed to repeat the previous task but with a different picture which was this time chosen to convey a less happy message (as a parallel to the corresponding typing task. The reasoning was the same as for the typing task (Task 4) previously described.

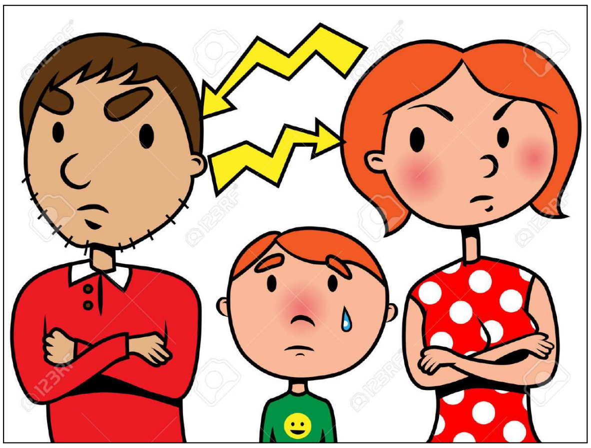

FIGURE 2.5: Handwriting experiment Task 3 - Picture

- Task 4: For the task, the participants were asked to copy a specified list of words (10 words, 50 characters). They had to try to complete the task within a time span of a maximum of 10 seconds. A countdown timer was displayed along with the text, paralleling Task 5 of the keystroke capture, and with the same motivation.

The handwriting capture session was arranged to take place very close to the keystroke session, in order to harmonize with the same underlying principle, potentially maximize the possibility of placing the participant in the most intuitive environment, while at the same time using completely different content to minimize the bias that potentially might otherwise exist in the experiment.

It is clear from the consideration of both part of this data capture exercise that emotional state is a changing and multidimensional phenomenon. This experiment overall was designed to create a window through which we might be able to explore the potential relationship between participants' mental state and their behaviour during writing and typing.

2.2 Experiment structure

2.2.1 Capture software

The experiment utilized two pieces of capture software, one for each modality that the study focused on. Both pieces of capture software provide the functionality to store data capture from each session separately under a unique subject ID. In order to ensure that that data capture was carry out anonymously.

2.2.1.1 Keystroke capture software

As was the case when considering the database, there is no publicly available capture software, which is universally used for this type of experiment. Therefore, the implementation of capture software that specifically fits the designed protocol was necessary. A custom piece of software for the keystroke capture was implemented using Java. The software captures each key press event and release event along with their respective timestamps in nanosecond precision (using the most accurate clock available on machine which can be guaranteed to have at least 10ms accuracy). Key event entries were added to linked list to allow constant-time insertions. These precautions were taken to minimized the effect of clock resolution on keystroke dynamics, as investigated by Killourhy and Maxion [184].

The software delivered the keystroke tasks to the participants exactly as instructed by the data capture protocol. Each session was continuous and uninterrupted. Figures 2.6 - 2.15 demonstrate the GUI of the software for each task:

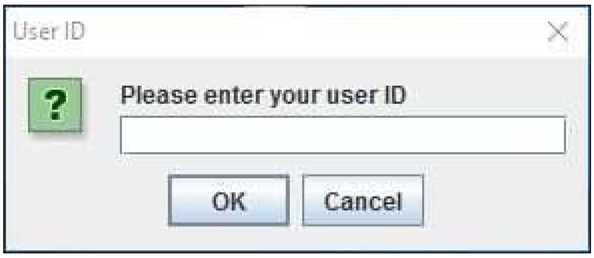

FIGURE 2.6: Participants were given a unique ID number to enter before they start their session

FIGURE 2.7: The screen before the first task begins. This is where the detailed instructions are given to the participants verbally, thereby minimizing interruptions during the task execution itself.

FIGURE 2.8: Task 1 - section 1 The correctly entered word is highlighted in green

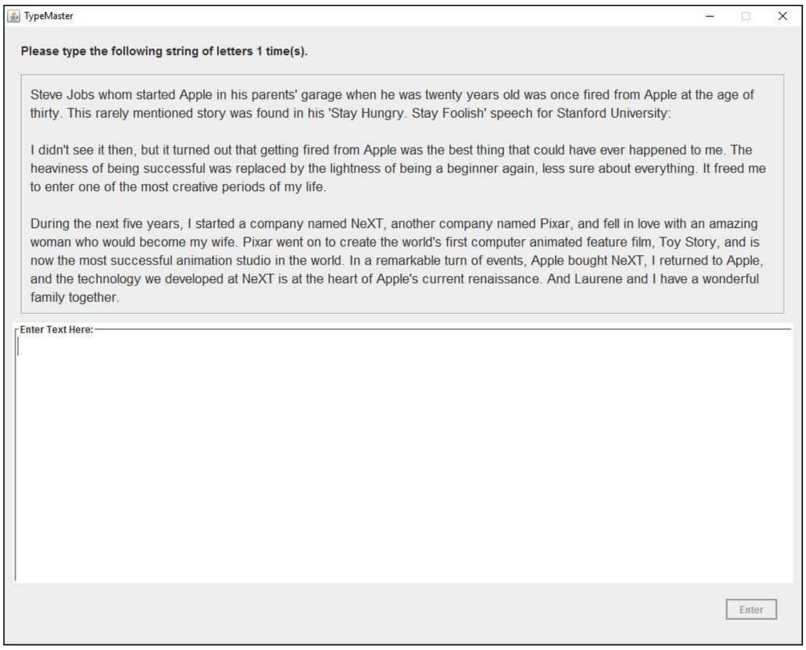

FIGURE 2.9: Task 1 - section 2 The participants are presented with paragraphs of text to copy

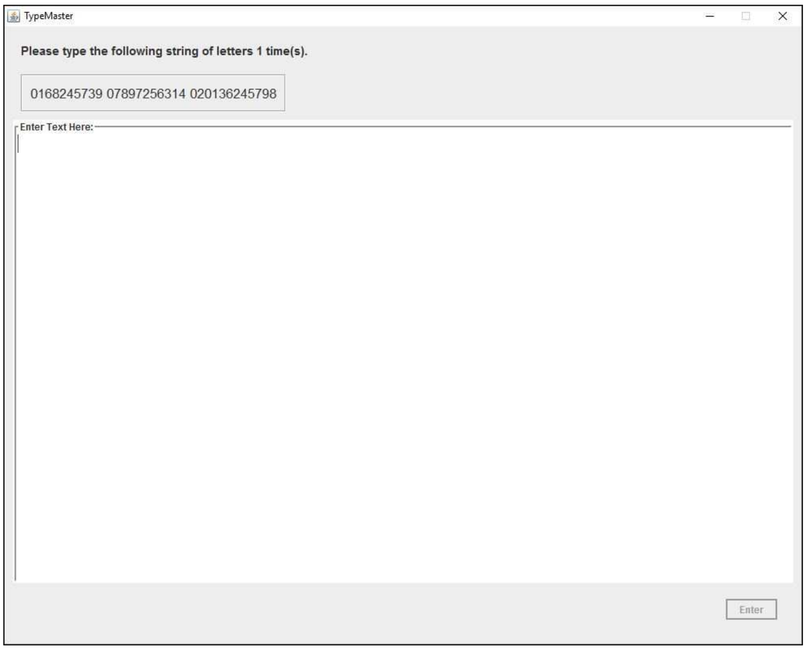

FIGURE 2.10: Task 1 - section 3 The number copying task - numbers was presented of top of the screen for subjects to copy

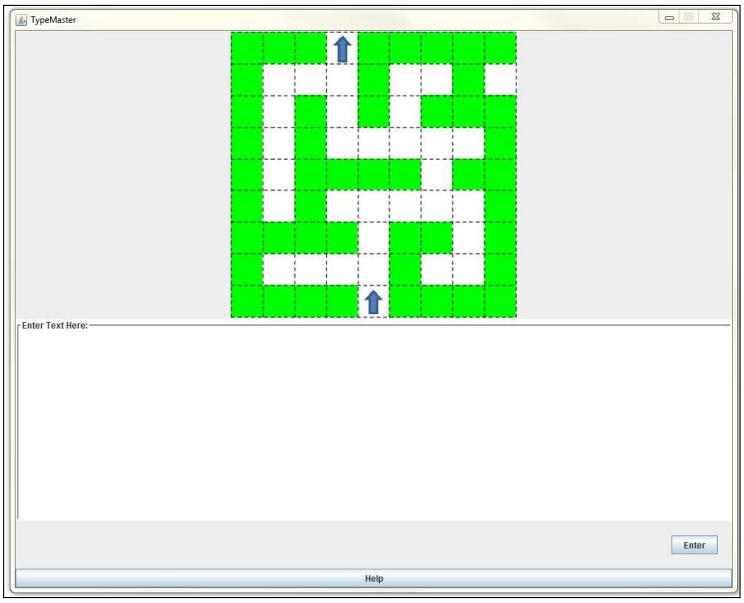

FIGURE 2.11: Task 2 - The image presented appears immediately above where the text is entered, so that participants can easily refer back to it

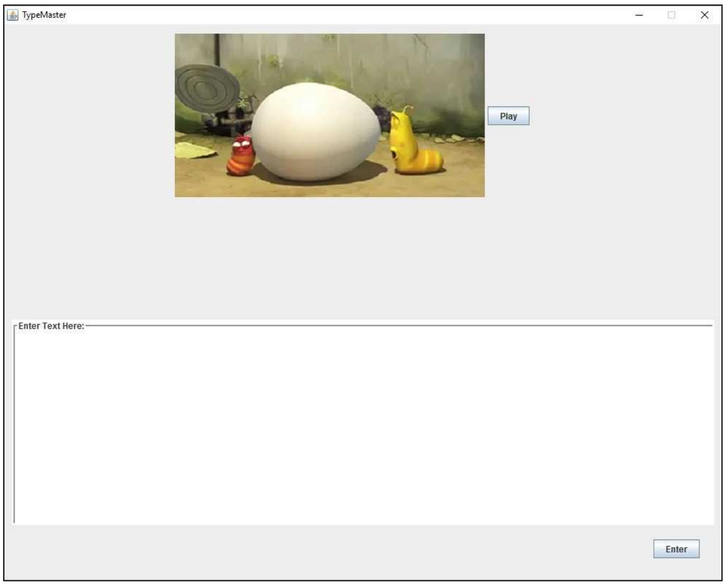

FIGURE 2.12: Task 3 - The participants can click the Play button to watch the clip from the beginning and type any word in the Enter Text Here text box

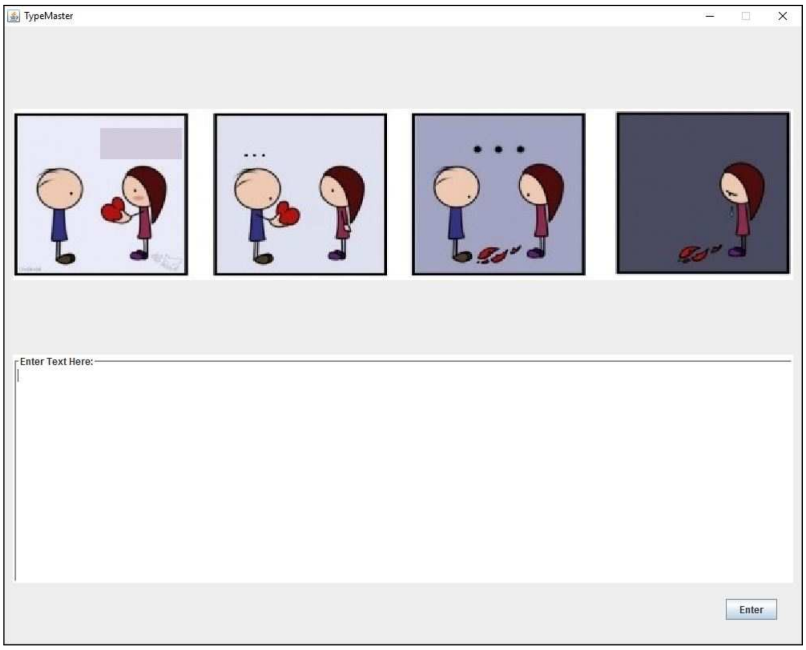

FIGURE 2.13: Task 4 - The series of pictures appear above the text input the area

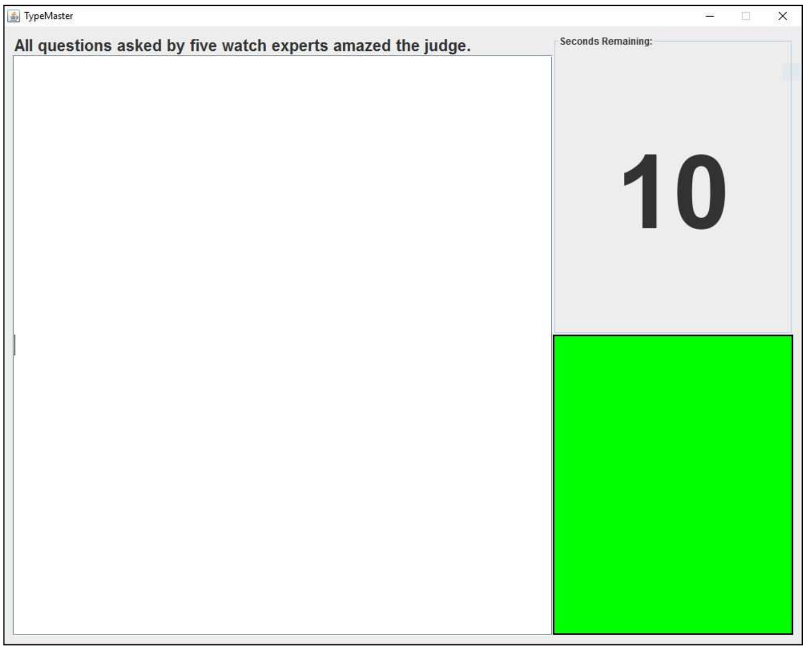

FIGURE 2.14: Task 5 - The timed task, where the timer starts counting down as soon as the first letter is entered

FIGURE 2.15: Task 5 - As the time runs out the colour of the light changes to red

2.2.1.2 Handwriting software

The handwriting capture software (MEDDRAW Data Capture tool) was originally developed within the Image Processing Research Group at the University of Kent, but has been progressively modified by successive research students. It is now a standard resource within the research group. The MEDDRAW Data Capture Program records the pen moment data from a digitising tablet. Data stored include a range of parameters: position, pressure, tilt and button status, which are stored in data pockets. Each data pocket is timestamped to microsecond accuracy. A detailed description of the data collected will be covered in Section 2.2.2.2.

As shown in Figure 2.16, using this software package, all the writing samples were captured using a digitising tablet with a paper overlay to provide familiar feedback during the writing process. Writing on a digitising tablet can be an unfamiliar activity to some people and they might not behave as they would normally when writing on paper.

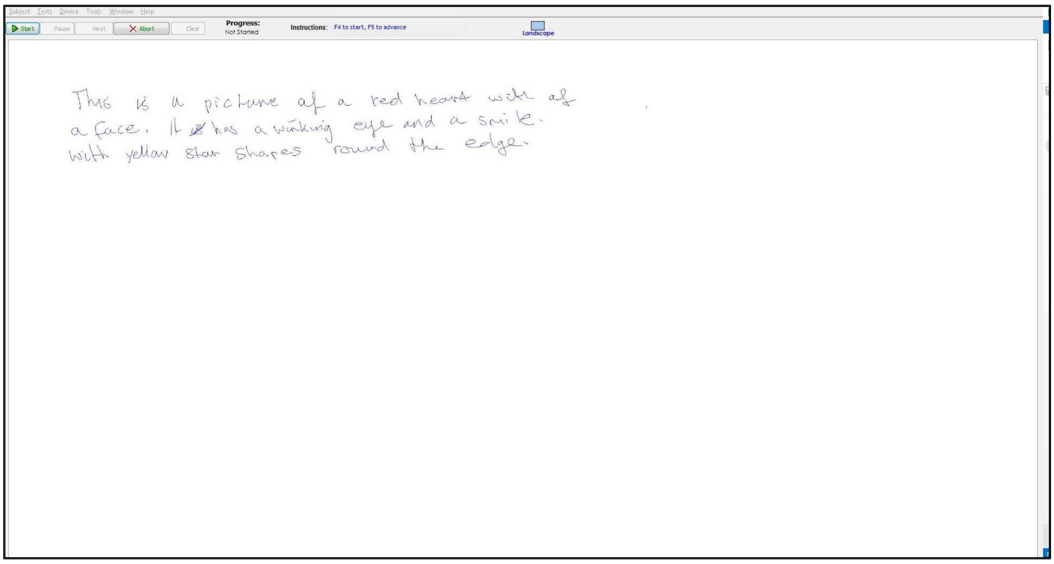

FIGURE 2.16: The software records data a task at a time. The captured handwriting data were played back simultaneously on the screen during the participant enrol their data.

2.2.2 Capture hardware set-up

Each handwriting session was performed on the tablet. Task materials were shown to the participants as hard copy pictures. The experimental set up will be discussed in detail in Section 2.2.2.2.

2.2.2.1 Keystroke experiment hardware set-up

The hardware set-up for the keystroke capture software is illustrated with reference to the Figures 2.17 - 2.18.

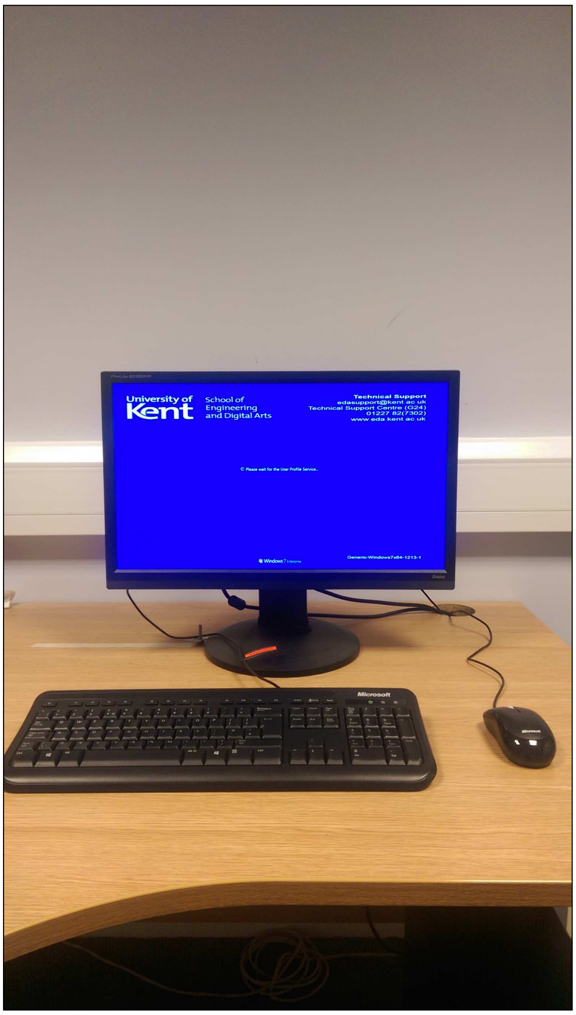

FIGURE 2.17: Hardware set-up for Keystroke capture including a standard monitor, and keyboard (QWERTY - Microsoft Wired Keyboard 400v1.0)

FIGURE 2.18: A participant during a keystroke capture session

2.2.2.2 Handwriting experiment hardware set-up

The set-up includes a "WACOM Intuos 3 PTZ-630 Graphic Tablet" and associated "inking WACOM pen". This is connected with the PC and transmits the captured data directly to the software described in Section 2.2.1.2. Figures 2.19 - 2.21 demonstrate the hardware set-up and a participant performing different tasks.

FIGURE 2.19: The WACOM tablet and the pen

FIGURE 2.20: A participant writing with the tablet pen on a paper overlay during the first handwriting task

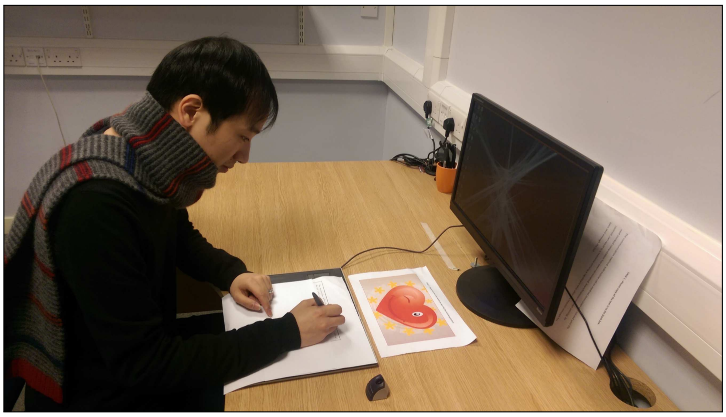

Figure 2.21: A participant writing in response to the presentation of the picture for the second handwriting task

After the data capture, few unforeseeable issues were encountered, which needed to be addressed before assembling the entries for the database itself, as follows:

- The participants were not given any limit on the length of text they provide for the free typing tasks (Task 2, 3, 4). It was necessary to make sure that subjects were not under any pressure for how many words they had to supply, which would not be their most intuitive behaviour. However, four out of a hundred participants supplied very short sentences with less than 10 words for the free typing task and this then became slightly problematic during the data processing phase, as those participants did not supply enough typing pattern samples for the minimum data processing requirements. Therefore, the data provided by the four participants were not comparable to their data from other tasks and data from other participants.

  This was compensated by removing the four insufficient supplied data participants during the data analysis for emotion prediction. Those participants were only a very small percentage (4%) of the whole database. This did not significantly affect our data processing process. Having discovered this issue early on in the experiment, the verbal script that the capture supervisor used was updated to include verbal encouragement for subjects to provide as much detail as possible in their responses.

- Participant filter: There was no filtering of subjects to exclude any willing participants before the capture session. As a result, there were participants who potentially could introduce "noisy" data into the database due to different circumstances. For example, there was a participant who had dyslexia and the data he/she provided had to be excluded, because the data collected from this subject was a complete outlier and distorted the database by introducing noise. A lesson can be learned from this, so that an appropriate participant filter can be put in place for any further data capture to avoid these potential problems.

- There is also a potential bias where a small percentage of participants might complete the emotional score questionnaire based on their individual perception of the experimenter's expectation. This potential problem was minimised by clearly explaining to the participants that they should score their emotional status solely based on how they feel at the time.

The points mentioned should be taken into account into any future data capture practice within relevant field of study.

2.3 Database specification

The database contains both keystroke and handwriting data samples from 100 participants, the largest number of participants included in any database available in this area of research (see the discussion in Section 1.3). This section will describe in more detail an initial analysis of the database in terms of the distribution of the data characteristics from the point of view of both demographic and emotional metadata for all the participants.

2.3.1 Analysis of subject characteristics within the database

- Age distribution

  Figure 2.22 illustrates the age distribution of the 100 participants included in the experiment. For our purposes (since age is a continuous process, and it is recognized that age-based effects are incremental) we have assigned subjects to one of four age bands corresponding to “Under 25 years old”, “25 - 40 years old”, “40 - 60 years old” and “above 60 years old”. Since age is a continuous variable, and since the granularity level at which to work is to some extent arbitrary, it is necessary to use this type of banding. The four bands used here were chosen based on common practice in many other biometrics studies. A useful analysis of this issue can be found in [111].

The histogram demonstrates that the recruitment was mainly distributed between the "under 25 years old" band, and the "25 to 40 years old" band. This reflects the fact that the recruitment of volunteers was mainly conducted within the University of Kent, with a consequence predominance of students, we did manage to avoid an over-reliance on the enrolment of a single type of participant (i.e. undergraduate students). This helps to minimize any potential bias overall, which age and the general background of the participants might have on the analysis of our subsequent experiments.

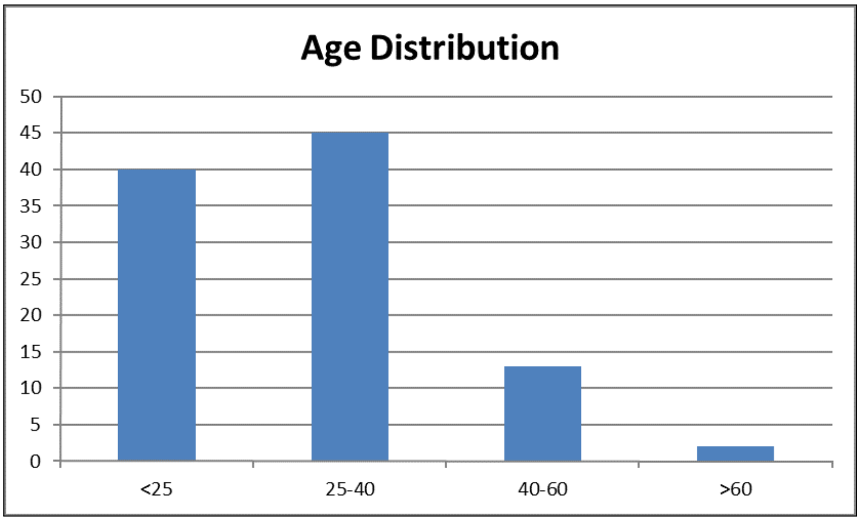

Figure 2.22: Age distribution histogram

- Gender distribution

  Figure 2.23 illustrates the gender distribution in the experimental population. It is clear that the database is approximately balanced in relation to gender. Specifically, the population consists of 45 female and 55 male participants. This demonstrates that the recruitment was broadly balanced in this regard and therefore, minimizes the potential bias gender may have on the subsequent experimental analysis.

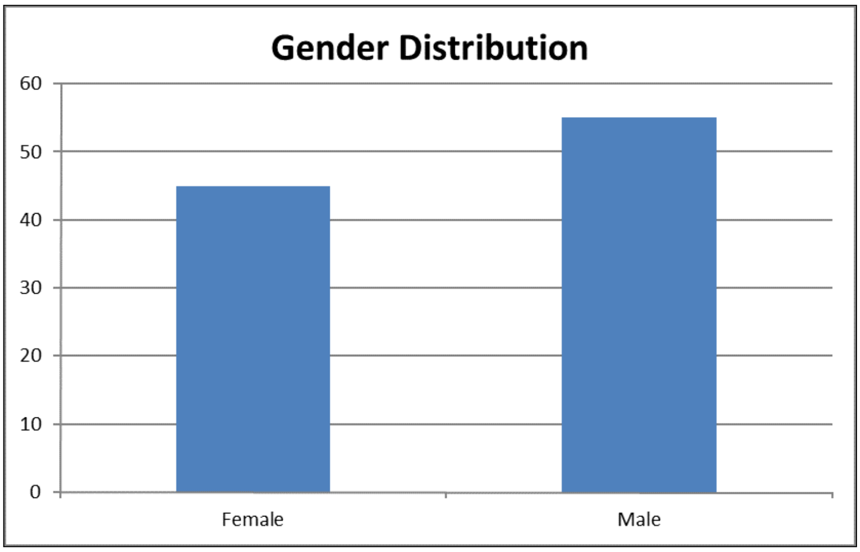

FIGURE 2.23: Gender distribution histogram

- Handedness distribution

  Figure 2.24 illustrates the distribution of handedness (whether a subject is left- or right-handed) in the experimental population. It can be seen that a significant majority of subjects are right-handed, as is common within a typical population, but this is not a factor of particular significance in this study, and we were careful not to attempt any analysis based on handedness as a variable.

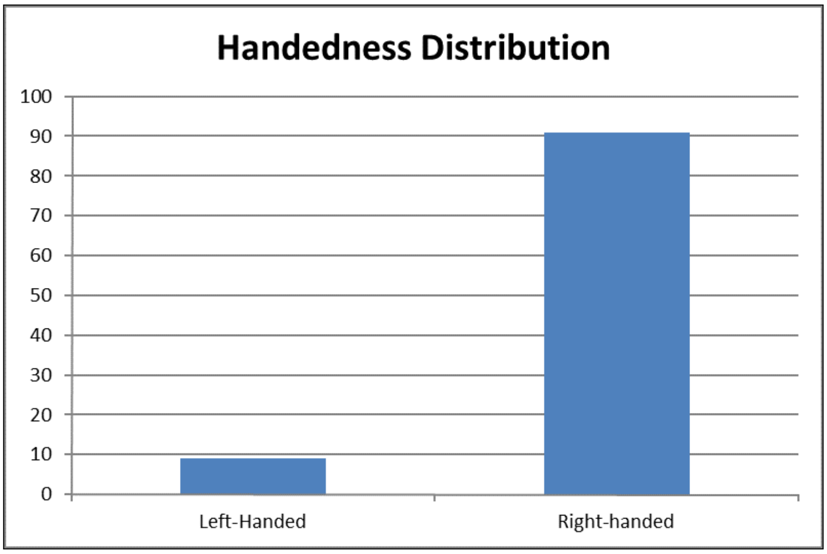

FIGURE 2.24: Handedness distribution histogram

2.3.2 Raw data and data structure

This section will describe the raw data capture and the stored format of the captured data within the database.

2.3.2.1 Keystroke data

Keystroke data, which include the Key Presses events and Key Release events were captured with the key code (ASCII code - the numerical representation of each key on the keyboard) and time-stamp. The data were stored separately into individual folders for each participant, with each folder named according to the unique user ID. Each folder contains .key format files that store the raw data for each task individually.

2.3.2.2 Handwriting data

Handwriting data were captured with 14 types of raw data element :

1. Timestamp - The time offset in microseconds since the start of the capture process.
2. X coordinate - Horizontal location of the pen on the tablet surface.
3. Y coordinate - Vertical location of the pen on the tablet surface.
4. Normal Pressure - Normal pen tip pressure, where the pen is directly pressing down.
5. Tangential Pressure - Pen tip pressure when the pen is pressing down on the tablet with an angle.
6. Status - the cursor state (in/out of input surface of the tablet)
7. Cursor - the cursor type that generated the data packet, the cursor type which will indicate the packet data items the cursor support.
8. Context - the ID of the context that generated the packet, which is referring to the window that is active for the input text to appear.

   Three data items relating to the orientation of the pen.

9. Azimuth - the clockwise rotation of the cursor about the z axis through a full circular range.

10. Altitude - the angle between the cursor and the x-y surface. Positive values specify an angle upward toward the positive z axis, which means the back of the pen is tilting up. Negative values specify an angle downward toward the negative z axis, which means the back of the pen is tilting down.

11. Twist - the clockwise rotation of the cursor about its own major axis.

    Three data items relating to the rotation of the pen.

12. Pitch - the pitch represents the slope of the cursor.

13. Roll - the roll represents the degree of rotation of the cursor

14. Yaw - the yaw represents the yaw angle of the cursor.

The data were stored separately into individual folders with each folder named by the unique user ID. Each folder also contains, for each task, a screen shot of the handwriting image in *.jpg* format and all the 15 data elements noted above in a formatted text *.tst*.

2.4 Conclusions

This chapter has covered, in detail and systematically, the data capture procedure, which is fundamental to the experimental results reported later and, indeed, without which, the experiments could not have been carried out. The description has provided information from the design of the protocol to the implementation of the protocol in the experimental procedures to be discussed later. A main focus of the chapter is to explain the logic behind each decision that was made on design and implementation. This is a very important fundamental building block for the further analysis, which will be presented in later chapters, and represented a significant effort in its realisation, and a major contribution of the study.

The other principal focus of this chapter is to introduce the software and hardware set-up for the experimental investigation. The set-up shows how we delivered our protocol into a practical data capture procedure. The set-up also incorporated the designed protocol as the guideline, thus creating the intuitive environment required by our specification. This also complements the goal of making more accurate the eventual emotion predictions. The database in its final form contains data acquired from 100 participants. This chapter has also analysed the distribution of the overall demographic data that were acquired for all participants during the data capture phase. These indicators might eventually play an important part when we draw conclusions from the analysis reported in later chapters.

Overall, through the data capture protocol, a very large amount of data material was collected, and this defined the starting point for the detailed exploration of the predictive capability of biometric data, which is at the heart of the study reported here. However, the raw data required further processing to facilitate our investigation, and a structured set of features needed to be developed for the data processing phase. Therefore, the next chapter will explain the project infrastructure from that perspective, and will discuss the technical implementation of this feature extraction process.

CHAPTER 3

Prediction of gender from handwriting data

Handwriting biometrics have been utilized mostly to determine the writer 's identity, as well as some other personal characteristics, for example, age and gender. However, a principal focus of this study is to explore the predictive capability of handwriting biometrics for characteristics which have been much less extensively explored in the literature. This type of prediction refers to what we have characterised as "higher-level mental states". Our study will focus on investigating the predictive capability of handwriting for determining the emotion/mental state experienced by the writer at the time of executing the writing in question. However, before getting explicitly into this area of emotion prediction with handwriting biometrics, a series of experiments on the prediction of a more frequently considered and more conventional characteristic (the gender of the writer) were carried out. This is partly because such an initial study can provide a rough benchmark in soft-biometrics prediction experiments for our data (other studies on gender prediction can be found, although these are not extensive, so this provides a basis for comparison while also adding to the literature of gender prediction) but also because this provides an excellent opportunity to reflect on the "quality" of the data captured.

This chapter will therefore introduce the investigations on gender prediction from digitised handwriting data and the results that were generated from this experimentation. This study also focuses on observing the impacts of data processing from two main different perspectives, which relate to the different feature types which can be utilised and the different handwriting content types which might be encountered, allowing an initial consideration of the impact these factors have on the gender prediction results. This initial study will therefore, provide some insight and evidence for achieving a better understanding of the practicality of utilising digitized handwriting data for gender prediction.

Section 3.1 will briefly review the relevant studies that use handwriting biometrics for soft-biometrics information prediction. In Section 3.2, a brief review of gender prediction and will also present the experimental set-up and gender prediction results by analysing the newly collected handwriting data (reported in Chapter 2). Section 3.3 will present effects of different features set have on gender prediction. Section 3.4 will make some concluding observations.

3.1 Handwriting biometrics

There have been many studies exploring handwriting biometrics in order to improve the identification of individuals ([37, 185] are representative examples) or to narrow down a list of possible identities while [78], at the same time, there have been an increasing number of research studies reported for soft-biometrics information predictions from the commonly used handwriting measurements

Although handwriting is nowadays less regarded as the dominant method for our personal communication, it is nevertheless still widely used in many important applications, some established and some new, and is likely to continue to be one of the principal modalities, especially given the increase in hand-held information platforms. The newly emerging most significant application is the use of handwriting data to determine the identity or other personal characteristics (such as age [108, 127] and gender [125, 127]) of the writer in biometrics applications. These studies can help in providing important information about the sample provider without have to acquiring any specific additional details of the person. There are also various other specific traits that can be valuable, for example, to determine the "handedness" [75, 76, 127] of a writer can be a very valuable in certain scenarios, such as forensic investigations [78], in order to identify, eliminate or narrow down a list of possible suspects in criminal investigations.

Other potentially important applications include providing assessment or diagnostic information in healthcare scenarios (such as in the management of Parkinson's disease in the assessment of patients' post-stroke [162]).

This chapter will therefore investigate and explore primarily one aspect of handwriting in this rather more general context, focusing particularly on the predictive properties of handwriting (especially, digital handwritten data) in relation to identifying the gender of the writer.

3.2 Gender prediction with soft-biometrics

As shown in Chapter 1, there has been some indication of public interest in the possibility of recognising soft characteristics such as gender from conventional biometrics data, for example, voice [73] , face [66] , keystroke [17] and also handwriting [125, 127] as introduced in Section 3.1. The ability to determine the gender of a subject from a facial image, voice, keystroke or handwriting has obvious practical importance, which suggests powerful and valuable application possibilities. For example, in [17], the author introduces an approach for assisting the user to determine gender of the person they are communicating with in a social network environment from the keystroke data others provide, facilitating gender and age prediction in crime investigation by analysing conversation recordings in [73]. The authors [66] presented algorithms that process facial video data which then can be used in human-computer interaction, surveillance monitoring, video content analysis, targeted advertising, biometrics, and entertainment.

The main handwritten data capture method reported in the literature has been to work with scanned images [125, 186] of the handwritten sample or, sometimes, to use samples acquired from a conventional electronic white board [187]. As demonstrated in Section 3.1, gender prediction from data collected directly from a handwriting capture tablet has not been as extensively investigated. Therefore, a comparison with handwritten data captured from other sources can help to improve an understanding of data from such a source, and in a sense benchmark the predictive performances that can be generated from data captured from those data. Thus, a principal, goal of the work reported in this chapter is to explore the gender prediction capability of data acquired from a digitized handwriting capture tablet.

By carrying out some experimentation on conventional soft-biometrics prediction using the newly acquired data not only provides a good opportunity to test the predictive capability of the dataset acquired (see Chapter 2), but simultaneously provides a new processing platform, allowing us to address the following three issues which have not been addressed in previous studies for gender prediction from other sources (such as those as mentioned above, namely the whiteboard or scanned handwritten samples):

- Gender prediction from handwriting biometrics has mostly been studied using other data capture methods [125, 187]. Hence, this study can usefully fill a gap in our knowledge by generating some information about gender prediction using handwriting-based behavioural data captured by digitising tablet.

- Studying data from this particular type of data source provides the opportunity to evaluate the results in order to provide a framework for comparison of whether and how different feature types influence the experimental outcomes, to complement studies when other sources of data have been analysed [187].

- Another key experiment is to explore the effects on predictive performance when using various different types of handwriting content. In particular, from a practical point of view, two different types of content are of special interest. The first concerns the use of *fixed content* where the subjects are all given a predefined piece of text to write (and thus all the test population write the same text), and *free content* where the subjects were prompted to produce text on a particular topic, but the provided content can vary for each participating subject (in other words, there are no constraints imposed on exactly what each subject writes). Different (subject-specific) content was adopted in the study reported in [125], but an analysis of differences is not undertaken. The proposed study will therefore provide an opportunity for a new approach which can investigate the influence of these different data types and activity scenarios on predictive performance.

In this section, a set of features, which are commonly reported in the literature, such as those specified in [78, 109, 188–191] are introduced, and the first issue noted in the list above will be addressed. The other two issues will subsequently be addressed in Section 3.3, and discussed in conjunction with similar issues concerning emotion prediction from handwriting biometrics (which will be investigated further later).

3.2.1 Experimental set-up for gender prediction

As shown in Table 3.1, there are 26 static features and 24 dynamic features extracted from the data captured for this study. The term “static feature” refers to features that are generated from the visual appearance (i.e. directly from an image) of the handwriting segments at the end of each task, for example, “F44”. “Dynamic features” are features that are captured during the writing process by the sensors from the digitized pen and the tablet, and which reflect characteristics of the execution of the signing process itself, for example, “F3”.

Feature name	Feature type	Feature description
F1	Dynamic	Total distance the pen travelled
F2	Dynamic	The total time taken for the writing
F3	Dynamic	The number of times pen moved away from the tablet
F4	Dynamic	Average pen velocity in x axis
F5	Dynamic	Average pen velocity in y axis
F6	Dynamic	Number of zero velocity sample points in x
F7	Dynamic	Number of zero velocity sample points in y
F8	Dynamic	Maximum pen velocity in x - Average pen velocity in x
F9	Dynamic	Maximum pen velocity in x - Minimum pen velocity in x
F10	Dynamic	Maximum pen velocity in y - Average pen velocity in y
F11	Dynamic	Maximum pen velocity in y - Minimum pen velocity in y
F12	Dynamic	Maximum pen velocity in x - Minimum pen velocity in y
F13	Dynamic	Average pen acceleration in x axis
F14	Dynamic	Average pen acceleration in y axis
F15	Dynamic	Number of zero acceleration sample points in x
F16	Dynamic	Number of zero acceleration sample points in y
F17	Dynamic	Maximum pen acceleration in x - Average pen acceleration in x
F18	Dynamic	Maximum pen acceleration in x - Minimum pen acceleration in x
F19	Dynamic	Maximum pen acceleration in y - Average pen acceleration in y
F20	Dynamic	Maximum pen acceleration in y - Minimum pen acceleration in y
F21	Dynamic	Maximum pen acceleration in x - Minimum pen acceleration in y
F22	Dynamic	Azimuth
F23	Dynamic	Altitude
F24	Dynamic	Pressure
F25	Static	Number of points comprising the image
F26	Static	Sum of x coordinate values
F27	Static	Standard deviation of x coordinates values
F28	Static	Maximum x coordinate value - Final x coordinate value
F29	Static	Initial x coordinate value - Minimum x coordinate value
F30	Static	Final x coordinate value - Minimum x coordinate value
F31	Static	Average of x coordinate values
F32	Static	Maximum x coordinate value
F33	Static	F31 - Minimum x coordinate value
F34	Static	Sum of y coordinate values
F35	Static	Standard deviation of y coordinate values
F36	Static	Maximum y coordinate value - Final y coordinate value
F37	Static	Initial y coordinate value - Minimum y coordinate value
F38	Static	Final y coordinate value - Minimum y coordinate value
F39	Static	Average of y coordinate values
F40	Static	Maximum y coordinate value
F41	Static	F39 - Minimum y coordinate value
F42	Static	Vertical centralness
F43	Static	Horizontal centralness
F44	Static	The width of the handwriting
F45	Static	The height of the handwriting
F46	Static	The width of the handwriting / The height of the handwriting
F47	Static	Area of the handwriting
F48	Static	The width of the handwriting / Area of the handwriting
F49	Static	The height of the handwriting / Area of the handwriting
F50	Static	The number of times pen passes through the mid-line

TABLE 3.1: Handwriting features

The features were normalised by a commonly adopted normalisation method for handwriting features which is mean and variance normalisation (also known as Z score, as shown in Equation 3.1). Each feature is normalised by their raw value "x" subtracting the mean of the feature "μ" and divided by the standard deviation (square root of variance) of the feature "σ". This method allows the feature values to represent how deviated they are from their mean, which enable features to better represent and distinguish between different types of behaviour. However, the disadvantages are the loss of behavioural information on absolute levels. This method was used because in our study, the ability to distinguish between different behavioural patterns has higher priority than maintaining the level of detail of the raw writing movement. The extraction software was already available within the Research Group[1]. Three classifiers were used for the gender prediction process, to generate data on the effects of different classification approaches. KNN (K=1) and SVM, and NaiveBayes classifier were used and a hold-out validation methodology adopted, where the first quarter of the subject population was used in testing and the rest as the training set, this method ensures that the same subjects are not used for training and testing to imitate a more likely to occur real life scenario where the samples that need to be analysed not existing in the known database.

$$z = \frac{x - \mu}{\sigma} \tag{3.1}$$

3.2.2 Gender prediction results

As can be observed in Figure 3.1, the gender prediction accuracy figures achievable (while using all these 50 features that were extracted from the data collected from the digital graphics tablet acquisition platform), are within the range of 60% - 75%, depending on the classification infrastructure utilised, where the best result was at 75%, achieved by the SVM classifier. This is comparable to and, indeed something of an improvement in comparison to other published results (53% - 62% also using SVM classifier) in studies reported in the literature on gender prediction using data captured with a conventional white board where

[1]Developed and provided by Dr Meryem Erbilek and thus available as part of an overall data processing tool for this study.

both dynamic and static features(static representation of the dynamic features) were included (total 29 features) [77].

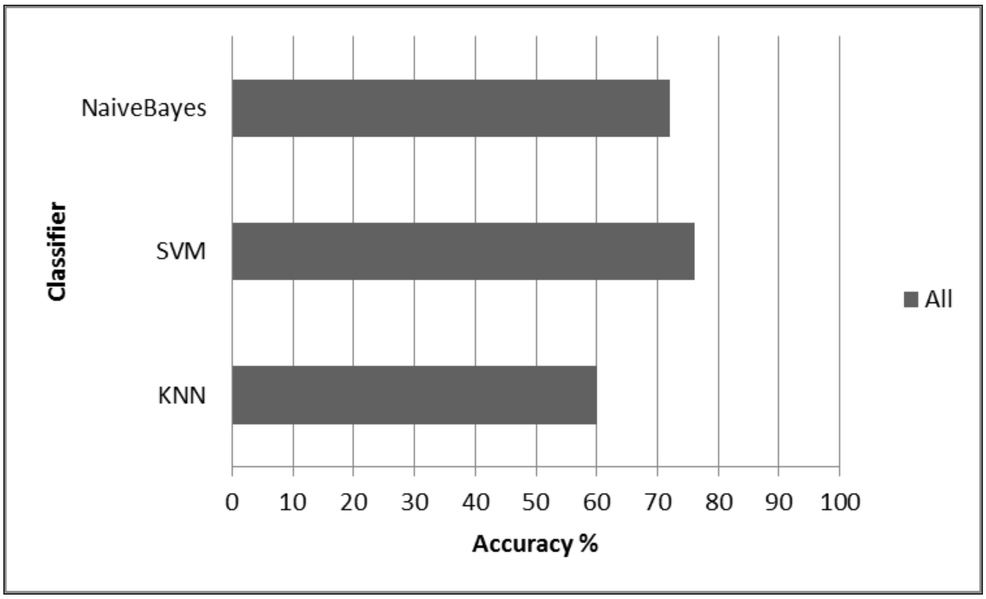

FIGURE 3.1: Predictive accuracy for fixed task, where the entire collection of 50 features presented in Table 3.1 were included.

This study, then, utilises a new data source (handwriting data captured dynamically from a digitising tablet), for gender prediction. The results demonstrate that the data captured share a similar level of predictive capability to what has been reported in other studies, but somewhat better in comparison with the main gender prediction studies previously reported in [77, 187]. These encouraging results present an opportunity for more in-depth gender prediction analysis using handwriting samples captured from a digitising tablet.

3.3 Effects of different feature sets in gender prediction

The process of gender prediction can be performed with different feature sets. In this experiment, 24 dynamic features and 26 static features are included in the overall feature set, as explained in Section 3.2.1. In order to investigate the effects of different feature sets in gender prediction, the data from Task 1 in our data acquisition procedure are used (Task 1 is where the participants were given a set of predefined text to copy.), as these provide greater richness in the handwriting data collected in comparison to other tasks.

For this experiment, the feature sets are divided into separate dynamic and static feature sets, as before. Figure 3.2 demonstrates the predictive performance when the three different classifiers (NaiveBayes, SVM and KNN) are adopted, and for three different feature sets, which consist of the dynamic features alone, the static features alone, and the full set of features combining both the static and the dynamic feature sets (the full set of features is also used in Section 3.2).

The predictive performances returned are within the range of 60% to 80% accuracy. As has already been noted in [13], the study demonstrates that "Dynamic features" generally return a better performance than "Static features", with the predictive performance around 64.25% for the dynamic features set and 55.39% for the static features set, while the best results are most commonly generated by combining the two types of features, in this case 67.57%. This phenomenon can also be observed in our study when SVM classifier was used, where about 68% was achieved by the static feature set, 72% by the dynamic feature set and about 75% for the combined feature set. Therefore, any firm conclusions that are drawn about the relationship between feature type and gender predictive performance should take the classifier type into account.

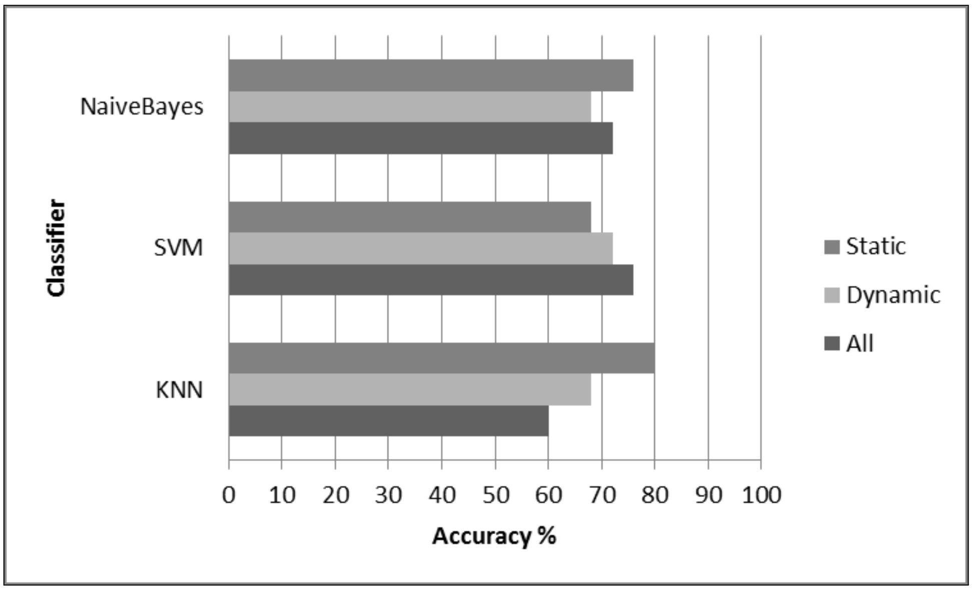

Figure 3.2: Predictive accuracy for Task 1 using different feature types.

By applying the same settings, where features were divided into static and dynamic categories, the same experiments were repeated with the data from Task 2 of our new data collection exercise (Task 2 is where the participants were asked to write a paragraph based on the picture they were presented, and the context is not constrained). As can be seen in Figure 3.3, the predictive performance obtained by using the NaiveBayes classifier confirms the observation commonly reported in the literature, that "Dynamic features" generally perform better than "Static features", while the best results are generated by combining the two types of features. For the other two classifiers, SVM and KNN classifier, the pattern reported in [187], where the combined feature set performs better than the static or dynamic feature sets alone , and the dynamic feature set generally performs better than the static feature set alone, has changed. The combined feature set still generates the best result achievable. However, when different classifiers and tasks are used, and it should be noted that no straightforward conclusions can easily be drawn with regard to the relationship between different feature sets, task type and gender prediction performance. By observing 3.2 and 3.3 together, a comparison of predictive performance in relation to different task types can be made. Task 1 is a constrained task and Task 2 is an unconstrained task. For Task 2, when using the same settings, the range of predictive accuracy performances attainable are within the range of 48% to 60%. Task 1, the performances are ranging from 60% - 80%, As the results show, there is a potential impact, even for the same subjects handwriting, of the type of task adopted, and the results can be influenced by using data drawn from different types of tasks.

It is apparent, therefore, that it can be a very challenging task to draw any definitive conclusions about the precise effects of feature type or task type on predictive performance for gender prediction. However, the effects that have been observed provide some important initial indicators of the factors, which need to be taken into consideration when undertaking gender prediction based on handwriting biometrics. The encouraging result above does reflect the potential for gender prediction with data captured from a digitized handwriting capture tablet.

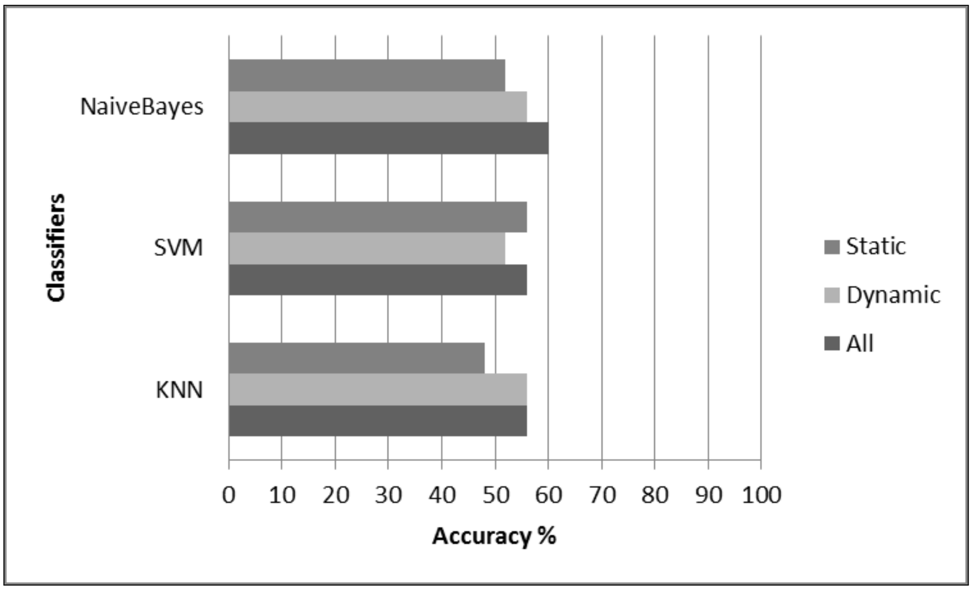

FIGURE 3.3: Predictive accuracy for Task 2 using different feature types.

3.4 Conclusion

An interesting yet conventional soft-biometrics characteristic prediction task (gender prediction) has been investigated in this chapter. This area has some significant potential value in a variety of applications, for example, in the context of forensic investigation. In the literature, most of the reported studies were based on data capture from more conventional capture methods. This study aimed to fill the gap of using handwriting data acquired by a digitising tablet for gender prediction. The experimental results produced show that a 75% predictive accuracy can be achieved with an SVM classifier, which is an improvement in comparison to the 62% predictive accuracy results published in [77], which was also using SVM classifier as mentioned in Section 3.2.2.

This chapter has also addressed two important issues related to the use of handwriting for gender prediction. The first concerns the question “Do different types of feature sets influence the gender prediction result?” and the second “Does the content of the writing influence the gender prediction result?” To answer the first question, the feature sets were divided into - static feature set, dynamic feature set and the combination of both feature sets. The predictive accuracy when using these three set are still around 68% - 75% with the SVM classifier when Task 1 data were tested. When the content of the handwriting is different, specifically where data from an “unconstrained” task were used, the predictive accuracy dropped to around 50% - 55%.

It appears that feature set type has, in gender prediction, smaller effect than the content of the task when the SVM classifier is utilised. At the same time, different classifiers also appear to have a different impact on the performance of gender prediction.

An important message is emerging from this type of study, where comparison between performances for gender prediction should not be made without considering the points above. This has important implications when optimizing the gender prediction process. The results presented in this chapter will raise the awareness of the issues that are particularly important in this type of research.

It appears that the dataset, which was specifically captured for our study, proves to be appropriate for the purpose of gender prediction, and this gives some confidence that it will be an appropriate dataset too for the other experimentation to be reported. Specifically, the main focus of this thesis is to investigate the predictive capability of handwriting biometrics and keystroke dynamics in the prediction of higher-level states, a topic which will be explored and discussed in the next two chapters.

CHAPTER 4

Prediction of higher-level states from handwriting data

Handwriting biometrics research has a long history. The most reported studies are based around utilising available handwriting information, which is assumed to be individually unique, to identify the data "provider" [192]. However, soft-biometrics are generally known not to be unique, yet still reflect an individual's characteristics. Age and gender prediction studies, reported in Section 1.1, are good illustrative examples. The main focus of this thesis is to expand on previous research studies on soft-biometrics characteristics (such as age and gender) prediction from "Handwriting biometrics" and "Keystroke biometrics", specifically to explore the possibility of predicting an individual's "higher-level mental state", and in particular those states which could be characterised as "emotional states". This chapter will present the experiments that were performed on emotion prediction, and the newly acquired database (described in detail in chapter 2), which contains data captured from handwriting segments produced under varying task conditions by a substantial number of writers, was used for all the experiments. The predictive capability of biometric data in relation to such "emotional states" based on acquired keystroke data, will be presented and discussed in Chapter 5.

Emotional states can refer to a wide range of emotions, and those emotions are multi-dimensional [193]. In order to capture the handwriting behaviour of individuals under different emotions, a number of different emotions were investigated via the questionnaire, as mentioned in Section 2.1.1. In order to investigate the predictive capability of handwriting biometrics in this initial study, two of these emotions were selected, these being emotions which are generally considered to be relatively straightforward to recognise in individuals. These two states are the degree of "Happiness" and the degree of being "Relaxed" being experienced by an individual. In our reported studies, the terms "higher-level mental state", "emotional state" and "mental state" all refer (interchangeably) to the acknowledgement of whether a participant is feeling "Happy" or "Not happy", "Relaxed" or "Not relaxed", and the extent to which the subjects are experiencing such a feeling.

As discussed in Section 1.1, there are studies to be found which report emotion predictions mainly based on other biometric modalities, such as face [143, 145, 194], voice [195], gait [196], and keystroke [152]. It is apparent that there were not many extensive reported studies on mental state or emotion prediction based on handwriting biometrics. This chapter will report and analyse some new and less considered experimental studies on prediction from the handwriting data which had been acquired during the data capture phase (reported in detail in Chapter 2). This chapter also introduce the operation of feature extraction, ground truth labelling and mental state prediction experimentations.

This chapter presents the results of the all the emotion prediction experiments that uses handwriting data, also initiate a discussion of the implications of the reported results, which address some of the key issues arising in this study. The discussion will provide a more insightful understanding of the challenges and opportunities which lie ahead for mental state prediction using handwriting data.

Section 4.1 will briefly review the relevant studies that use handwriting biometrics for emotion prediction. In Section 4.2, the feature extraction process will be presented with a detailed list of handwriting feature used. Section 4.3 will present ground truth labelling method with the emotion score distribution charts. Section 4.4 will present the emotion prediction methods and results. Section 4.5 will further investigate three of the main areas (ground truth labelling method, nature of handwriting features and types of tasks) where different settings could potentially influence the predictive performance. Section 4.6 will make some concluding observations.

4.1 Handwriting biometrics

There have been many studies exploring handwriting biometrics in order to improve the identification of individuals [37, 185] or narrow down a list of possible identities while [78], at the same time, there have been an increasing number of research studies reported for soft-biometrics information predictions from the commonly used handwriting measurements.

The reported studies of soft-biometrics have mostly been focused on investigating characteristics such as participants' age [108, 127] or gender [125, 127]. These studies can help in providing important information about the sample provider without have to acquiring any specific additional details of the identity of that person. There are also various other specific traits that can be useful, for example, to determine the “handedness” [76, 108, 127] of a writer can be a very valuable in certain scenarios, such as forensic investigations [75], in order to identify, eliminate or narrow down a list of possible suspects in a criminal investigation,as mentioned in Section 3.1.

These reported studies have shown that the sort of characteristics mentioned above can be predicted with a sufficient degree of accuracy for many practical applications. However, there has been much less work to explore these ideas further, taking them beyond the prediction of the simple characteristics described. The possibility of extending the predictive capabilities to higher-level mental states, for example, by using handwriting behaviour or other biometric characteristics can be beneficial in many practical scenarios. For example, the application to forensic investigations [78], as mentioned above, would be beneficial, and this type of prediction would self-evidently also potentially be very valuable in some healthcare scenarios [79], and in many other applications. The higher-level mental states of interest can involve a broad range of emotional states, but “Happy” and “Relaxed” are the main emotional states which the present reported work has focused on. The prediction of such states will be able to provide extremely valuable information in human behaviour evaluation.

In the study reported in this thesis, we narrow down the domain of the investigation and specifically focus on predicting whether a subject is “Happy” or “Not happy”, “Relaxed” or “Not relaxed”, based on the data derived from handwriting fragments provided by participants in the experimental investigation. Although this is a preliminary study, at the time these experiments were carried out and the results published, this was the first formal research study directly focusing on emotion prediction from handwriting [79, 197]. The study has identified and addressed some of the principal important issues arising, in order to assist future work by providing evidence-based guidelines on how further in-depth research should be developed with a consideration of those issues. We believe that this work has laid down some fundamentally new initial evidence about the potential predictive capabilities of handwriting biometrics. The next section will explain the feature extraction process where features are extracted from raw handwriting data collected from the data capture specified in Chapter 2.

4.2 Handwriting feature extraction

To start the initial analysis, a template which includes twelve simple features was developed to represent each user's handwriting behaviour pattern, where the features were extracted from each user's writing samples from each task. Each task's set of features were mean and variance normalised. All the features that the template contains are commonly used in handwritten signature research [191, 198]. A list of the features utilised is shown in Table 4.1. As can be observed in the second column, there are two different types of features. There is an interesting and important issue about handwriting which relates to the different ways in which the data can be captured, leading to the availability of different feature types. “Dynamic features” and “Static features”. “Static features”are referring to features which characterise the overall form and appearance of the handwritten samples. Such features can be extracted following either “on-line” (which capture the handwriting segments as well as the execution of the writing) or “off-line” (which only capture the appearance of the handwriting segments) acquisition methods. As explained in Section 3.2.1, Static features are notable for the fact that they provide no time-based information, but reflect simply the appearance of the writing sample. They are thus available directly from off-line capture and can also be reconstructed from on-line acquisition. “Dynamic features” are features which characterise the actual execution of the writing activity and which explicitly reflect time-related information. Such features can only be acquired directly by means of an on-line acquisition process. There is also another category of so-called “pseudo-dynamic” features, which are dynamic features which can be generated from static data by inferring the dynamics from a static representation. However, these features are limited in number and, indeed, are most commonly used in human document analysis where they are observed by human inspection of the appearance of handwriting. This feature category is therefore not included in the experiments reported in this thesis.

4.3 Emotion ground truth labelling

After analysing the previously reported ground truth labelling methods, it appears that there is not a uniform approach in this area. For example, in [150] the

Feature number	Feature Type	Feature description
F1	Dynamic	The total time taken to execute the handwriting
F2	Dynamic	Average pen velocity in x direction
F3	Dynamic	Average pen velocity in y direction
F4	Dynamic	Average pen acceleration in x direction
F5	Dynamic	Average pen acceleration in y direction
F6	Static	Standard deviation of x coordinates values
F7	Static	Standard deviation of y coordinate values
F8	Static	Vertical centralness of the handwriting
F9	Static	Horizontal centralness of the handwriting
F10	Dynamic	Pen Azimuth
F11	Dynamic	Pen Altitude

Table 4.1: Extracted handwriting behaviour features

authors present the subjects with both sound and graphic contents, each content element being designed to "induce" one of the six commonly defined basic emotion categories (neutral, anger, fear, happiness, sadness and surprise). This approach was fundamentally different to what our study set out to achieve, since the underlying foundation of our study is to avoid any assumptions about the emotions that the participants were experiencing, but instead to determine emotional state more objectively by asking the participants to report their emotional state solely based on how they were feeling at each key point in the experiment. Therefore, in order to achieve this, a self-assessment questionnaire method was adopted, where the participants were given the freedom to reflect how and what emotion they were experiencing at the time of task execution. This approach can also be found in studies such as that reported in [175].

As mentioned in Chapter 1, this preliminary study was the first to explore the possibility of emotion prediction by using handwriting data. It was important to limit possible "noise" in the data that can confuse the classifier. Therefore, one of the most critical goals is to create a clearer separation of emotional state at the ground truth labelling phase. This will increase the likelihood that the labels assigned to participant samples genuinely reflect the emotional state that they were experiencing at the time of participation and, therefore, provide a more meaningful and accurate reflection of their state for the purposes of investigating the predictive process.

In this method of ground truth labelling, a threshold is introduced, the value of the threshold representing a point on a graded emotion score scale from (1 - 10).

For example, if the value of the threshold is 5 for the "Happy" emotion score, this means that all the participants who scored 1 - 4 on the "Happy" scale will be deemed to be experiencing a feeling of being "Not happy", and will be labelled as "Not happy", while a score of 6 - 10 will be labelled as "Happy". Generally, any participants who scored the 5, will be discarded from the prediction, therefore, create a clearer separation between the two states to be distinguished, and define a clear two-class classification task. In other words, for this initial study, the threshold is used to define two distinct for the subsequent experimentation. In order to have a clear overview before the value of the threshold was investigated in detail, the distributions of the emotional scores for each task were produced as charts:

- Happy

  Figures 4.1 - 4.4 illustrate the emotion score distribution for the "Happy" emotion across the 4 tasks. In Figure 4.1, for Task 1, which is the fixed writing task, it appears that for this first task, the most populated emotion score categories are 7,8, and the scores allocated roughly approximate to a normal distribution. Figure 4.2 (Task 2), the task requiring the participants to describe the "happy heart" picture (as shown in Figure 2.4), shows that the distribution of the emotion scores are very similar to Task 1, where the most populated emotion score categories are 7 and 8. However, here a few more subjects shifted into emotion score category 10. For Task 3, where the participants describe the "parents arguing" picture (as shown in Figure 2.5), the peak of the distribution appears to be shifted from 8 to 7 but is otherwise very similar to the previous case. In Task 4, the time-limited task, the distribution becomes more evenly spread over categories 5,6 and 7. A few participants even drift into the very lower end of the scale 1,2,3 and 4.

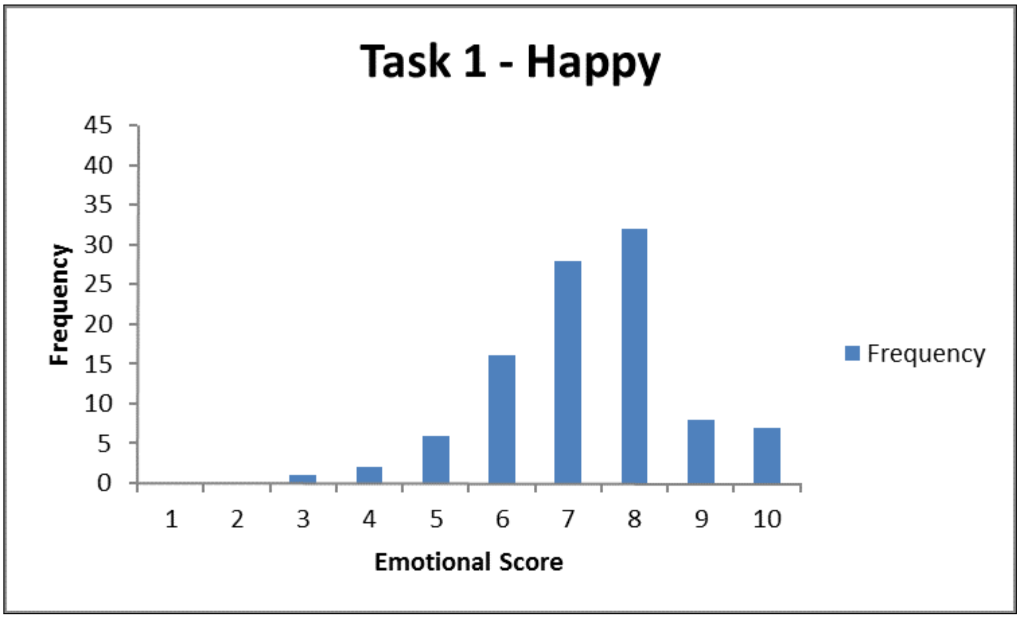

FIGURE 4.1: "Happy" score distribution for handwriting Task 1

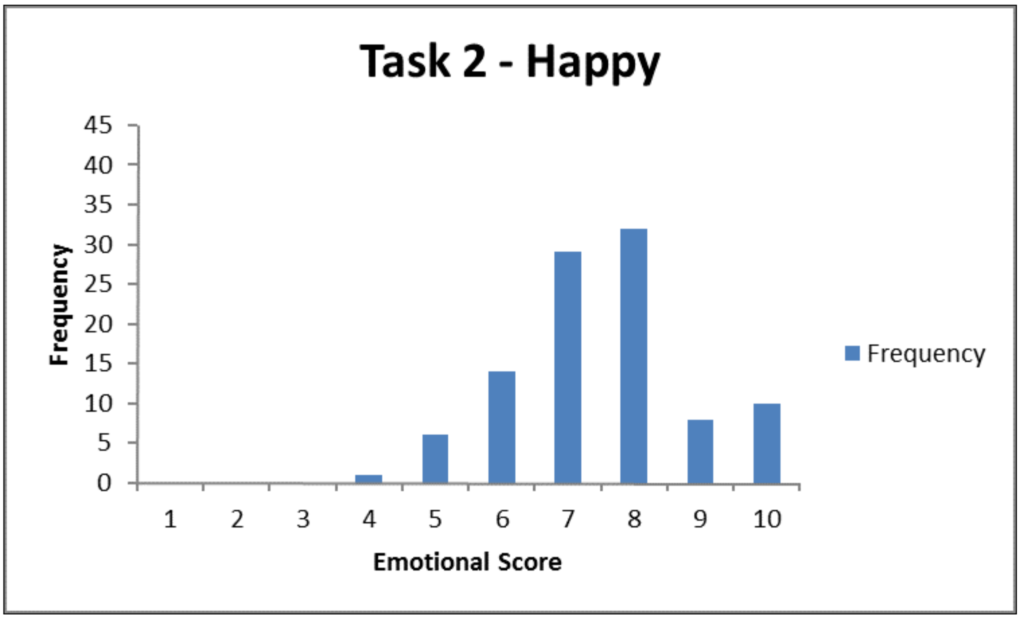

FIGURE 4.2: "Happy" score distribution for handwriting Task 2

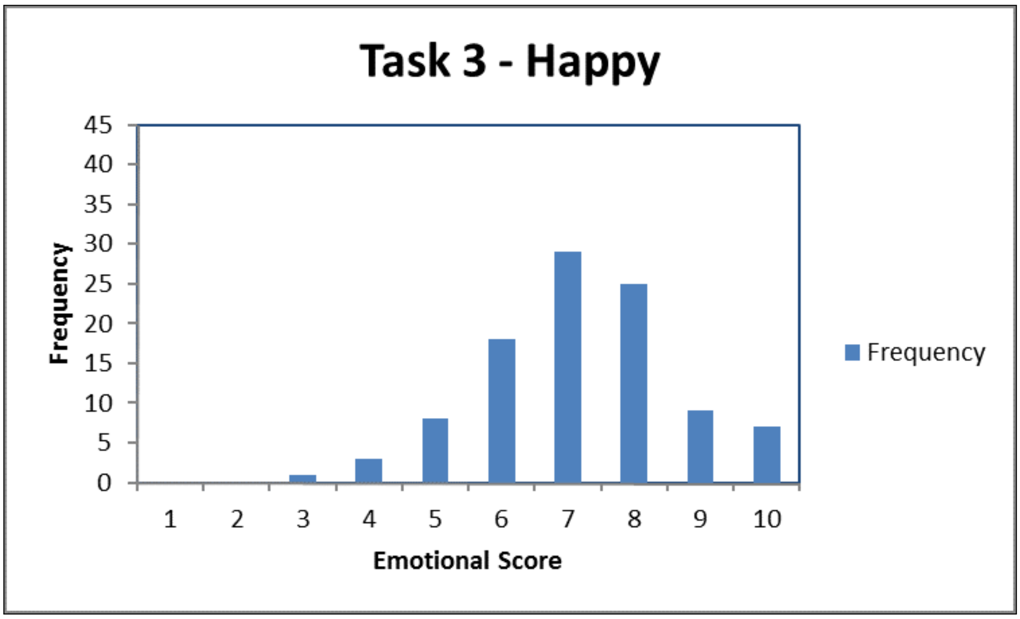

FIGURE 4.3: "Happy" score distribution for handwriting Task 3

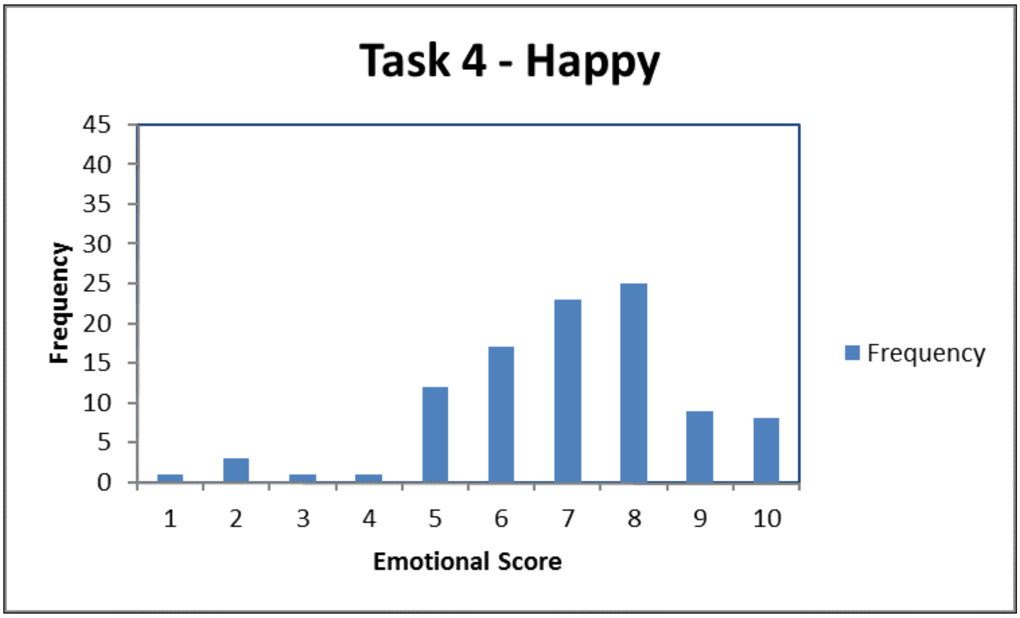

FIGURE 4.4: "Happy" score distribution for handwriting Task 4

- Relaxed

  Figures 4.5 4.8 present the emotion score distributions for the "Relaxed" emotion prediction across the 4 tasks. Task 1, the fixed writing task, shows that after the first task, the most populated categories are 6,7 and 8. For Task 2, where the "happy heart" picture (as shown in Figure 2.4) was shown to participants, the "Relaxed" emotion score distribution is shifted towards the higher end of the scale and the peak of the distribution changed from 7 to 8 from Task 1 to Task 2. For Task 3, the "parents arguing" picture (as shown in Figure 2.5) task, the peak of the distribution shifts from 8 to 7 but otherwise remains largely the same. Task 4, the timed task, caused a few participants even to drift into the very lower end of the scale with values of 1,2 and 3. The rest of the distribution pretty much remains the same as for Task 3.

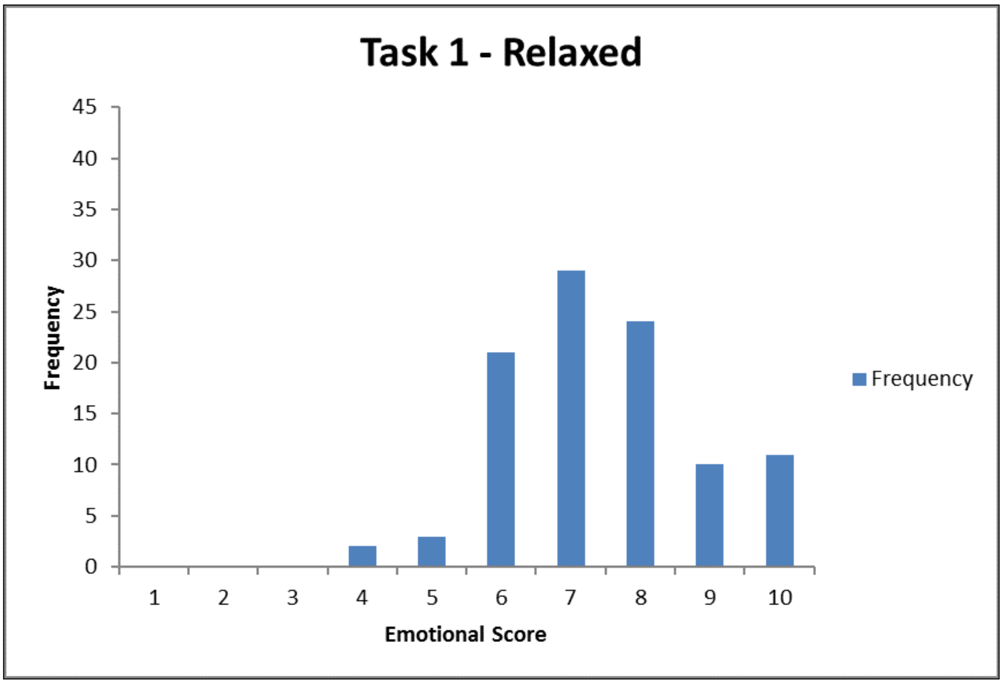

FIGURE 4.5: "Relaxed" score distribution for handwriting Task 1

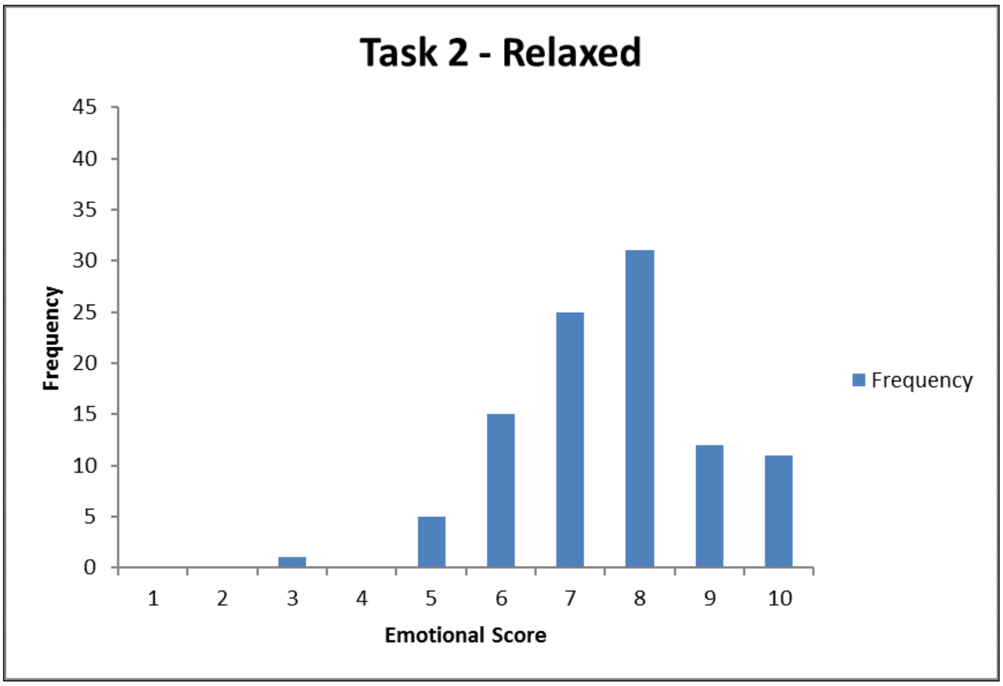

FIGURE 4.6: “Relaxed” score distribution for handwriting Task 2

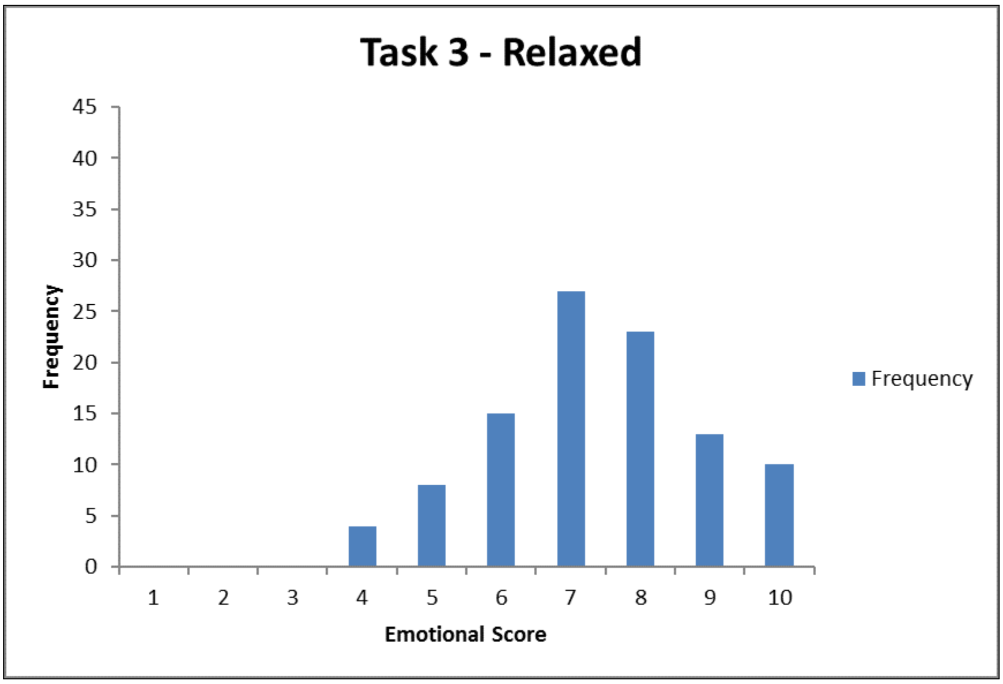

FIGURE 4.7: “Relaxed” score distribution for handwriting Task 3

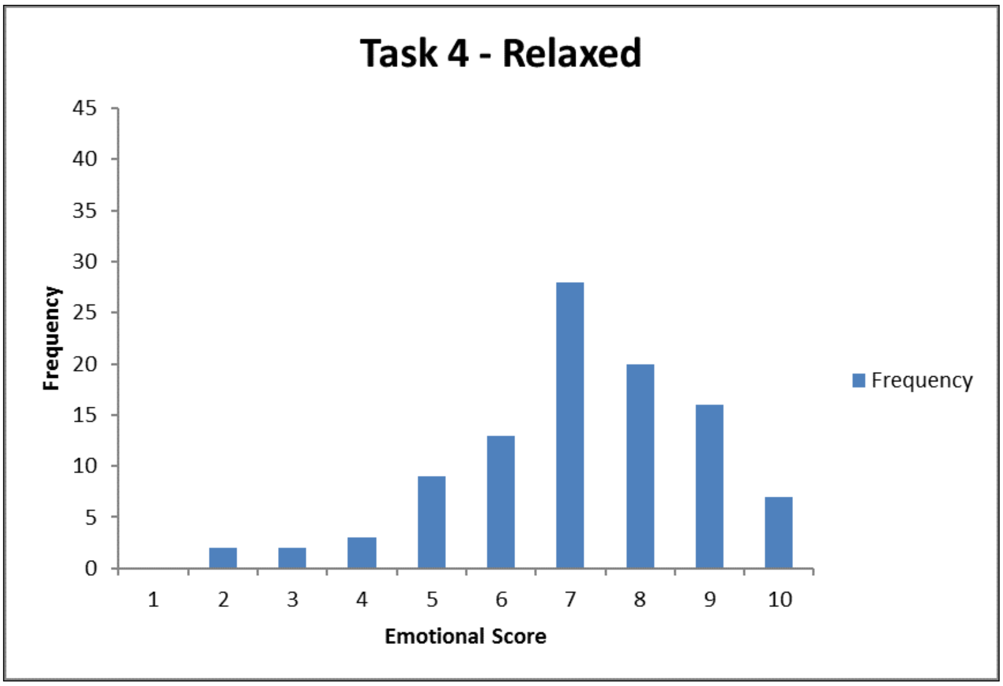

FIGURE 4.8: "Relaxed" score distribution for handwriting Task 4

The scores all appear to be distributed approximately as a normal distribution, where the peak values for both emotions are around 7 and 8. The threshold "8" was applied for this study, as this threshold value can create two classes where one class ("Happy" subjects) are clearly separated from the other class, thereby increasing the likelihood of finding clearly distinguishable handwriting behaviour patterns. This threshold was used in all the published work based on this method arising from this study. However, a more detailed study using other options for the threshold values was also carried out, and a detailed discussion of these specific experiments will be presented in Section 4.5.1.

4.4 Emotion prediction

After generating the handwriting features from the acquired raw data and producing the corresponding labels for each user and each task, the immediate following phase of the data analysis process is to take the features extracted, and introduce the labelled data into the emotion prediction classification process itself, in order to explore any correlations which can be found between the mental states of the subjects and their handwriting behaviour, as captured by the feature distributions.

In the experimental study to be reported, three different classification methods were utilised for processing the handwriting features, the classifiers adopted consisting of a K-NN(K=1) approach, the Jrip classifier and an SVM (support vector machine), using a "leave one out" cross validation methodology. This allowed a variety of different classification methodologies to be investigated. Briefly, the important characteristics of these classifiers are as follows, more details to be find in [197]:

- K-Nearest Neighbour (KNN) classifier: The KNN [199] algorithm defines a simple classifier which does not require a sophisticated training phase. Its only requirement is that the labels of classes represented by the samples are available. Thus, to find the nearest neighbour, an appropriate distance metric (simple Euclidean distance is used in this study) is calculated between each test sample and all training samples. The class label corresponding to the K minimum distances in this list determines the assigned class label (since K=1 in this study, this reduces to a process of using the label of the sample which is closest to the training set).

- Jrip classifier: The Jrip [200] classifier is an optimised IREP (Incremental Reduced Error Pruning) classifier, It is based in association rules with reduced error pruning (REP), a very common and effective technique found in decision tree algorithms to decrease the error rates of a dataset with noise. The IREP uses a "divide to conquer" approach and the training data is split into a growing set and a pruning set. First, an initial rule set is formed that over the growing set, using a heuristic method. This over-large rule set is then repeatedly simplified by applying one of a set of pruning operators (typical pruning operators would be to delete any single condition or any single rule). At each stage of simplification, the pruning operator chosen is the one that yields the greatest reduction of error on the pruning set. This process is repeated until there are no unacceptable errors. Our implementation uses a delayed pruning approach to avoid unnecessary pruning, resulting in a Jrip procedure.

- Support Vector Machine (SVM): The SVM [201] classifier aims to minimize the structural risk, which means the SVM tries to increase the performance when trained with known data based on the probability of a wrong classification of a new sample. The SVM classifier is based on maximizing the margin between two classes with the decision surface/hyperplane based on an induction method, which minimizes the upper limit of the generalization error related to uniform convergence. Therefore, this hyperplane divides the training set into positive and negative groups, and selecting the surface which keeps more samples.

These algorithms were applied to each task for both the “Happy” and “Relaxed” emotions. The predictions are subject-independent, where only one sample from each subject is available for each task for analysis. Figure 4.9 presents a comparison between the predictive accuracy achieved by each classifier for the attempt to predict the “Happy” emotion. Figure 4.10 presents a comparison between the predictive accuracy achieved by each classifier for the attempts to predict the “Relaxed” emotion.

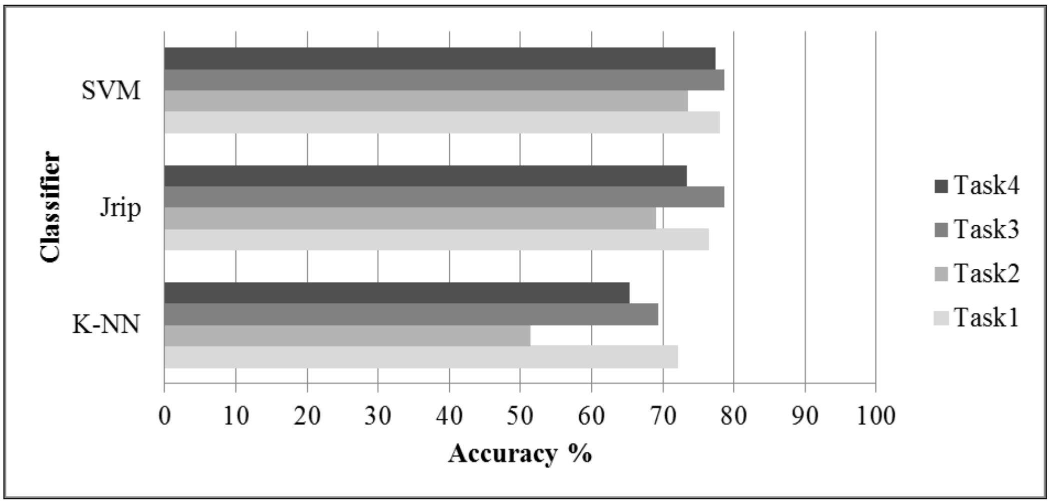

Figure 4.9: Prediction accuracy for the “Happy” emotion for Task 1- 4 by the three chosen classifiers

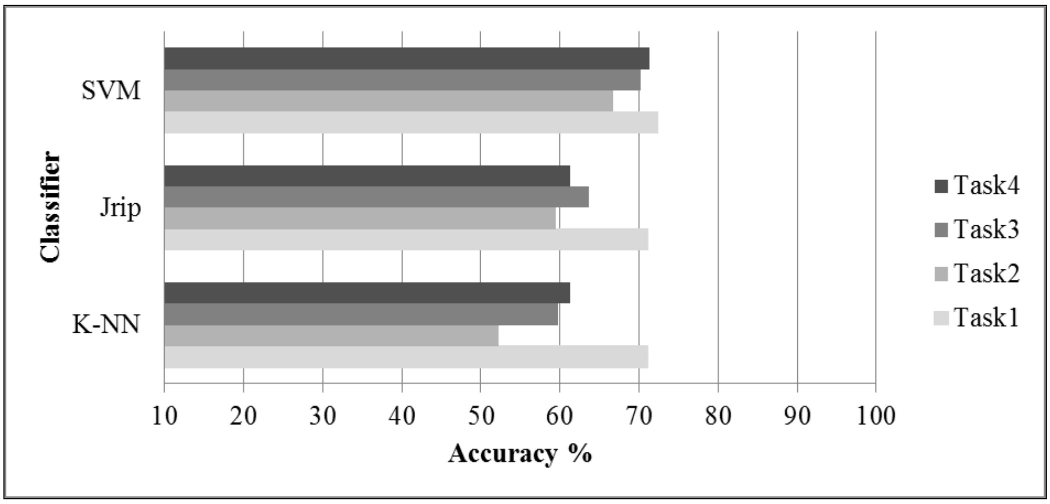

FIGURE 4.10: Prediction accuracy for the "Relaxed" emotion for Task 1- 4 by the three chosen classifiers

As demonstrated in Figure 4.9, the best prediction accuracy was achieved by using the support vector machine (SVM) classifier, which is the case for almost all tasks, the performances showing close to 80% correctly predicted "Happy" subjects. The Jrip classifier was more task dependent, where tasks 1 and 3 return around the 80% accuracy, but the other two tasks are around 70% accuracy. The KNN classifier showed the most variable performance, ranging from 53% - 72% accuracy, and the predictive accuracies obtained are the worst when compared to the other two classifiers tested.

In Figure 4.10, the "Relaxed" emotion prediction is generally similar to the case for the "Happy" emotion prediction. The SVM classifier achieved the most consistent predictive accuracy, which is around 70%. However, for the Jrip and KNN classifiers, the prediction accuracies achieved were around 60%, expect for Task 1, where the predictive accuracy is around 72% across all three classifiers.

The initial results raised a few issues providing important insights, challenges and lead to the following observations:

- The results suggest that from analysing human handwriting behaviour patterns, it is possible to identify and predict the "higher-level" states previously described which reflect the current emotional state of the subjects.
- The classification infrastructure will influence the prediction accuracy.

- The predictive capability can be influenced by the parameters of the writing task. For example, Task 1(the fixed typing task), where the content of the text was chosen to include a full range of different elements which characterise the language structure, is a case in point. Task 1 seems to offer a richer feature template and lead to better performance than both Task 2 and Task 3(free typing tasks), where the content is not fixed and the text provided can vary considerably, depending on the individual choice of specific text.

- As the contents of Task 2 and Task 3 are uncontrolled, therefore, the predictive accuracy for these two tasks varies more than fixed-text Task 1. To be more specific, one of the main factors to consider is the amount of raw data available. To be clearer, both the variability of the task (with respect to the generated content) and the amount of raw data influence the predictive accuracy. However, it appears that content variability is the main influence. Figure 4.11 and Figure 4.12 demonstrate the distribution of amounts of raw data that were collected from the 100 participants during Task 2 for the "Happy" and "Relaxed" emotions. The populations were divided by the number of words that the individuals provided and, at the same time, the total number of participants in each group are broken down into groups for which the emotional state was "incorrectly" and "correctly" predicted. As the distributions suggest that the amount of the raw data can influence the predictive capability, this allows a notional "optimum" amount of raw data to work with to be suggested.

- The performance difference between tasks can be reduced to some extent when a suitable and adequately powerful classifier is introduced during the processing phase. This point is illustrated in Figure 4.9 and Figure 4.10, where the similar level of predictive performance was achieved across the four tasks by SVM classifier.

As mentioned in Chapter 2, the tasks were designed using the format which, it might be expected, could "nudge" the participants towards one emotion of interest over another. For example, Task 2 might be expected to emphasise a more "Happy" frame of mind, while Task 4 was designed to stimulate a less "Relaxed" emotional response. Figure 4.13 and Figure 4.14 presents the emotion score distribution for "Happy" and "Relaxed" emotions across all tasks, which does show the distributions are changing based on the nature of the task.

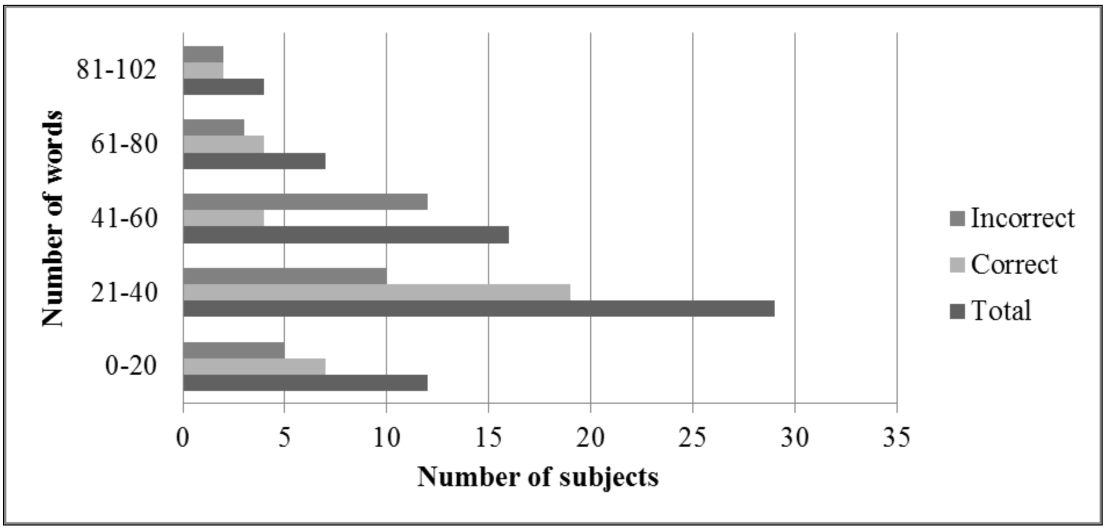

FIGURE 4.11: Number of words vs correct prediction of assigned group for Task 2 "Happy" emotion

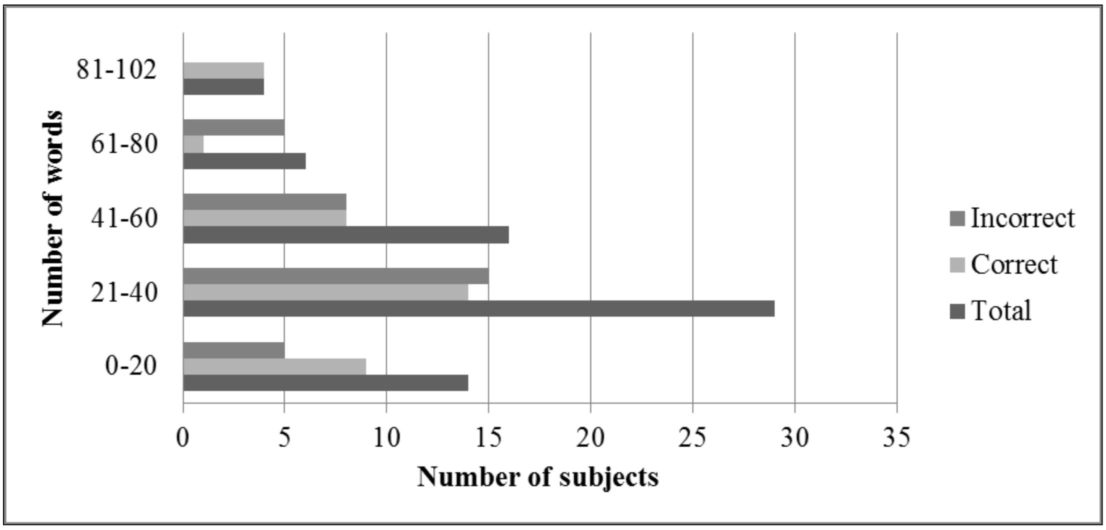

FIGURE 4.12: Number of words vs correct prediction of assigned group for Task 2 "Relaxed" emotion

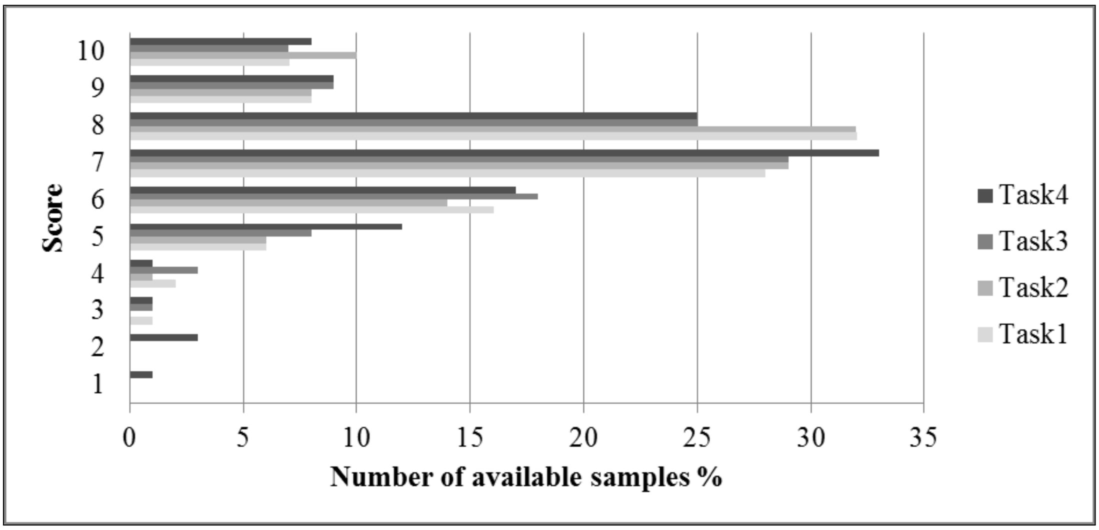

FIGURE 4.13: Score distributions for all tasks and for "Happy" emotion

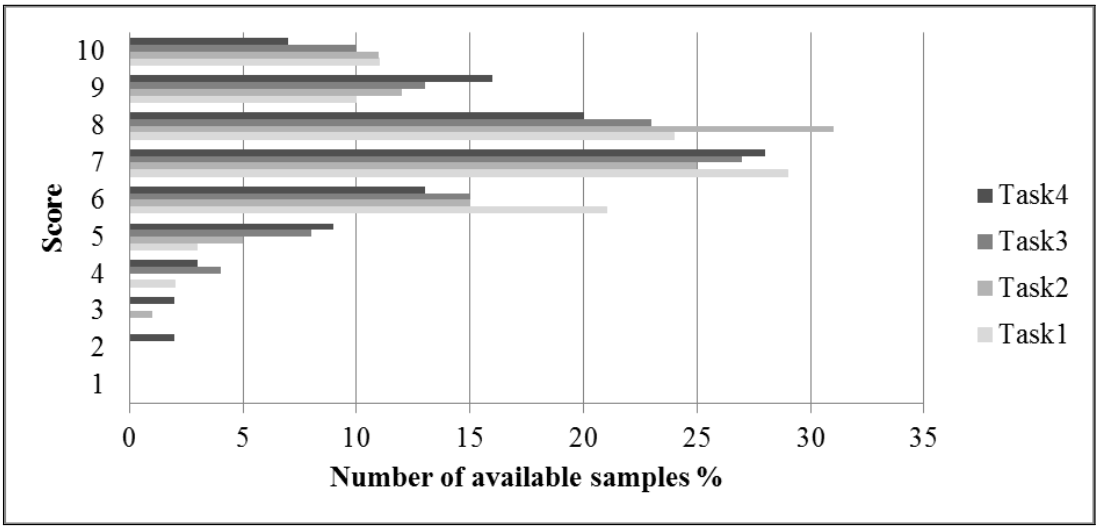

FIGURE 4.14: Score distributions for all tasks and for "Relaxed" emotion

This initial analysis presents some rather encouraging results while, at the same time, raising some very valuable and constructive points that could be very beneficial towards the further development of this type of study.

In order to explore this further, a set of more in-depth experiments will be presented in the next section, where three variables of the emotion prediction process are investigated with different settings, in order to observe the impact that each one can have on the achievable predictive performance.

4.5 Further analysis

In order to achieve a more comprehensive understanding of the results presented here, the following further investigations of the approach adopted were carried out and analysed:

- Threshold value assignment and the effects which different threshold values have on the predictive experimentation.
- The effects of different types of feature sets (especially in relation to static and dynamic feature distinctions) on the predictive experimentation.
- The effects of using different tasks datasets (especially in relation to differences between using fixed and unconstrained written content) on the predictive experimentation. This will also allow an observation of the effectiveness of using inter-task training/test datasets.

4.5.1 Threshold assignment

As mentioned in Section 4.3, in the present study, a threshold value on the numerical emotion scale for the “Happy”and “Relaxed” emotion is adopted, where the chosen threshold value divides the population into two classes. For example, for all the of the emotion prediction experiments reported above, the threshold was set at point 8 on the scoring scale for both the “Happy” emotion and the “Relaxed” emotion. Thus, the “Happy” or “Relaxed” label would be assigned to subjects who rated themselves 9 or 10 on the available scale, and an “Not happy”

Threshold accuracy (%)	Gr1	Gr2	Predictive Accuracy %
6	9	75	89.3
7	25	47	65.3
8	53	15	77.9

Table 4.2: Accuracy of the “Happy” prediction while applying different threshold values for Task 1

or “Not relaxed” label would be assigned to subjects who rated themselves from 1 to 7. The remaining subjects are excluded in the experimentations.

Table 4.2 demonstrates the different results obtained for three different choices for this threshold value, when using the SVM classifier to process samples from the dataset relating to Task 1. This task is the fixed task where the textual context is seen to be the richest out of the whole set of tasks, and thus helps to illustrate the principal point of interest investigated in this experiment. In Table 4.2, each row presents respectively, the threshold value, the number of subjects falling into Gr1, the number of subjects falling into Gr2 and the predictive accuracy recorded. The first row indicates that the predictive accuracy achieved was 89.3%, when the threshold value was set at 6 and Gr1 has 9 participants and Gr2 has 75 participants. The second row present the predictive accuracy achieved when using a 7 as the threshold value, where Gr1 has 25 participants and Gr2 has 47 participants, was 65.3%. The third row shows the case when a value of 8 was adopted as the threshold value, which is the setting that was adopted for the emotion prediction experiments reported in this chapter.

The results obtained from the experiments where different threshold were assigned for the “Happy” emotion clearly reflect the fact that applying different thresholds at the data labelling phase certainly influence the predictive accuracy. When the threshold is set at 6, the predictive accuracy increased to almost 90%, which is a significant improvement while, at the same time, when the threshold value change to 7, the predictive capacity dropped substantially. This demonstrates that interpreting the emotion score appropriately is an important and challenging task, especially with limited data.

Another important factor is that the number of subjects in Gr1 and Gr2 change significantly when the threshold values are different. This phenomenon could potentially have an impact on the predictive accuracy when changing the threshold. The closest to having a balance between Gr1 and Gr2 (where Gr2 is still almost twice the size of Gr1) is when the threshold value is set to 7, which might suggest that these settings better reflect the likely predictive capability achievable in real life scenarios.

In this field of study, establishing an objective ground truth is undoubtedly a very difficult task, and the limitation of data availability needs to be addressed before any more explicit results can be discussed and conclusions can be drawn. This will be taken up again in a later chapter.

4.5.2 Effects of different feature sets in emotion prediction

As mentioned in Section 4.2, Two types of features can be extracted, which are static features and dynamic features, and they were combined when the emotion prediction were performed. In this section, the influence of the type of features, on emotion prediction accuracy will explored and compared.

Table 4.3 illustrates the performance comparison when different types of feature sets are adopted. For illustrative purposes the experiments were carried out for predicting the "Happy" emotion for Tasks 1 - 3, using the KNN classifier. Four commonly used static features were used as part of an initial feature set for the investigation of emotion prediction, which are *Standard deviation of x coordinate values, Standard deviation of y coordinate values, Vertical centralness of the handwriting, Horizontal centralness of the handwriting* defined in Section 4.2, predictive results were generated for the "Static features" column and the eight commonly used dynamic features, which are *The total time taken to execute the handwriting, Average pen velocity in x direction, Average pen velocity in y direction, Average pen acceleration in x direction, Average pen acceleration in y direction, Pen Azimuth, Pen Altitude, Pen Pressure on capture tablet*, were defined in Section 4.2, were used to process the results for "Dynamic features" column. Table 4.1 presents the list of specific features with each feature type labelled. In Table 4.3, each row presents results for a task. Task 1 is the fixed task where the subjects copy the same predefined text, and Task 2 and Task 3 are the unconstrained tasks, where the content is unconstrained.

It is seen that for Task 1, the predictive accuracy increases as more features are included for the classification process, which is broadly to be expected [202, 203].

Task type	Accuracy %		
	All features	Static features	Dynamic features
Task 1	72.1	57.4	69.1
Task 2	51.5	54.4	55.9
Task 3	69.3	73.3	65.3

Table 4.3: Accuracy of "Happy" prediction with different feature sets for Task 1 - 3

When only the static features are used, the predictive accuracy drops by about 15% in comparison to the results obtained, where a mix of all features were used. When only the eight dynamic features were used the performance achieved was around the similar level of when mix of all features were used. This confirms what most studies have shown, namely that the dynamic handwriting features have a greater impact on reflecting the characteristics of the handwriting behaviour than the static features alone. Thereafter, three simple additional static features (one related to x/y perspective and two relating to x/y axis local pen excursions) were added, to create a more comprehensive feature set. With these additional features, the predictive accuracy for the "Happy" emotion for Task 1 improved from 57.4% to 64.7%. This shows great potential for the possibility of optimising the adopted feature sets for the further development of the predictive technique in the future.

For the unconstrained tasks (Task 2 and Task 3), the pattern of how the predictive performance is influenced is not as clear-cut as for the fixed task. One of the main reasons for the inconsistency is that the free tasks provide a greater variability in the actual data collected and, therefore, the predictive performance is more dependent on the individual responses.

4.5.3 Inter-task operation

As shown earlier, the achievable predictive accuracy can be influenced by the nature of the feature sets or the emotion labelling methods. The experiments discussed were designed to contain different types of task, specifically those represented by Task 1, where all participants are asked to provide the same text, and also Task 2 and Task 3, which are in a free format, where the participants are not

Train	Test	Accuracy %
Task 1	Task 1	72.1
	Task 2	69.1
	Task 3	63.2
Task 2	Task 1	55.9
	Task 2	51.5
	Task 3	61.8
Task 3	Task 1	66.7
	Task 2	50.7
	Task3	69.3

Table 4.4: Accuracy of the "Happy" prediction using feature sets from different tasks in testing and training sets

constrained in terms of the content they wish to provide but instead can write whatever they wish. The approach adopted to generate the results was to treat each task as an individual dataset, and all the experiments were task-centred, in the sense that only data generated within a single task are used for training and testing. Therefore, it is a useful exercise to investigate if there are any effects on performance when using data from a mixture of the tasks. This is valuable and of practical value since, often in real life scenarios, the test samples and the training samples may be gathered under completely different conditions and in differing environments.

Table 4.4 displays the predictive performance achieved when training samples and testing samples are selected from handwriting data collected from the same individuals, but from different tasks. Again, the experiment was set up for the prediction of the "Happy" emotion and using the KNN classifier.

The results presented again demonstrate the benefits of the richness of the data collected in Task 1(fixed task). When data from Task 1 are used in testing, the predictive accuracies are typically higher than when data from the other two tasks are used in testing. The predictive performance clearly varies, which could also be expected as this is consistent with the observations noted before, when the free tasks (Task 2 and Task 3) are used in testing, irrespective of which task the training data are drawn from. Again, there are some valuable lessons here which can be learned in relation to the possible adoption of the proposed techniques in practical scenarios, where the availability of testing samples can often be limited or where little control over the source data can be exercised, but where there may sometimes be an option in relation to obtaining training data.

This experimentation therefore provides some insight into how the predictive capability of analysing handwritten fragments can be more fully exploited by optimising the development of a practical strategy to improve confidence in "higher-level" mental state prediction from handwriting biometrics.

4.6 Conclusion

This chapter has described the experiments and emotion prediction results obtained from analysing handwriting data across various different tasks and experimental conditions. Feature extraction was the first step: by generating a basic set of features that were commonly used in other studies reported in the literature, an initial set of preliminary results was extracted and analysed, in order to provide some insight into mental state prediction from handwriting biometrics. This is especially valuable because it is an area which has not been extensively researched to date. At the time when our work was initially published, there were no reported studies to be found in the literature which directly addressed emotion prediction from handwriting data that captured by digitising tablets.

Another challenging task at the feature extraction phase is the ground truth labelling, and this chapter has presented the approach adopted and the way in which emotion-indicative scores could be assigned to the data that were captured. Emotion score distribution charts were generated to demonstrate and characterise the allocation of the emotion data. A notion of "optimum" parameter settings for ground truth labelling was introduced and developed, based on the investigation of the influences that different parameters of labelling method can induce with respect to the predictive performance attainable.

After defining the parameters of the experiments and the feature sets generated, three different classification infrastructures have been investigated for emotion prediction. The preliminary experiments reported raised a number of important insightful issues and challenges. At the same time, the initial results were also encouraging, achieving almost 80% predictive accuracy for the “Happy” emotion prediction and almost 75% predictive accuracy for the “Relaxed” emotion prediction.

In the last section, we have addressed three of the main areas where different settings could potentially influence the predictive performance. These included studying the impact when different ground truth labelling threshold value settings, different categories of feature sets and different categories of tasks are used for emotion prediction. The study has shown that, in order to develop an “optimised” prediction strategy, it is important to take those options for the experimental parameters into consideration, as these points will have significant effects on the confidence when drawing any meaningful conclusion from the predictive performances. These results also demonstrate the performance bounds likely to seen in practical adoption of these techniques in different practical scenarios, perhaps where the prevailing conditions are predetermined and not controllable for analytical purposes.

The predictive capability of handwriting biometrics can be potentially beneficial in many real life scenario, as reported in Chapter 1, and it is becoming increasingly apparent that the field of forensic investigation, to take a major example, can benefit significantly by adopting new methods and technology which cross traditional boundaries. A typical forensic investigation situation, where without a specific identification description, any reasonable traits prediction from analysis the handwriting sample of the individual in question, can be tremendously helpful. For example, even if a precise subject cannot be identified, by using the traits such as gender or the emotional state that the subject was in, can help eliminate some confusing information, narrow down the search domain, and so on. Acknowledgement of an individual's emotional state, such as whether they were likely to be “Happy” or “Not happy”, “Relaxed” or “Not relaxed”, might all provide an extra dimension of information which could be very valuable while investigating human behaviour or interpreting particular forensic scenarios, thus assisting the efficiency and effectiveness of forensic investigations.

The experiments reported have taken us a few step closer towards establishing a more comprehensive (“optimised”) approach for investigating mental state prediction by using handwriting biometrics, and have also shown there are likely to be great benefits to be gained from more in-depth studies in the future. In the next chapter, another hand-orientated biometric modality (Keystroke biometrics) will be investigated, in order to provide a more comprehensive overview of "higher-level" mental state predictive capability by using hand oriented behaviour biometrics.

CHAPTER 5

Prediction of higher-level states from keystroke data

Handwriting biometrics have been shown to have significant potential for predicting higher-level mental states, as presented in Chapter 4. This is one positive step towards finding out the answer to a main question that this thesis is focused on, which is "To what extent can mental status be predicted based on hand-oriented behavioural biometrics?" In order to provide a more broadly-based answer to the question, another one of the main hand-oriented behavioural biometrics has also been studied, namely "Keystroke Dynamics". The biometric modality based on keystroke dynamics - is also a behavioural biometric. This behavioural biometric reflects the behaviour pattern executed by individuals while they are interacting with a computer keyboard, and is therefore potentially of considerable significance in many practical situations.

Keystroke dynamics, like any other biometric measurements, are frequently adopted for individual identification purposes [18]. But also, as presented in Chapter 1, there have been a small number of previous studies which utilise keystroke dynamics for emotion prediction. The results reported are suggesting that Keystroke Dynamics have the predictive capability for "higher-level mental states prediction". This type of predictive task using keystroke data is still an area that is not extensively explored, but the positive prediction results reported in the literature [173, 204, 205] lay down the fundamentals for our proposed research. It was clear that there were no publicly available databases for predicting emotions from keystroke data and each of reported studies have developed their own database, with the largest database containing samples from only 50 participants, as demonstrated in Table 1.1. The uniform principle of data collection proposed and the limited size of available databases present even more challenges in this area of research, as these have made benchmarking results and drawing conclusions a very difficult task. As presented in Chapter 2, the data capture principle that was designed for this research study is aiming to address both these issues.

In this chapter, the newly captured database which contains sample from 100 participants will be used for the investigation of higher-level mental state, or emotional state prediction. The database, as described in Chapter 2, contains double the number of samples the biggest database previously reported. From the encouraging emotion prediction results, presented in Chapter 4, which were generated from the handwriting biometrics modality, it is interesting and valuable to uncover the predictive capability of the keystroke samples collected from the same 100 participants.

For the investigation of the results, this chapter will report the feature extraction process, including the keystroke features and ground truth labelling. The features generated will then be processed with some commonly used classification infrastructures to perform the mental state prediction.

After observing and discussing the results, the last section of this chapter will also describe experiments that explore the impact on predictive accuracy when different number of features used and, different amount of raw data used to generate features.

Section 5.1 will briefly review the relevant studies that use keystroke dynamics for emotion prediction. In Section 5.2, the feature extraction process will be presented with a detailed list of keystroke feature used. Section 5.3 will present ground truth labelling method with the emotion score distribution charts. Section 5.4 will present the emotion prediction methods and results. Section 5.5 will present a further investigation of the predictive accuracy achievable when two of the fundamental characteristics of the features are tested with different parameters. Section 5.6 will make some concluding observations.

5.1 Keystroke biometrics

As presented in Chapter 1, a number of published studies have investigated emotion prediction based on keystroke data, as reported in, for example, [150, 173, 175, 176]. In the work reported in [173], the study utilized background key event collection software to record the keystroke data. The database used in this experiment included just 12 participants. The experiments were set for recognizing emotions include confidence, hesitance, nervousness, relaxation, sadness and tiredness. The study adopted a 5- point Likert scale for emotion labelling. The best achieved predictive accuracy was ranging from 77 to 88% when predicting confidence, hesitance, nervousness, relaxation, sadness and tiredness, and also showing a result of 84% when predicting anger and excitement. In [175], the study also used key event collection software to collect a database containing 24 subjects, this study achieving an accuracy of around 75% for predicting cognitive stress. For emotion labelling, this study adopted a 11-point Likert scale.

The authors adopted a rather less standard method in [150], which utilized a keyboard with a pressure sensor. The collection software recorded both the key events and pressure sequence. The data were labelled according to the specific task content presented to the participants. This study managed to collect data from 50 participants and achieved overall predictive error rate ranging from 4.4% - 8.4% when predicting emotions including: neutral, anger, fear, happiness, sadness and surprise. In [176], the authors recorded the pressure sequence and key events, but also recorded mouse movements during the whole session. The database collected for this study consisted of 24 participants. Emotional state was determined by means of a 7- point Likert scale. This study reported 83% of subjects showed increased key pressing pressure when under stress.

The literature provides some useful insights into the process of predicting emotion from keystroke data and has shown some encouraging predictive performance results while, at the same time, the reported studies have also left a few areas that need further exploration.

First of all, the size of the databases adopted in the studies reported range from only 12 subjects, up to a maximum of 50 subjects. Such databases, the smallest of which is, frankly, unacceptably small. The largest is, at best, modest in size therefore incorporate significant limitations in relation to drawing any representative conclusions from the results. Therefore, the newly collected keystroke database compiled specifically for the current project, and incorporating data from 100 participants, makes the size of our database double the previously largest database. This therefore provides a much better opportunity to achieve a more robust and representative predictive performance.

In [150], the experiment was set up so that the emotion labels were generated by assuming that the participants will feel the specific recorded emotion when the pre-labelled tasks were presented to the participants. This is probably not the best approach as task content can have a very subjective effect on the participants' emotional state, which cannot necessarily be reliably predetermined. Therefore, our experimental procedure adopted a 10 point Likert scale questionnaire in order for the participants to reflect their emotional state on an individually determined basis.

5.2 Keystroke feature extraction

The keystroke data were captured with the specifically designed software, which was described in detail in Chapter 2, in order to collect the raw data of keystroke behaviour. The software was set up to collect two items of timing data for a single keystroke, which are, first, the time-stamp of the moment a key on the keyboard is pressed and, second, the time-stamp of the moment at which that key is released, as demonstrated in Figure 5.1, where the "t" axis represent the continuous time axis. P1 represents the time-stamp of "Key 1" been pressed and R1 represents the time-stamp when "Key 1" has been released. The values assigned to the time-stamp value are typically very large numbers, given that they are collected with the precision of milliseconds or in our case nanoseconds, therefore the time-stamp data are generally normalized for feature extraction. Therefore, the "Duration" and "Latency" are calculated as the basic representation of keystroke data. As presented in Figure 5.1, "P1R1" represent the "Duration" of the "Key 1" (which is the later time-stamp value R1 subtracted from the time-stamp value P1), while "R1P2" represents the "Latency" before the second key "Key 2" is pressed (which is the later time-stamp value P2 subtracted from the time-stamp value R1).

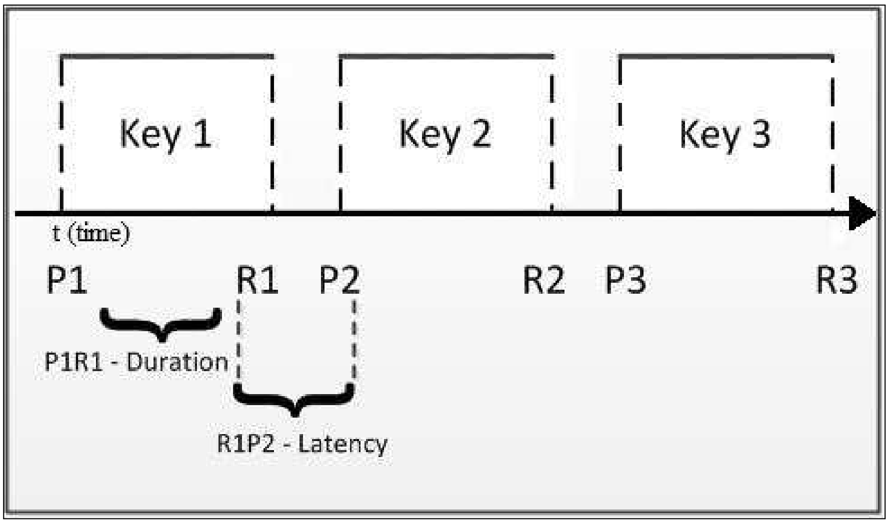

FIGURE 5.1: "Duration" and "Latency" measures to characterise keystroke patterns

As shown in Table 5.1, there are 29 features extracted from the newly captured database for this study. "Duration" and "Latency", as described above, are used as well as other timing features (such as “R1R2” and “P1P2”) to represent the pattern of the keystroke behaviour. In addition to the timing features, five commonly used behavioural habit-related features (such as “Correction rate” and “Input rate” are used. Correction rate is obtained by dividing the total number of times the “Delete” and “Backspace” keys are pressed in a task, by the total number of key presses in the overall task. “Input rate” is calculated by dividing the total number of key presses in a task, by the total duration of the task.) These features were included to build a more comprehensive behaviour pattern template. In all, a total of 29 features were defined and are extracted from each individual's keystroke activity.

Feature code	Feature description
Di_P1R1_Mean	Mean of the duration for the first key in a digraph.
Di_P1P2_Mean	Mean of the time take to get from the first key to the second key in a digraph.
Di_R1R2_Mean	Mean of the time take after release the first key to release the second key in a digraph.
Di_R1P2_Mean	Mean of the latency between the first key and second key within a digraph.
Di_P1R1_SD	Standard Deviation of the duration for the first key in a digraph.
Di_P1P2_SD	Standard Deviation of the time take to get from the first key to the second key in a digraph.
Di_R1R2_SD	Standard Deviation of the time take after release the first key to release the second key in a digraph.
Di_R1P2_SD	Standard Deviation of the latency between the first key and second key within a digraph.
Tri_P1P2_Mean	Mean of the time take to get from the first key to the second key in a tri-graph.
Tri_P1R1_Mean	Mean of the duration for the first key in a tri-graph.
Tri_R1P2_Mean	Mean of the latency between the first key and second key within a tri-graph.
Tri_P2P3_Mean	Mean of the time take to get from press the second key to the press the third key in a tri-graph.
Tri_P2R2_Mean	Mean of the duration for the second key in a tri-graph.
Tri_R2P3_Mean	Mean of the latency between the second key and third key within a tri-graph.
Tri_R3P3_Mean	Mean of the duration for the third key in a tri-graph.
Tri_P1R3_Mean	Mean of the total duration of the whole tri-graph.
Tri_P1P2_SD	Standard Deviation of the time take to get from the first key to the second key in a tri-graph.
Tri_P1R1_SD	Standard Deviation of the duration for the first key in a tri-graph.
Tri_R1P2_SD	Standard Deviation of the latency between the first key and second key within a tri-graph.
Tri_P2P3_SD	Standard Deviation of the time take to get from press the second key to the press the third key in a tri-graph.
Tri_P2R2_SD	Standard Deviation of the duration for the second key in a tri-graph.
Tri_R2P3_SD	Standard Deviation of the latency between the second key and third key within a tri-graph.
Tri_R3P3_SD	Standard Deviation of the duration for the third key in a tri-graph.
Tri_P1R3_SD	Standard Deviation of the total duration of the whole tri-graph.
Total_keyEvent	Total number of key stroked during the task.
CR	Correction rate - number of times they correct a letter/Total_keyEvent
FKR	Function key rate - number of times they used a function key/Total_keyEvent
IPR	Input rate - Total_keyEvent/total duration
PR	Pause rate - Total of latency / Total_keyEvent

Table 5.1: Keystroke features

5.3 Emotion ground truth labelling

In new dedicated database we have captured, the emotion scores were acquired for both the handwriting tasks and the keystroke tasks. As reported in Section 4.3, for handwriting, a "threshold" method was applied and details of the implementation of this method is presented in Chapter 4. This method is shown to be effective when applied for emotion prediction with the handwriting data from our newly implemented database. Therefore, the same approach is also adopted for the emotion ground truth labelling process for keystroke data.

5.3.1 Keystroke emotional scores distribution

For the keystroke-based experiments, the emotion scores were captured from the end of Task 2, for clarity, Tasks 2 - 5 are respectively presented as Tasks 1 - 4 in Figure 5.2 - 5.9, respectively. All the emotion scores are plotted in a bar chart form, in order to present the distribution of the emotion scores that were captured.

- Happy

  Figures 5.2 - 5.5 illustrate the emotion score distribution for the "Happy" emotion across the four tasks. Task 1, the maze task, shows that the most populated category is 8. For Task 2, the video clip task, the population for score 8 is significantly increased by about 30% in comparison with the previous task and fewer participants fell into the categories 2,3,4 and 5. In Task 3, the picture task, Categories 7 and 6 show an increase in the number of participants included where Categories 10 and 8 show a clear decrease in the number of participants. Categories 7 and 8 are the most populated categories. For Task 4, the timed task, it appears that category 8 is still the most populated category, although more participants fall into category 7. A few participants even fall into the category 1 and 2. The fact that participants are seen to be shifting across different "Happy" levels are perhaps indications that the task was having the effect of "nudging" the participant in a particular way and inducing some effect on the participants' emotional state. These figures demonstrate that the participants were reacting to the tasks differently. This provides the opportunity to study and analyse the relationship between the behaviour pattern and the changes in emotional state.

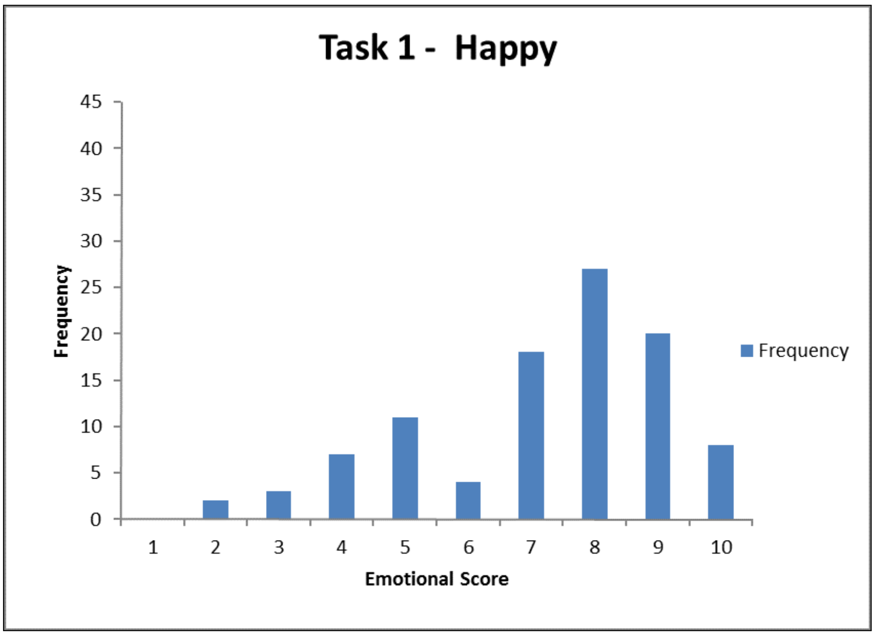

FIGURE 5.2: "Happy" emotion scores for the Task 1 - the "Maze Task"

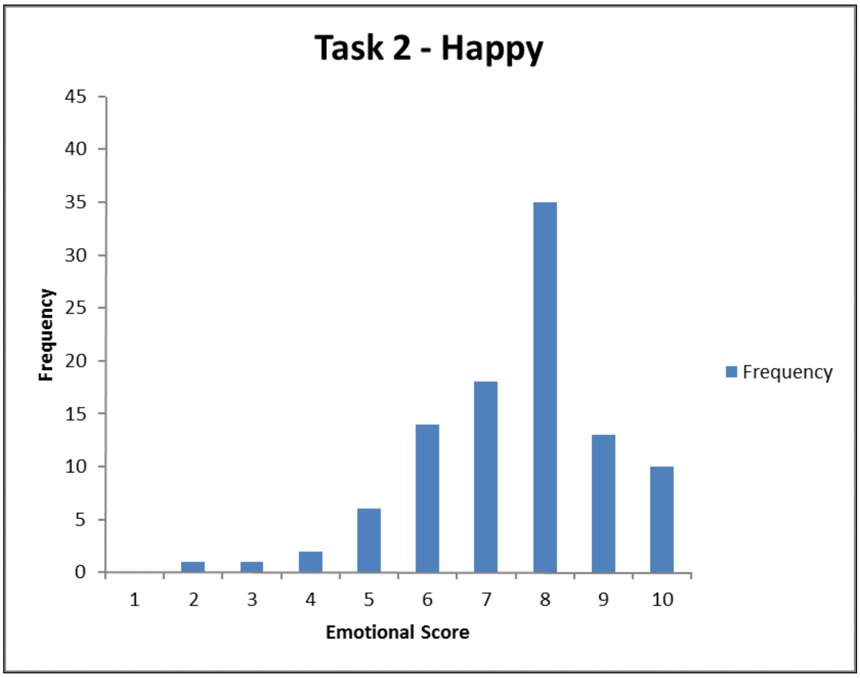

FIGURE 5.3: "Happy" emotion scores for the Task 2 - the "Video Task"

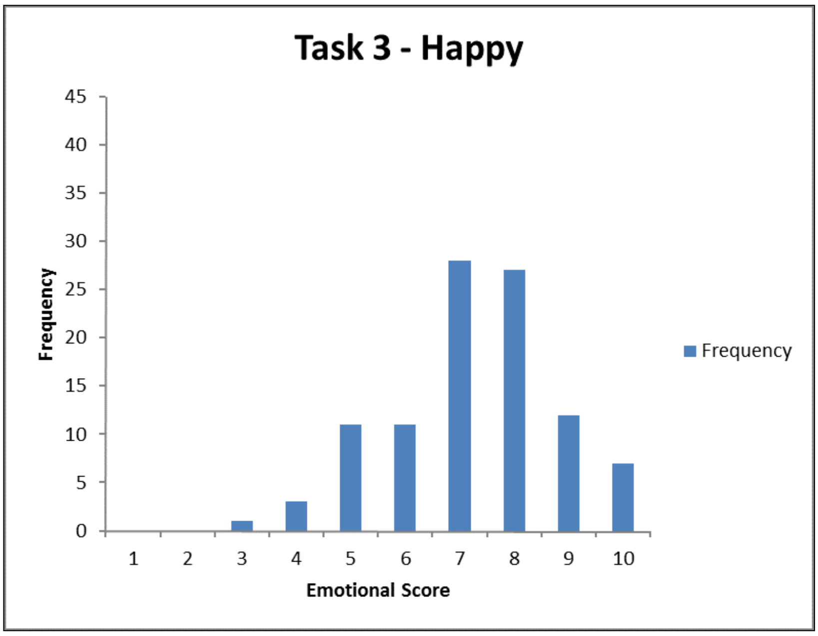

Figure 5.4: "Happy" emotion scores for the Task 3 - the "Picture Task"

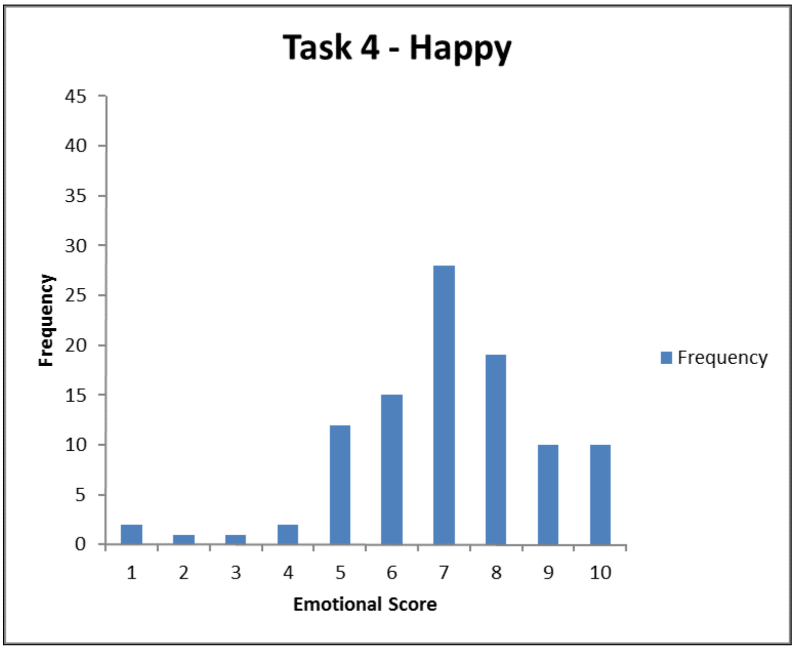

Figure 5.5: "Happy" emotion scores for the Task 4 - the "Timed Task"

- Relaxed

  Figures 5.6 - 5.9 illustrate the emotion score distribution for the "Relaxed" emotion across the four tasks. Task 1, the maze task, shows that the most populated category is eight. After Task 2, the video clip task, the population scores shift from the lower end of the scale, points 3,4 and 5, toward the higher end of the scale, points 7,8 and 9, where categories 7 and 8 are the most populated categories. In Task 3, the picture task, more participants shift from score 10, 9, 7 to 8 and 5 when compared with the previous task, where category 8 is clearly the most populated category. In Task 4, the timed task, there is a great reduction of numbers from categories 8 and 7. It appears that the lower end of the scale, points 2,3,4,5 and 6, all showed an increased number of participants. However, categories 7 and 8 are still the most populated categories.

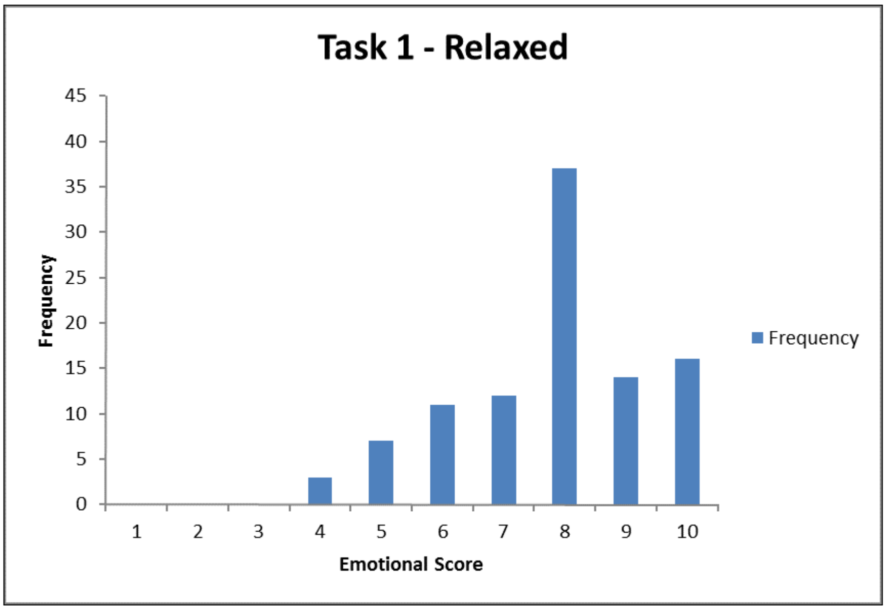

FIGURE 5.6: "Relaxed" emotion scores for the Task 1 - the "Maze Task"

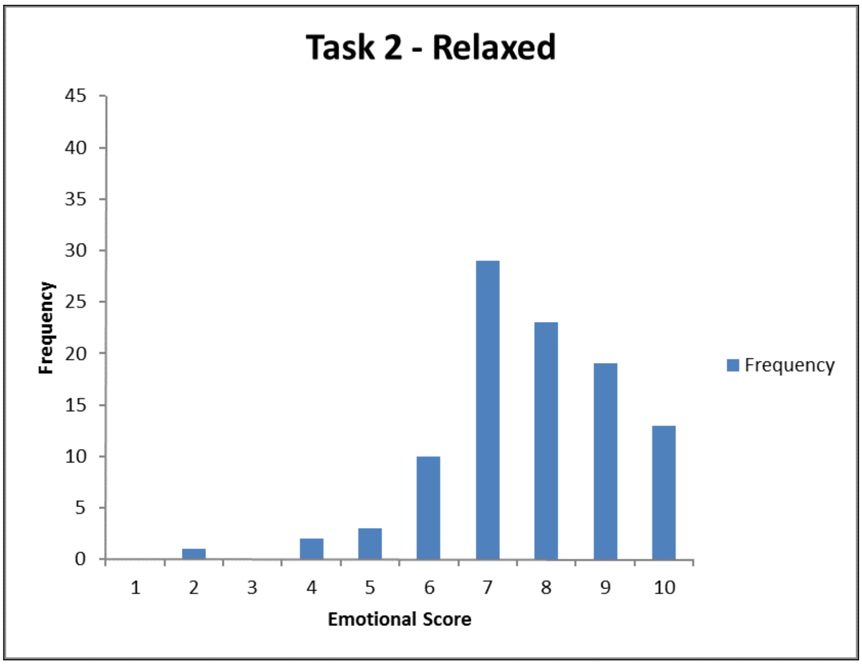

FIGURE 5.7: "Relaxed" emotion scores for the Task 2 - the "Video Task"

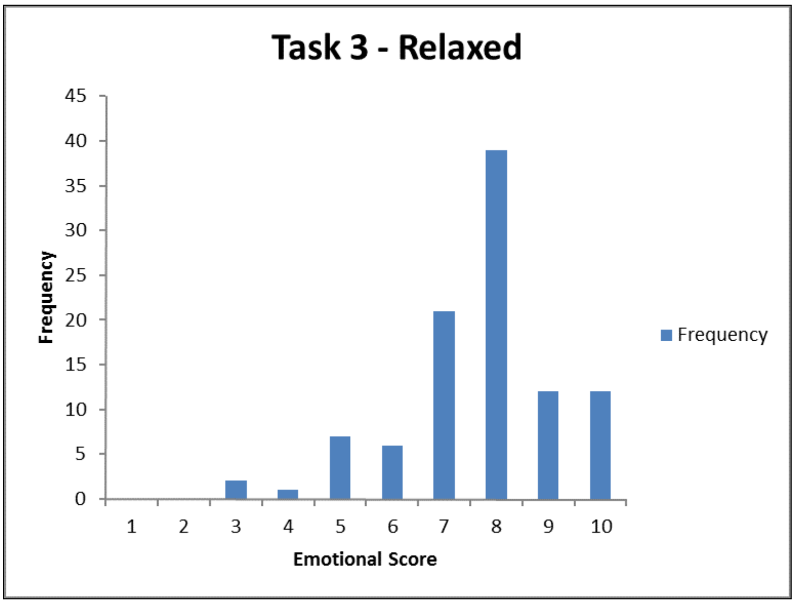

FIGURE 5.8: "Relaxed" emotion scores for the Task 3 - the "Picture Task"

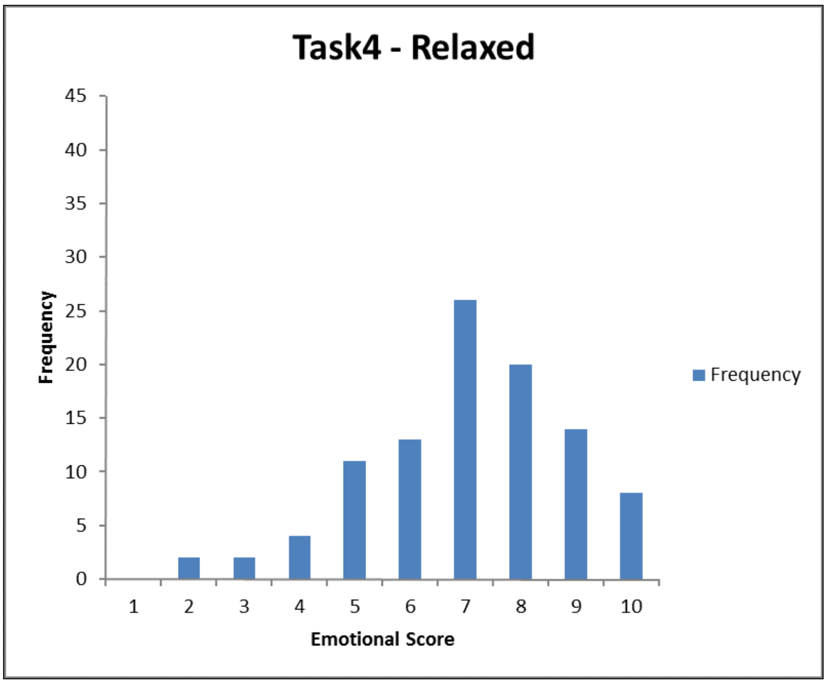

FIGURE 5.9: “Relaxed” emotion scores for the Task 4 - the “Timed Task”

The emotion scores distribution show that the peak score for both emotions are around 7 and 8 (mostly 8). Therefore, in order to achieve a clearer “separation” between the two classes, the threshold “8” is adopted for keystroke emotion ground truth labelling. As explained in Section 4.5.1, this method takes a threshold value and in order to increase the likelihood of distinguishing between the behaviour patterns, only the participants outside of the category of the chosen value are considered during the data processing phase. As discussed earlier, this method was designed to increase the likelihood of identifying keystroke behaviour patterns that represent different emotional states.

All of the publications that derived from the studies reported here used this labelling method, but it is clear that the issue of ground truth labelling is a major challenge in this area of study. Therefore, a more detailed study investigating other potential labelling methods and some discussion of these alternative methods will be presented in Chapter 6.

5.4 Emotion prediction

In this initial experiment, the main focus of the study is to investigate the predictive capability of keystroke data, which were captured during the compilation of our newly developed database. For this preliminary study, the initial process described earlier is implemented, in order to turn the pseudo-continuous emotion rating into just two classes which define either the presence or absence of the chosen emotion.

Data from two of the tasks were selected to conduct this set of experiments, Task 3 and Task 5, where Task 5 includes predefined text for the users to copy and Task 3 allows the participants to choose the content they want to provide.

In **Task 3**: As previously described, participants were shown a short video clip (cartoon) and asked to use their own words to describe the clip. Although no assumptions are made about the emotional state which this might or might not induce in subjects, the cartoon was chosen to be amusing and entertaining to watch.

In **Task 5**: Again, as previously described,participants were asked to type out a list of pre-specified words under a counting down clock, and to complete the task within a limited time span of 10 seconds. The task was designed to create an environment where the participants might feel some degree of pressure in task execution.

In this experiment, two different classification algorithms, the simple KNN (K=1) classifier (i.e. simple nearest neighbour classifier) and the Jrip classifier were used for processing the keystroke data, in order to explore the potential correlations between emotion and keystroke behaviour. The Weka data processing tool was used. Fuller descriptions of these classifiers can be found in [199, 200]. The KNN classifier with (k=1) was used with the leave-one-out methodology [206]. The leave-one-out method in this case, at each iteration, uses a sample from the subject of interest as the test sample and samples from the rest of the population are used as training samples. The first experiment was executed with the four non timing features shown in Table 5.2. The results from this experiment will provide a general baseline for the predictive process, and the prediction accuracy results achieved are shown in Table 5.3.

Feature Number	Feature Description
F1	Correction Rate
F2	Function Key Rate
F3	Key Pause Rate
F4	Input Rate

TABLE 5.2: Four common keystroke features

Task type	KNN	
	Happy	Relaxed
Task 3	53.9	46.8
Task 5	63.0	56.3

TABLE 5.3: Emotion prediction results (accuracy %) with KNN classifier

It can be observed that the results range from 46.8% to 63%. This experiment used a very small number of features and a simple classifier for the purpose of trying to establish an initial baseline performance benchmark, to provide a basis for comparison against the second experiment, which is reported in the next section.

Table 5.4 presents the performance obtained when a somewhat more powerful classifier is adopted, specifically the Jrip classifier [200]. The same leave-one-out evaluation methodology is used in this experiment. On the left of each column of predictive results achieved by the Jrip classifier, the corresponding result from KNN classification experiment is also included in Table 5.4. The comparison of results obtained by both classifiers suggests that the selection of the classifier can noticeably affect the predictive performance. In this case, the more powerful Jrip classifier is seen to generally improve the predictive accuracy. Indeed, specifically for the prediction of the "Relaxed" state based on execution of Task 5, the predictive accuracy achieved improved by 15%.

Another important aspect of the results presented in Table 5.4, is that the results obtained from Task 5 are generally better than the results generated from Task 3 when predicting "Happy" and "Relaxed". This observation is actually

Task type	Happy		Relaxed	
	KNN	Jrip	KNN	Jrip
Task 3	53.9	58.5	46.8	51.9
Task 5	63.0	63.0	56.3	71.3

Table 5.4: Emotion prediction result (accuracy%) with both classifiers (for comparison)

consistent with the previous findings reported in [17], where the use of unconstrained content (provided in Task 3) means that there is no control over the details, characteristics and amount of the data available, and this generated a performance poorer than for tasks (in this case, Task 5) where consistent content in relation to the available keystroke data is collected across all the participants.

These initial results, generated from the first set of experiments where a small feature set was used, demonstrate that there are potential opportunities for predicting emotions based on the keystroke data captured in the new database developed for this study.

In the next section, a second set of experiments will be described which include more features. The experiment will also explore the effect on predictive accuracy of using different volumes of raw keystroke data.

5.5 Further analysis

In this section, a further and extended analysis of the keystroke data is presented, where all the reported work on prediction is carried out for illustrative purposes using the Jrip classifier. In the first experiment, three different feature sets are processed with this classifier to predict the familiar "Happy" and "Relaxed" emotions. All three feature sets are taken from the overall feature set generated, and which was recorded in Table 5.1.

As can be observed in Table 5.5, the three feature sets are referenced, for ease of identification according to the number of features they contain, as follows:

Classifier	Jrip classifier					
Task type	Happy			Relaxed		
Feature set	SET _4	SET_25	SET_29	SET _4	SET_25	SET_29
Task 3	58.5	67.7	67.7	51.9	58.4	58.4
Task 5	63.0	66.7	70.4	71.3	72.5	72.5

Table 5.5: Emotion prediction (accuracy %) with different feature sets

- **SET_4**: this set contains the Correction Rate, Function Key Rate, Key Pause Rate and Input Rate features, a total of 4 features.
- **SET_25**: this set consist the mean and standard deviation for four di-graph timing features (commonly used in keystroke analysis) and eight tri-graph timing features plus the total key event, a total of 25 features.
- **SET_29**: this set consists of the combination of the **SET_4** and **SET_25** feature set, a total of 29 features.

Table 5.5 provides a comparison of the predictive accuracy achieved when these three feature sets are used for emotion prediction with the Jrip classifier. The results show that, when the feature set SET_25 is used, in comparison with the result produced when the feature set that contains four features is used, there is a clear improvement, substantially in some cases, of the emotion prediction accuracy. The SET_29 feature set contains 29 features. The feature set is produced by a combination process by concatenating both the previous feature sets. The results generated are presented in the SET_29 column in Table 5.5, and in most cases the performance achieved is still similar to the case when the "25" feature set is used, although a further 3.7% performance improvement can be seen in Task 5 when predicting the "Happy" emotion.

It can be seen from Table 5.5 that, when more features are used, the performance appears to improve until it reaches a "plateau", at which point the predictive performance stops showing any clear sign of improvement. It is also necessary to investigate whether such "threshold" can be seen in the amount of raw data that is used for generating the features, and also to analysis the features generated by different amounts of raw data to observe any possible effect on

Classifier	Jrip							
Task type	Happy				Relaxed			
Raw data volume	300	400	500	Full	300	400	500	Full
Task 3	67.2	78.7	75.4	67.7	51.9	58.9	50.0	58.4

TABLE 5.6: Emotion prediction results (accuracy %) with different amount of keystroke data

predictive performance. The "amount of raw data" here refers to the number of digraphs produced during task execution. Table 5.6 shows the predictive performance achieved when different amounts of raw data are used from Task 3. The numbers in the "Raw data volume" column represent the number of digraphs used for feature (SET_29) extraction. For example, "300" refers to the first 300 digraphs the user produces in Task 3, and "Full" describes the case when the whole collection of digraphs is used for feature extraction.

It is clearly demonstrated in Table 5.6 that when changing the volume of raw keystroke data available, the predictive performances also change significantly, although it should be noticed that for Task 3, the content provided by each participant is unconstrained, meaning that performance achieved is likely to be somewhat unpredictable. The results in Table 5.6 suggest that in any practical scenarios in which keystroke data is used in a predictive context, the amount of data used can also be a variable that potentially will have an impact on the performance analysis. In this case, an unconstrained task was investigated, but a constrained task, where the content is uniform and each user provides an equal amount of data, can potentially be more productive for investigating the effect that changes in the amount of data has on emotion prediction performance. Task 5 was not included in this experiment as it was considered that the task contained too little data for meaningful comparisons to be made with Task 3. The work reported in [78] has shown this effect more clearly when dealing with a different modality.

5.6 Conclusion

This chapter has presented an investigation of the possibility of predicting emotions from keystroke data, and has included the pre-processing phase where the features and the emotion labels are generated. For this experiment both the KNN and the Jrip classifiers are used as the data processing tool.

Section 5.4 has demonstrated the predictive accuracy results generated from our newly collected database by using keystroke data. Two of the tasks were taken as the principal point of interested for the investigation. The results show that when we take a substantially larger database with a more complete data capture procedure, the best attainable prediction accuracy is around 73%. However, these results are based on a dataset of around 100 subject, significantly larger than those used in comparable studies, and therefore more likely to represent realistic scenarios. [173] reports a predictive accuracy of around 87% but on the basis of a dataset of only 12 subjects. It should be especially noted here that the size of the newly collected database that was used for generating the results is twice the size of the biggest database previously used in reported studies.

The last section of this chapter has provided a further investigation of the predictive accuracy achievable when two of the fundamental characteristics of the features, the number of features used and the amount of raw data used to generate the feature values, are tested with different parameters. These parameters are, first, the number of features used and, second, the amount of raw data used to generate features. It appears that the predictive results show a substantial improvement when first of increasing the volume of data in terms both of number of features used, and the amount of raw data. However, the situation is not simple, and continuing to increase feature number and amount of data quickly reaches a plateau in relation to performance. The finding therefore indicates that these two parameters should be considered when generating feature sets for predicting emotions from keystroke data. It is clear that there are other possible factors, which might affect the predictive accuracy, such as the nature of the features used. As can be seen in Section 5.5, where a mix of features with different characteristics have been tested, it is clear that adding features is not the only factor which affects the predictive capability. However, this part of our reported study has provided an important foundation for future investigations to further address the impact of different feature sets on emotion prediction using biometric keystroke data.

In Chapter 4 and Chapter 5, therefore, we have presented emotion prediction results that were obtained from two different hand-based biometric modalities, handwriting and keystroke data respectively. At the same time, in Section 4.5 and Section 5.5, multiple experimental parameter characteristics have been further explored in order to observe the impacts they have on emotion prediction.

It can be seen that most of the potential issues addressed are based on further analysis of the three fundamental "building blocks" in our study. These include feature extraction, ground truth labelling, and choice of classifier. It is apparent that there are limitations on what performance can be achieved and, at the same time, opportunities to extend the boundaries of the three fundamental factors of this area of study. A set of additional experiments, that tested and explored the boundaries of those factors, will be presented in the next chapter.

CHAPTER 6

Other issues in predictive biometrics

The previous chapters have reported the core message of an extensive study of predicting higher-level soft-biometrics information. The results presented in Chapter 4 and Chapter 5 have shown some encouraging indicators of the capability of predicting higher-level mental states from biometric information based on handwriting and keystroke data. At the same time, the earlier chapters have presented and discussed some of the important challenges and opportunities for developing an effective implementation framework and an experimental approach to establishing an appropriate procedure to investigate this. The previous chapters also have drawn out three principal key issues, which significantly influence the performance which can be achieved in relation to predicting these higher-level mental states/emotions.

Firstly, as demonstrated in Section 4.4 and Section 5.4, the variations within different possible feature sets utilised can influence the predictive performance. As reported in Chapter 1, the feature sets that most of the reported emotion prediction studies use are seen to be fairly standardised and traditional.

Secondly, it can also be seen, in Section 4.4 and Section 5.4, that the selection of the classifier which is to be used for the predictive process itself can have a clear impact on the predictive results.

Thirdly, one of the elusive and difficult challenges that almost any emotion prediction study faces is the "Ground Truth Labelling" task. In Section 4.5.1 presented a series of experimental studies exploring the idea of imposing different thresholds by means of which to interpret the emotion scores that were collected from the participants, and to observe the impact which the corresponding representations have on the predictive accuracy.

It is readily apparent that emotion prediction performance can be significantly impacted by any of the three fundamental building blocks specified above and, in order to achieve some further insights into the nature and limits of the techniques we have been developing, it is imperative to take a closer look at these key issues.

However, in addition to the core material presented previously, the study has also generated other, less immediately central work, and in particular, some further development of the three issues noted above has arisen. This chapter provides an opportunity to address some of these additional ideas and attempt to address some of these less central findings so far discovered. In an important sense, this chapter therefore can be seen as a bridge between the rigorous formal studies so far reported, and some areas on which future work might most usefully be focused. However, while the previous (core) material has been fully worked through and discussed in detail, the nature of this chapter is therefore such that a number of further developments the of issues identified above are reported rather more briefly, and without the degree of rigour with which earlier experimental analysis was carried out. This part of the study nevertheless sheds some light on very relevant issues, which are related to the core theme of the overall study. This chapter should consequently be seen as enhancing the main body of work reported, providing extra information about some of the additional issues which have arisen during the course of the study, and pointing out some of the possibilities for carrying out some important new developments in the future.

Section 6.1 briefly investigates some possibilities for generating some new feature sets, which provide a new dimension for the representation of the behaviour patterns of interest. In Section 6.2, a more powerful and flexible novel classifier, a multi-agent-based classifier which exploits the power and flexibility of intelligent agent technology, will be investigated as an alternative paradigm on which to base the processing of the feature set. In Section 6.3, two further alternative labelling methods (specifically, different ways of handling the transformation from the emotion continuum as captured by the Likert scale to the two-class emotional labelling) will be presented, discussed and assessed. It will be seen later that this is a particularly important area for further work in the future. Section 6.4 will make some concluding observations.

6.1 Newly developed keystroke feature sets

Most conventional keystroke features commonly used currently were developed more than two decades ago. The majority of studies in this area of research have adopted the same features, including the small number of studies aimed at predicting emotions from keystroke data, which consist of standard features, as reported in Chapter 1. These feature sets typically include key press and key release timing features and some studies also include variation of features such as, features extracted from the content of the data (i.e. verb rate), features extracted from the use of the keyboard (i.e. backspace rate) and so on. There are a number of studies that try to address this issue by modifying the keyboard to include additional features of a pressure sequence [150], features such as mean and standard deviation of the pressure reading for one key press. It is also helpful to consider an alternative to introducing special-purpose hardware and instead to examine possibilities to focus on developing new features, which are viable alternative methods without any modification, required on the keyboard. Most keyboards that are currently used by the general public in ordinary use are not modified, and therefore to ensure that this study can be seen as relevant to the widest number of possible applications, we have explored further the basic features to seek for any further useful developments.

A traditional set of keystroke features normally requires the computation of the mean and standard deviation of the timing characteristics (as previously noted) as indicators of how the participants are behaving on the keyboard. However, this method only represents some of the characteristics of the participant's typing pattern. In order to be more specific when we are looking for the pattern changes under different emotional conditions experienced by the participants, a closer look at these features is required. This means targeting a group of commonly occurring digraphs across the entire population within the database.

Instead of using every key and corresponding timing data that were collected, only the most frequently occurring digraphs were selected, where those digraphs each appeared in the typing of every participant throughout all tasks. This provides a more uniform template to compare between each participant. It should be noted that the most frequently occurring digraphs are database dependent, as this information could vary depending on the data collected from the individual participants. For example, for our study, the top four most frequently used digraphs are: the transitions between the "Space" and "A" keys, the "Space" and "T" keys, the "A" and "N" keys, and the "E" and "Space" keys. For each task, the four timing features for each occurrence of those frequently used digraphs are collected and a mean value measured.

The second method for generating new features that was developed utilized the physical layout of the keyboard. For this method, the aim is to determine the distance between each key. Therefore, by incorporating the distance between each key with the timing features (first key press to the second key press), this new feature reflects the "speed" of the typist's finger travelling from a specific key to another specific key.

Even this method did not require any modification of the keyboard, as mentioned earlier. The procedure by which the measurement of the required distances was determined is shown in Figure 6.1 and Figure 6.1.

Figure 6.1 shows the reverse side of the keyboard that was used in the data collection procedure, showing the white circles painted to the base to enable accurate imaging of the overall layout. Then the highlighted base of the keyboard was electronically scanned. The scanned image was processed to find the edges of each key. The centroids of highlighted circles were calculated, and then the distance between each key determined by measuring the distance between these centroids. The distances were stored by using the distance between the key "A" and "D" as a unit distance, and the rest of the distances are then a ratio representation based on the unit distance.

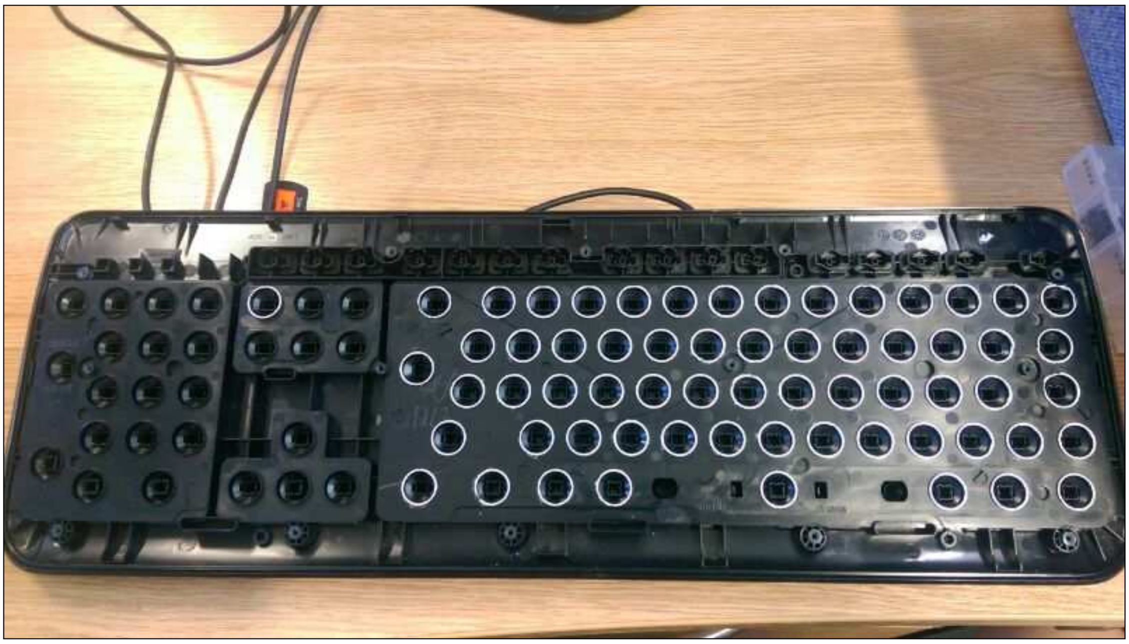

FIGURE 6.1: The disassembled keyboard with white paint highlighted keys

Task 3	Conventional 29 features feature set	New 20 features feature set
KNN	63.1	62.3

Table 6.1: "Happy" emotion prediction (%) result with both conventional feature set and newly developed feature set

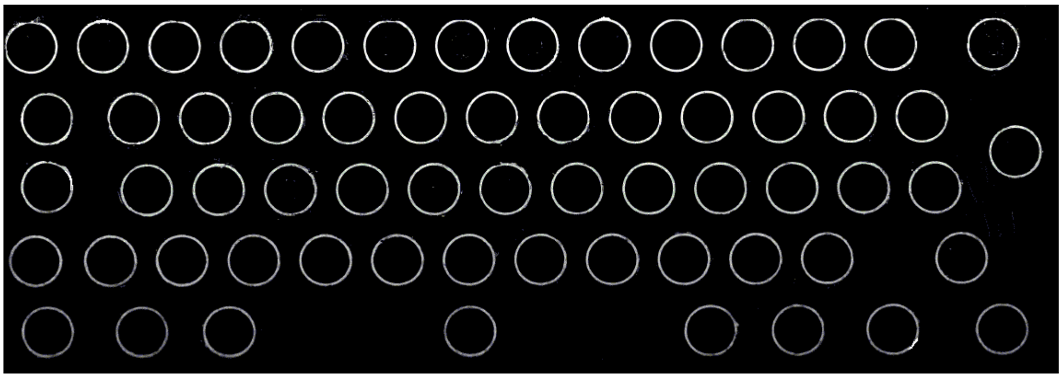

Figure 6.2: Scanned and filtered key layout

The newly developed feature set, which consist of the selected top digraphs' timing features and the "Speed" feature are tested with the KNN (K =1) classifier. The results are compared with the results generated from using the more conventional features set out in Table 6.1, where features based on Task 3 and using the "Happy" emotion prediction were used as an illustrative example.

As the results presented in 6.1 show, the newly developed feature set has not exhibited a significant improvement from the results that were generated from the conventional feature set. However, this feature set has achieved the same level of predictive accuracy using only a part of the raw data for feature generation. Therefore, computationally, it might be argued that this is more efficient in cases where a significantly larger dataset is required to be processed.

6.2 An intelligent multi-agent classifier algorithm

Previous research studies within the author's Research Group have provided evidence that an intelligent multi-agent classifier paradigm can help improve the identification accuracy in a range of biometrics-related applications [182, 183] including in soft-biometrics studies [17]. This section will take the newly generated set of features and explore the effectiveness of the combination of the new features and these more intelligent and flexible novel multi-agent classifiers which was previously developed within University of Kent School of Engineering and Digital Arts for other biometrics data processing. [1]

Here we simply summarise the findings of the experiments, which have been published in [207]. For this experiment, we selected 20 newly developed keystroke features from Task 3 as an illustrative example, the feature set including the top four most commonly used digraphs' four timing features and "Speed", as described in Section 6.1. The emotion adopted for the experiment for the prediction task was to predict whether a participant is "Happy" or "Not happy".

6.2.1 Base classifiers for the multi-classifier system

Multi-classifier systems (MCSs) aim to combine a set of base classifiers in order to achieve a better predictive accuracy than is possible using the individual base classifiers alone [208]. The advantage of integrating multiple classifiers is allow the MCS to consider different techniques to solve the classification problem in a cooperative way, where different classifiers might be more capable of solving different classification problems. The base classifiers are combined with different strategies. The strategy combines where each classifier's outputs information and compute an overall decision for the system. The final predictive result are computed by decision-making methods which can also adopt different strategies or algorithms [209].

In order to adopt a multi-classifier system, the base classifiers first need to be defined. For the proposed system, the following three base classifiers were chosen, providing a diversified range of classification algorithms:

- Multilayer perceptron (MLP) [210]: This algorithm is a Perceptron neural network with multiple layers [211]. The output layer receives stimuli from the intermediate layer and generates a classification output. The intermediate layer extracts the features, their weights being a codification of the features presented in the input samples, and the intermediate layer allows

[1] The author wishes to acknowledge the work of Dr Marjory Abreu in developing the classifiers referred to here, and to thank her for permission to use the software she developed.

Classifier	Accuracy
MLP	61.0
RBF	66.1
SVM	62.7

Table 6.2: Accuracy (%) of the individual classifiers performing the emotion prediction task using the keystroke database

the network to build its own representation of the problem. Here, the MLP is trained using the standard back-propagation algorithm to determine the weight values.

- Radial basis function classifier (RBF) [212]: This algorithm adopts an activation function with radial basis, and can be seen as a feed forward network with three layers. The input layer uses sensory units connecting the network with its environment. The second layer executes a non-linear transformation from the input space through the output space performing the radial basis function.

- Support vector machine (SVM) [201]: This approach embodies a functionality very different from that of more traditional classification methods and, rather than aiming to minimise the empirical risk, aims to minimise the structural risk. In other words, the SVM tries to increase the performance when trained with known data based on the probability of a wrong classification of a new sample. It is based on an induction method, which minimises the upper limit of the generalisation error related to uniform convergence, dividing the problem space using hyperplanes or surfaces, splitting the training samples into positive and negative groups and selecting the surface, which keeps more samples.

Before the testing of the proposed multi-classifier system, the three base classifiers are applied individually to the newly developed keystroke features to determine the each achieves alone. The emotion prediction performance achieved with each classifier, with the parameters settings, is presented in the Table 6.2.

A multi-classifier system can adopt different combination strategies, such as sum, and vote, each representing different methods of integrating the prediction decisions, which each base classifier makes. The sum method operates directly on the soft outputs of individual classifier for each prediction probabilities. The integrated decision is obtained by applying the maximum value selector to the class dependent averages of these outputs. The vote method operates on class labels assigned to each subject by the respective classifier. The labels are obtained by evaluating the decision outputs using the maximum value selector. The vote method output is a function of the votes received for each class from each base classifier. These approaches are described in greater detail in [213]. In order to generate a benchmarking performance, for this investigation a conventional "sum" method of combining the classifiers was tested and evaluated and this method generated a prediction accuracy of 68.10%. As would be expected, this is an improvement in comparison to any of the individual classifiers' performances.

The next section will present the proposed new agent-based system and will present the results achieved with this system.

6.2.2 An intelligent multi-agent classifier algorithm

The multi-agent classifier was developed to improve the flexibility of the decision making process of a multi-classifier system, where base classifiers are systematically organised with an agent-based architecture. Each agent in the multi-agent system is able to carry out the classification task and make its class prediction. Agents can communicate and work in cooperation to reach the same goal, the interations follow their negotiation protocols. Each agent has four mian modules, which include "*Controller module*", "*Decision-making module*", "*Negotiation module*" and "*Classifier module*". Therefore, when a new biometric sample is passed into the multi-agent system, all the agents perform the same prediction on the sample. The *Controller module* passes the required information to the *Decision-making module* and the *Decision-making module* then accesses the *Classifier module* to produce classificationr results and the degree of confidence measure. Then, instead of using the output integration methods (such as sum or vote) the agents negotiate with each other in order to reach an agreement on the prediction of the input sample, as is explained in detail in [182, 207]. The negotiation process can be embedded within different configurations and three different configurations were implemented for this study, as follows:

- Sensitivity (Agent-S) negotiation uses a decrease in the **confidence level** of the classifiers (which is part of the agent) and is considered with a sensitivity analysis during the testing phase. This analysis can be achieved excluding and/or varying the values of an input feature and analysing the variation in the performance of the classifier method. The main aim of this analysis is to investigate the sensitivity of a classifier to a certain feature and to use this information to guide the negotiation process. This analysis is performed with respect to all features of the input patterns in the emotion prediction classifiers.

- Game theory (Agent-G) uses the systematic description of the results, which can be carried out through the use of strategic games. A strategic game is a game in which a player chooses a plan of action only once and at the same time as his opponent. The agents will only have two options, which are **keep** and **change**, where **keep** means to keep the result of their own prediction and **change** mean to agree to the other agent's prediction result. In order to help the players to make their decisions, a payoff matrix is used, in which each cell represents the payoff values which the players will have in a situation where these actions are chosen. The cell with the highest value is chosen.

- Auction (Agent-A) uses a **cost** for each agent which is based on the sum of the differences between the winner confidence and the confidences related to the chosen classes of the other agents. In this technique, all agents are considered as buyers, trying to reach an agreement in relation to an input pattern.

The results generated using the multi-agent configurations are presented in Table 6.3.

As the results demonstrate, in comparison to the other classifiers's predictive performance reported in Section 6.2.1, by using the proposed multi-agent system with the strategy (Agent-S), a 7% improvement on the previously best performance achieved by the "sum" method is produced, and at the same time, 13% improvement when the same feature set was used with a more traditional classifier as reported in Section 6.1.

Thus, it can be seen that when the more intelligent and flexible multi-agent system is adopted, not only it can improve performance in more traditional

Strategies setting for the multi-agent system	Predictive accuracy
Agent-G	71.8
Agent-A	71.0
Agent-S	75.3

Table 6.3: “Happy” emotion prediction (%) with multi-agent system using the keystroke database

biometrics data processing such as identity prediction [182], but can also benefit emotion prediction based on the analysis of biometrics-based data.

6.3 Newly developed ground truth labelling

The main approach adopted previously to interpret the emotion scores is to find the best threshold to separate the population into two classes, as presented in Chapter 4 and Chapter 5 for emotion prediction from both the handwriting data and the keystroke data.

For this method, the labelling is based on the threshold value. For example, if the threshold is set at “8” for the “Happy” emotion and Participant A scored him/herself with a value of “6” for Task 1, Participant A would then be labelled as feeling “Not happy” for Task 1 because the score is less than the threshold of “8”. Similarly, in a situation where Participant A self-scores with a value of “9”, this participant would fall into the category of being considered to be in an “Happy” state.

A principal reason for adopting his method is that it is significantly more objective than the method adopted in the study reported in [150], as we have indicated earlier. However, we have also considered a rather different approach which can reflect the emotion experienced by a participant from a different perspective. It was recognized during the data capture process, when capturing the emotional scores, that participants mostly appear to determine their current emotion score based on the score they assigned in relation to the previous task undertaken. After observing this pattern, the second approach was therefore developed. In fact, this new approach developed into two sub-methods that both take more account of the change in the participant's emotion scores, in order to determine their emotional label.

6.3.1 Dynamic Labelling Method

For this method we introduce an index which we designate the "Emotion Changing Index", ΔEmo(E). Equation 6.1 demonstrates the method of calculation of the index value.

$$\int_{2}^{a} \Delta Emo(E) = \varepsilon(a) - \varepsilon(a-1) \tag{6.1}$$

In this equation, ΔEmo(E) is the emotional score "ε" for emotion "E" at task "a" and then subtract the emotional score "ε" for emotion "E" of the previous task "a-1". The limit of the integral are between "2" and the number of the task "a". This method require the emotional score from the task "a" and the previous task. Thus the value of "a" starts at "2".

After each task, the participants were asked to assess their own emotional status. A pattern became obvious while the participants were filling in the questionnaire provided, which is that all of the participants tend to refer back to the previous scores they assigned when determining their current states. For example, Participant A might rate himself/herself "8" for the "Happy" scale in the case of he/she felt happier after Task 2 (assuming their previous score for "Happy" emotion was, say, "6" for Task 1). Participant B might rate himself/herself "10" for the Task 1, and in Task 2 the score drops to, say, "9" if he/she is less happy. In this case, as a result of the Threshold labelling method, Participant A would be excluded from the "Happy" class for Task 2, but was clearly feeling happier than Task 1. Participant B would be labelled as "Happy" for Task 2, although also showing clear signs of being less "Happy" than they were during Task 1.

In order to compensate for this potential mislabelling, we have developed what may be referred to as the "Dynamic Labelling Method", where we only take the positive or negative change in emotional score into account, rather than dealing with absolute values. This means that if ΔEmo(E) > "0" then the label is positive, while the label is negative when this change is negative. For example, if Participant A scored "3" in the first task for "Happy" and then moves to "7" after the next task, Participant A would be labelled as "Happy" for the second task. This method takes into consideration the fact that the scale might mean something subjectively differently to each participant. Participant A scoring "5" for Task 3 might be the same as Participant B scoring "8" for Task 3 in the context of the what they are experiencing emotionally. The ΔEmo (E) measure is more reflective of the change of their emotional experience.

6.3.2 Mixed Labelling Method

After consider both the labelling methods discussed in the previous sections, it is apparent that both methods potentially have their own particular advantages and disadvantages.

The Threshold Labelling Method widens the gap between the two classes of participants, which creates more separation of the behaviour patterns of participants while comparing the two contrasting classes. The principal downside of this method is that a good percentage of participants might be excluded from the data processing phase in order to create that gap between two classes. It can be noted selecting "8" as the threshold around which to base the separation of the two classes, this process has settled on one of the most populated score categories, as has been demonstrated in both Chapter 4 and Chapter 5 in the section when emotional score distributions were examined in detail.

The Dynamic Labelling Method, on the other hand, includes more users into the data processing phase, although it also introduces more "noise" into the data, as certain participants will be assigned to the opposite class when they are slightly less happy but still "Happy", or slightly happier but still "Not happy", Thereby creating a somewhat more complicated classification challenge.

Therefore, a "Mixed Labelling Method" has been developed after examining the previous two methods. The aim of this further new method is to balance the trade-off between eliminating too many participants and including too many participants as the previous two methods do. The "Mixed Labelling Method" uses the principle of the Dynamic Labelling Method to execute the first round of labelling with the ΔEmo(E) $>$ "1" or ΔEmo(E) $<$ "-1". This effectively means that we first look for highly "emotionally stimulated towards happy" or "emotionally stimulated towards not happy" participants. For example, if the

		29 common features		
Relaxed	Classifier	Threshold label method	Dynamic label method	Mixed label method
Task 3	KNN	57.4	83.3	81.2
Task 4	KNN	66.0	60.0	67.5
Task 5	KNN	78.4	55.0	63.2

Table 6.4: Predictive results (%) when three different labelling methods

ΔEmo(E) for Participant A is "+3" where "E" is the "Happy" emotion, then Participant A will definitely be labelled as "Happy" regardless of where the scores lie on the $1 - 10$ scale, and vice versa. For the rest of the participants who fall into the range of "-1" ΔEmo(E) "1", we then apply the "Threshold Labelling Method" with the threshold index set at "8".

This method is able to exclude fewer participants than either of the previous methods, which provides an alternative but perhaps fairer way of establishing ground truth labelling. In order to observe the impact of the newly developed labelling methods, keystroke data from Task 3, Task 4 and Task 5 are selected to predict whether a participant is "Relaxed" or "Not relaxed". Table 6.4 provides the predictive results with the different labelling methods and traditional feature set, using a simple KNN (K =1) classifier to execute the predictive process.

These results show that it is perhaps difficult to draw any really definitive conclusions from these results. The results suggest, however, that there may be potential benefits to the use of alternative methods for ground truth labelling, for example in the case of Task 3, where the predictive performance improves with both new alternative methods.

However, it is clear that there is more work, which needs to be done in the future to explore the possibility of a better or more accurate ground labelling method for emotion prediction studies. The study presented in this section does no more than set out some basic ideas for further developments.

6.4 Conclusion

This chapter has identified three of the fundamental principal key issues that can be commonly encountered with any behavioural biometrics.

We have presented a set of newly developed features. These extend the underpinning ideas of the more traditional features by selecting only the common digraphs that most participants used and only building the template around those digraphs. This method uses a small section of the data captured (i.e. he four most frequently occurring digraphs in our newly collected database), which we believe provides a greater focus on the typing behaviour template than a traditional feature set.

Another method of generating new features was also developed by measuring the distance between each key and using this additional information to calculate the speed of participants' finger movement from one key to the subsequent key when typing certain combinations of keys. These new features have been tested with one of the classifiers that was used in the previous experiments. As demonstrated in this chapter, this feature set has achieved the same level of predictive accuracy. However, computationally, it might be argued that this is more efficient, as this feature set only used a part of the raw data for feature generation. Therefore, the new feature set can be effective when a significantly larger dataset is required to be processed, or when raw data availability is limited.

The newly generated features also have been tested with a more powerful classifier structure, a multi-agent based classifier system. This allow a comparison to be made with those previously considered data processing infrastructures. The predictive performances achieved show promising signs of improvement when different combination strategies are applied. The proposed system shows a significant improvement over the best result achievable using the base classifiers alone and a conventional multi-classifier system using the conventional standard typical combination strategy, as reported in Section 6.2.1. It is clear that by adopting a multi-agent system, better predictive performance can be achieved. Also by optimizing the settings, further performance gains are also possible.

By comparing the emotion prediction results produced in [79], when the traditional feature set is used, with the predictive accuracy reported in Section 6.1 when the new feature set is used and Section 6.2 when a more conventional multi-classifier system strategy is used, the newly developed multi-agent system and features still achieved the best predictive accuracy of 75.3%. This is consistent with the studies reported in [182, 183]. The newly developed features demonstrated a similar level or predictive capability when compared with the traditional feature set in Section 6.1. However, the newly developed features require less raw data to be used when generating features, which can prove to be computationally more viable or can be effective when raw data availability is limited. Both the newly adopted data processing infrastructure considered here and the feature set show some potential benefit and lay down the foundations for further developments.

The last section of this chapter investigated the “Ground Truth Labelling” phase of the analysis, which is one of the most challenging tasks in the predictive aspects of behavioural biometrics, especially for emotion predictions. Two new methods have been presented and tested using the traditional keystroke feature set. As can be observed, even with the same basic emotion scores, the assigned labels can be very differently interpreted. The different labelling methods considered can have a significant impact on the predictive performance, and this emphasises the importance of providing a fair and meaningful ground truth labelling in this area of study. The methods developed in this study can provide a starting point for considering in more detail, and with greater rigour, a better ground truth labelling methods, although it is clear that substantially more work needs to be undertaken in this area.

The three principal issues that are investigated in this chapter can all be seen to have an impact on the predictive performance attainable. However, different configurations can either increase or decrease the predictive accuracy, but the experimental results and analysis presented give some initial indicators about how performance may eventually be optimised in specific practical applications.

Some of the potential areas that can benefit from this type of study will be considered briefly in the next chapter, along with a consideration of possible future directions for this work and some overall conclusions about the study which has been reported in this thesis.

CHAPTER 7

Conclusions and suggestions for further work

This Chapter will provide a brief review of the studies reported in this thesis, summarise the progress that has been made, further discuss the contributions made and identify further challenges and opportunities for future research which arise from this study. Section 7.1 will discuss the main issues that have been identified and present a summary of how these issues have been addressed, and will provide an overview of the contributions made to the field of research. Section 7.2 will present some observations about both immediate and long-term future work which this study has highlighted. This section also will present some supplementary developments that have been made to support further investigations. Section 7.3 will conclude this chapter and the thesis.

7.1 Summary of contributions

This thesis has reported a comprehensive study on extending the predictive capability of hand-oriented behavioural biometrics and, to be specific, behavioural data collected from typing and writing activity. The thesis has presented an analysis of the issues relating to the acquisition of biometric data from both modalities, and so-called “higher-level” mental state prediction based on the data collected. Presented in this study is a robust data collection principle designed to collect data when participants are exposed to different experimental environments, a systematic study of the predictive performance of the data acquired, and the implementation of detailed experiments to investigate valuable issues which can enhance the robustness of this higher-level state prediction technique.

A detailed review of the state of art of higher-level soft-biometrics information prediction studies has been reported in Chapter 1. This review presented a detailed overview of the available literature that has focused on the prediction of both lower-level soft-biometrics information and higher-level soft-biometrics information. It is apparent that lower-level soft-biometrics information prediction has been more extensively studied than the prediction of higher-level soft-biometrics information. It can be seen that lower-level information prediction has already shown promising success in the field of theoretical research and the transformation to important practical applications. The review has focused particularly on the less extensively studied higher-level information prediction, which is inspired by the successes of lower-level information prediction, and has considered some initial studies which have been reported in the literature. The initial studies on different modalities collectively suggest that higher-level mental states can, indeed, be reflected in the behavioural patterns exhibited by individuals. As presented in the review, facial expression is one of the most developed modalities for higher-level mental state prediction, but the review also presented studies relating to other, behavioural, modalities, such as handwriting, keystroke dynamics and voice. The literature suggested that handwriting and keystroke are especially lacking more extensive investigation with regard to their predictive capability. Therefore, this thesis has focused on extending and analysing the predictive capability of these two hand-oriented behavioural biometrics. However, these two particular modalities are also interesting because they represent the most commonly used interactive operations by means of which individuals interact with machine-based systems and, indirectly, therefore, with each other and with external services.

The review highlighted one of the main significant weakness of the studies based on these two modalities, which is that the size of the databases typically adopted are generally too small, which makes it very difficult to draw any firm conclusions from the currently available studies. The review also presented four areas that have not been thoroughly addressed in previous studies, which include considering the size of database, ground truth determination, task definition and processing infrastructure. where the non-standardised approaches can generate some fundamental problems. The literature review has identified these issues and provided useful insights into the techniques and methods developed for higher-level mental state prediction. This provide an opportunity for the study carried out in this thesis to be more targeted in the development of the methods adopted.

The literature review and the discussion of the issues in Chapter 1 laid down the foundation of the development of the core principal objectives of this thesis, and has provided an initial starting point for the development of a more efficient and effective methodology for investigating higher level soft-biometrics information prediction with hand-oriented behavioural biometrics. The discussion of the previously reported databases in Chapter 1 has provided the initial structure for the design of a new proposed data collection protocol and experimental infrastructure, which have been described in Chapter 2.

Chapter 2 has covered the data capture procedure for both modalities. This chapter also presented, in detail, the logic and reasoning behind the design of the protocol and the implementation of the protocol in the experimental procedures.

It presented the software and hardware set up for the data acquisition system, and demonstrated how the set up incorporated the designed protocol into the system implementation, thus creating an intuitive environment for experimental subjects to provide their behavioural data from handwriting and typing activities.

The database is an important fundamental building block for the further analysis of the predictive capability of both the behavioural soft-biometrics modalities targeted, and represented a significant effort in its realisation, which is one of the major contributions of this study. The database contains keystroke and handwriting data acquired from 100 participants, which is double the size of the largest keystroke database previously reported in the literature. To present the database with greater clarity, a visual representation of the distribution of the overall demographic data of the entire population is also shown in this chapter. An initial analysis of the distributions has highlighted some indicators that can help to draw conclusions from the actual data analysis and mental states prediction.

Before the data collected are used for higher-level mental state prediction, Chapter 3 presented an initial context linking the present work to previous related studies by focusing on predicting a conventional soft-biometrics characteristic (gender prediction in this case) using the handwriting data acquired in our newly developed database. This study filled a notable gap in the available literature of handwriting gender prediction by using data collected by a digitising tablet. The predictive result reported around 13% improvement in comparison with the experiment result reported in [77], which used a similar data processing algorithm(SVM).

This chapter reported a further investigation of two important issues related to the use of handwriting for gender prediction. The two issues are, first, exploring the potential impact of different types of handwriting feature set and, second, the content of the handwriting data supplied, and the effect which these factors have on the gender predictive performances. The chapter reported two experiments implemented to explore the potential impact. In the first experiment, features are divided into three feature sets (static feature set and dynamic feature set and the combination of both feature sets). These feature sets are processed separately and the gender prediction results suggest that in gender prediction, the more important factor that determine the predictive performance is the content of the task. An important message, which emerges from this chapter, is that when comparing gender prediction results, it is necessary to consider influencing factors such as the two points introduced here.

The gender prediction experiments presented in this chapter provided some valuable experience prior to the exploration of the main focus this thesis, which is to investigate the predictive capability of handwriting biometrics and keystroke dynamics in the prediction of higher-level states. Chapter 4 therefore reported, in detail, on experiments on emotional state prediction from handwriting data.

This chapter described the analysis of the handwriting data obtained for emotion prediction, which is one of the core aspects of analysis in this thesis. Chapter 4 also presented the steps taken for pre-processing of the raw data and a description of the feature sets used for the emotion prediction. The feature extraction was carried out by generating a basic set of features which are commonly used for handwriting data collected by means of a digitising tablet, while recognising that most previous studies have been developed for identification tasks, which make the study reported the only study directly addressing higher-level mental state prediction from handwriting data that are captured by this means. This is another of the core contributions of this thesis to this general area of research.

This chapter also, presented a detailed discussion of the ground truth labelling methodology adopted, which is a very important and potentially difficult areas and, indeed, is one of the most challenging tasks in this type of work, as discussed in Chapter 1. Chapter 4 presented the approach adopted in this study and how emotion-indicative scores could be assigned to the data that were captured. The ground truth labelling method adopted for this study avoided the more problematic method where the labels were automatically assumed based on the task content, rather than the actual responses obtained from the subjects themselves. The method of using a 10-point Likert scale was adopted to determine an indication of the actual emotions being experienced by subjects, consistent with similar Likert scale approaches which can also be found in other emotion prediction studies presented in Chapter 1. A list of emotion score distribution charts were presented to provide a visual representation of the allocation of the emotional scores, and a notion of optimum parameter settings for ground truth labelling was introduced and developed. In order to examine the chosen parameters, an investigation of the influences that different parameters of the labelling method can induce with respect to the predictive performance attainable were also presented and analysed in this chapter.

Chapter 4 reported the emotion prediction results obtained from three different classification infrastructures across the four chosen tasks. A set of initial results achieving almost 80% predictive accuracy for the "Happy" emotion prediction and almost 75% predictive accuracy for the "Relaxed" emotion prediction was obtained. These preliminary experiments therefore showed very encouraging prediction performance, which fill a gap in the current literature of using handwriting data collected from digitising tablet to predict emotional state. The results provide evidence and confirm the value of the core focus of this thesis, which is to extend the predictive capability of hand-oriented soft-biometrics. The study also raised a number of important insightful issues and challenges, which have also been investigated.

Chapter 4 also addressed three of the main areas that could potentially influence the predictive performance. These three areas include the ground truth labelling threshold value settings, different categories of feature sets and different categories of tasks. The work carried out concluded that, in order to further develop an optimised prediction strategy, it is essential to take these issues into consideration. These results also demonstrated the performance bounds that were found when exploring the different options, which are likely to arise in the practical adoption of these techniques in different functional scenarios. Chapter 4 briefly described some real life scenarios to exemplify the practical applicability of emotion prediction by using handwriting data. The second main focus of this thesis is to explore the predictive capability of keystroke data, results for which were presented in detail as the main topic of Chapter 5.

Chapter 5 therefore continued to explore the predictive capability of hand-oriented biometrics by focusing on the analysis of the keystroke dynamics modality. This chapter reported the list of features generated from the keystroke data collected in the acquisition of the newly developed database. For this set of experiments, the same ground truth labelling method was carried over from the handwriting data processing described in Chapter 4, where emotional score distribution charts were listed and analysed, in order to find the optimum setting to maximize the likelihood of separating the population into two clearly defined groups representing the different emotional states of interest.

The predictive accuracy results using keystroke data were reported in this chapter, where two of the tasks were taken as the principal point of interest for the investigation. The best attainable prediction accuracy was seen to be around

73% and these results are based on a dataset that is the double the size of the largest database previously reported, and therefore more likely to represent realistic scenarios. The findings of this chapter also provide evidence that with a significantly larger and a more complete database, higher-level emotional states can still be predicted with keystroke data. This again confirms the value of the core purpose which this thesis attempting to achieve, which is to find out to what extent hand-oriented biometrics can be used for higher-level mental state prediction.

This chapter has provided a further investigation of the predictive accuracy achievable when two of the fundamental characteristics of the keystroke features that have not yet been addressed in the literature, the number of features used and the amount of raw data used to generate the feature values, are tested with different parameters. It is apparent that there are a number of factors such as these which can have an impact on the predictive accuracy attainable. This chapter has provided an important foundation to further address the impact of different feature sets on emotion prediction using biometric keystroke data. It can be seen that we have made considerable progress from conventional approaches to predicting gender to predicting higher-level emotional state in Chapter 3, Chapter 4 and Chapter 5, where each of those chapters addressed important issues raised by the experimentation undertaken and the analysis of the resulting data. Chapter 6 identified three of the fundamental principal key issues that can also be commonly encountered with most behavioural biometrics, and provided a set of additional experiments, that tested and explored some main important aspects of those areas.

This chapter presented a set of newly developed features that were generated by processing the keystroke timing data selectively, to reflect the timing features for the most frequently occurring finger movement across all subjects, in order to provides a greater focus on the typing behaviour template than a traditional feature set. An additional feature was also generated by measuring the distance between each key and combining this additional information with the newly generated timing feature to calculate the speed of participants fingers travelling between one key to the subsequent key when typing certain combinations of keys.

The newly generated features have been tested with the same data processing infrastructure that was used in the previous experiments. As demonstrated in this chapter, this feature set has achieved the same level of predictive accuracy, but with the difference that this feature set only used a part of the raw data for feature generation. Therefore, computationally, it might be argued that this is more efficient in cases where a significantly larger dataset is required to be processed, or can be effective when raw data availability is limited.

This chapter also presented a study of the predictive performance achievable when a more powerful classifier structure, a multi-agent based classifier system, was used. The multi-agent system is based on a multi classifier system, where prediction results from different classifiers are strategically combined to achieve a better predictive accuracy. The proposed multi-agent system was designed to improve the flexibility of the results combination strategy. This system showed a significant improvement in predictive performances in comparison with the best results achievable using conventional multi-classifier system using a standard typical classifier combination strategy. It is clear that by adopting a multi-agent system, better predictive performance can be achieved, which is consistent with their application to other biometric processing tasks.

Chapter 6 presented a further investigation of the ground truth labelling method since, as reported in Chapter 1, the methodologies typically adopted are not standardized and some are even rather problematic, making this one of the most challenging tasks in the predictive aspects of behavioural biometrics, especially for emotion predictions. This chapter presented two new methods of label assignment. The new methods focused on the strategy of interpreting the basic emotion scores since, even with the same basic emotion scores, the assigned labels can be very differently interpreted. The new methods were tested using the traditional keystroke feature set, and the results indicate that different labelling methods considered can have a significant impact on the predictive performance. This observation emphasises the importance of providing a fair and meaningful ground truth labelling method in this area of study. The new methods that were developed in this study therefore provide an initial base for further development of more rigorous and better ground truth labelling methods in the future.

The research reported here suggests that there is great potential in the higher-level soft-biometrics information prediction using hand-oriented biometrics. With a larger database, the prediction results confirmed the findings in the literature and suggest that a firmer conclusion can be drawn from the prediction results with more confidence. Equally, there are some important issues which need to be further explored in order to eventually optimise predictive performance for practical applications. This thesis has afforded the author the opportunity to expand the less-extensively studied areas of soft-biometrics, and develop work that provides further advancement and insightful messages to aid future development of higher-level mental state prediction biometric systems.

7.2 Future work

This thesis has reported a comprehensive study which has not only provided some new insights into the implementation of emotion prediction systems based on hand-oriented behavioural biometrics, and expanded the scale of the databases available for this type of study compared to those reported in the literature, confirming the predictive capability of hand-oriented behavioural biometrics and enhancing the confident in the conclusions draw from the emotion prediction results. It has also raised and addressed some important issues that can be further developed in future work.

The database acquired also contains other emotional state scores, for example boredom, and other demographic information such as handedness, hand size. Thus, the studies of predicting those types of information can be the targets of following investigations, as the acknowledgement of such information can help many real life scenarios, for example, forensic investigation, access control or internet security. As covered by Chapter 6, there are three main areas that suggest themselves as immediately requiring more extensive studies, including development of new feature sets, applying more powerful classification techniques, investigation of a more complete and further consideration of robust ground truth labelling methods.

An exemplifying study of feature sets was developed to aim to better reflect the behaviour pattern, and the results provided some interesting findings and suggested the need for further experimentations. Alternative classification processes would be an area for further development, perhaps developing further techniques such as classifier fusion, or more optimum settings for the reported classifiers and the multi-agent classification system, can be applied to improve the prediction performance. The initial investigation into the alternative ground truth labelling methods was established in Chapter 6. The results provide some insight into the impact of different methods, and suggest that more extensive work can be developed in this aspect of the system.

Another intriguing area of interest is the combination of different modalities for soft-biometrics information prediction, which have been seen in the literature, and the newly collected database can provide a unique opportunity for investigating the combination of the handwriting and keystroke data for conventional identification, lower-level soft-biometrics information prediction and higher-level soft-biometrics information prediction.

The extensive analysis of the keystroke data reported in Chapter 5 demonstrated that the data captured are sufficient for the emotion prediction, although the design of the data collection tasks can be refined, and more data can be effectively collected. In fact, at the final stage of this study, with a fortunate funding opportunity, a further data collection for keystroke study was implemented. The additional data capture refined the task implementation for keystroke, enabling richer data can be captured without completely changing the structure of the previous version of the data acquisition system. Although this further data collection opportunity arose too late for inclusion into this reported study, the further data acquired can be integrated in the future with the data from first capture. The additional database contains a further 60 subjects, which will be a major enhancement of this facility. A pin microphone is also fitted into the keyboard, and with the updated capture software, and the sound data are also recorded in parallel with the keystroke timing data, allowing opportunities for developing a further new exploration of relevant features, as introduced in [16], it should be also noted that the by fitting the microphone into the keyboard give us a closer capture range than the set up reported in [16], where the keystroke sound are recorded by a camera which is set on top of the screen.

The research reported in this thesis has laid down some ground work and proposed many intriguing directions for future work in higher-level soft-biometric information prediction.

7.3 Chapter conclusion

A summary of the research carried out and contributions to higher-level soft-biometrics information prediction has been reported in this chapter. A collection of potential future work is also highlighted for further development.

The author hopes that the work presented in this thesis has provided a data acquisition basic baseline standard for the area of hand-oriented soft-biometrics information studies, highlighting some insightful facts about the data processing progress, emphasising the importance of a range of diverse issues and demonstrating some guidance on the direction to solve the issues, and eventually make this technique more effective in real life practical applications to better serve our future life.

APPENDIX A

Support documentation for data acquisition

University of Kent

PARTICIPANT INFORMATION SHEET

Study of interactive modes in social media – data collection

Purpose

You are being invited to take part in a research study to help us to understand the way in which people interact with modern computer systems. The aim of this project is to build up a database which can provide us with a wide range of information about the typing patterns of individuals when using a normal keyboard and the way in which handwritten input is produced as an alternative to keyboard entry of typing patterns with mental states label. Before you decide to participate it is important for you to understand why the data collection is being carried out and what it will involve. Feel free to ask the researcher if there is anything which is not clear or if you would like more information. Take time to decide whether or not you wish to participate. Thank you for reading this.

Parts of the Study:

The study will involve essentially three parts, as follows:

PART A:

In Part A of the experiment we will record your typing patterns as you type predetermined fixed text. All that is required is for you to type on a standard keyboard.

PART B:

In Part B of the experiment we will record your typing patterns in a task where the text is not predetermined. To do this we will show you various images which we ask you to type some information about. All that is required is for you to type on a standard keyboard.

PART C:

In Part C of the experiment we will ask you to undertake some similar, but shorter tasks where you will use an electronic pen (with exactly the same feel as a normal biro pen) to write freehand on an input tablet.

We do not expect the whole test to take longer than around 40 minutes, and we will pay you £15 for your participation.

What will happen to the samples I provide?

The data that you donate will form part of a database which will be owned and maintained by the University of Kent to support its research in the area of human-computer interaction.

When you participate your samples will be stored on a secure Departmental server and linked to a reference number rather than your name, and there is no way that we will be able to identify you personally from the information stored. Participation in the data collection process is voluntary and you are permitted to withdraw at any time, without giving any reason.

University of Kent

PARTICIPANT INFORMATION SHEET

What will happen to the results of an evaluation of the database?

The results of the evaluation will be documented and are likely to be published in the scientific literature to help others benefit in the future from the knowledge we have gained. However, no participant will ever be identified individually - indeed, it is not possible for us to do this since all the data will remain anonymous. Copies of any publication will be available via the contact point noted below.

Contact for further information:

Cheng Li

School of Engineering and Digital Arts

University of Kent

Canterbury

Kent, CT2 7NT

Email: cl382@kent.ac.uk

University of Kent

PRE – EXPERIMENT PARTICIPANT DETAILS

Participant Information	
First Name:	
Surname:	
Email:	
Contact telephone number:	
Researcher to fill in	
Participant Identification Number for this project:	
Date of Capture:	

Please note: This sheet will be destroyed within 7 days of the date of experiment taken.

University of Kent

DATA SHEET

ID number: ________

Please answer the following questions:

1. Into which age band do you fall:

   ◯ <25 ◯ 25-40 ◯ 40-60 ◯ >60

2. What is your gender:

   ◯ Male ◯ Female

3. Please tell us if you are left or right handed, or ambidextrous:

   ◯ Left-handed ◯ Right-handed ◯ Ambidextrous

4. Please give an indication of your level of familiarity with using a keyboard:

   ◯ Little experience (rarely) ◯ Average experience (most days) ◯ Highly experienced (very frequent)

5. Please select your hand size.

   ◯ Small ◯ Medium ◯ Large

6. What is your highest level of educational qualification:

   ◯ School ◯ Undergraduate ◯ Postgraduate ◯ Prefer not to say

THE VERBAL SCRIPT

Pre-recording:

“Welcome to our experiment today. Please make sure you have read and understand the information sheet. If you have any question please feel free to ask now.”

After they confirm that they understand and happy with all the information they have been provided in the information sheet:

”Now please read the consent form and sign it if you are feeling comfortable with it.”

After they signed the consent form:

“The experiment consists of 7 tasks and it will take you from 15 – 25 mins to complete”

Task 1 – Section 1:

“This task is a ‘word repetition task’. Please copy all the words in the blank box at the bottom of the window. Once you finish typing in all the words we will automatically move on to the next task.”

Task 1 – Section 2:

“This task is a ‘paragraph repetition task’. Please copy all the text in the ‘Enter Text Here’ box.”

When they finish typing all the text:

“Now please click on ‘Enter’ button to move on to the next task.”

Task 1 – Section 3:

“This task is a ‘number repetition task’. Please copy all the number in the ‘Enter Text Here’ box.”

When they finish typing the number 3 times:

“Now please click on ‘Enter’ button to move on to the next task.”

Task 2:

“This task is a ‘Media description task”. Please use your own words to describe the escape route for the maze: The yellow arrow is the starting point of the maze and the blue arrow is the finishing point of the maze. Every square can be described as a step. For example: To escape the maze I would enter and move forward 4 steps and turn right. ”

When they finish describing:

“Now please click on ‘Enter’ button to move on to then next task.”

University of Kent

THE VERBAL SCRIPT

Task 3:

“This task is a ‘Media description task”. Please watch the media provided and use your own words to describe it.”

When they finish describing:

“Now please click on ‘Enter’ button to move on to then next task.”

Task 4:

“This task is a ‘Media description task”. Please watch the media provided and use your own words to describe it.”

When they finish describing:

“Now please click on ‘Enter’ button to move on to then next task.”

Task 5:

“This task is a ‘Timed task’. For this task you will have 10 seconds to type in all the words. Time will only start counting down when you type in the first letter:”

Bibliography

[1] S. C. Dass, "Fingerprint-based recognition," *International Statistical Review*, vol. 81, no. 2, pp. 175–187, 2013.

[2] T. Divan and V. Gulhane, "A fingerprint matching technique using minutiae based algorithm for voting system: A survey," in *2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. IEEE, mar 2015, pp. 1–5. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7226054

[3] A. N. Kataria, D. M. Adhyaru, A. K. Sharma, and T. H. Zaveri, "A survey of automated biometric authentication techniques," in *2013 Nirma University International Conference on Engineering (NUiCONE)*. IEEE, nov 2013, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6780190

[4] P. Pakutharivu and M. V. Srinath, "A Comprehensive Survey on Fingerprint Recognition Systems," *Indian Journal of Science and Technology*, vol. 8, no. 35, pp. 1–7, 2015.

[5] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65–77, 1992.

[6] A. S. Tolba, A. H. El-baz, and A. A. El-Harby, "Face Recognition : A Literature Review," *International Journal of Signal Processing*, vol. 2, no. 2, pp. 88–103, 2006. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.108.9118{&}rep=rep1{&}type=pdf

[7] T. Wu, S.-y. Fu, G.-s. Yang, and X.-k. Kuai, "Survey of the facial information cognition: Identity, affective and cross-culture analysis," in *2012 IEEE 11th International Conference on Cognitive Informatics and Cognitive Computing*. IEEE, aug 2012, pp. 504–509. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6311200

[8] B. Mishra, S. L. Fernandes, K. Abhishek, A. Alva, C. Shetty, C. V. Ajila, D. Shetty, H. Rao, and P. Shetty, "Facial expression recognition using feature based techniques and model based techniques: A survey," in *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, no. Icecs. IEEE, feb 2015, pp. 589–594. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7124976

[9] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, "Image understanding for iris biometrics: A survey," *Computer Vision and Image Understanding*, vol. 110, no. 2, pp. 281–307, 2008.

[10] N. Kaur and M. Juneja, "A review on Iris Recognition," in *2014 Recent Advances in Engineering and Computational Sciences, RAECS 2014*, 2014.

[11] F. M. Alkoot, "A review on advances in iris recognition methods," *International Journal of Computer Engineering Research*, vol. 3, no. 1, pp. 1–9, 2012. [Online]. Available: http://www.academicjournals.org/ijcer/PDF/Pdf2012/Alkoot.pdf$\delimiter"026E30F$nhttp://www.academicjournals.org/ijcer/abstracts/abstracts/abstract2012/Alkoot.htm

[12] E. Chandra and C. Sunitha, "A review on speech and speaker authentication system using voice signal feature selection and extraction," in *2009 IEEE International Advance Computing Conference, IACC 2009*, 2009, pp. 1341–1346.

[13] M. Fairhurst, M. Erbilek, and M. Da Costa-Abreu, "Selective review and analysis of aging effects in biometric system implementation," pp. 294–303, 2015.

[14] M. C. Fairhurst, *Age Factors in Biometric Processing*, Fairhurst, Ed. Institution of Engineering and Technology, oct 2013. [Online]. Available: http://digital-library.theiet.org/content/books/pc/pbsp010e

[15] S. Bhatt and T. Santhanam, "Keystroke dynamics for biometric authentication-A survey," *Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, PRIME 2013*, pp. 17–23, 2013. [Online]. Available: http://ieeexplore.ieee.org/ielx7/6482756/6496435/06496441.pdf?tp={&}arnumber=6496441{&}isnumber=6496435$\delimiter"026E30F$nhttp://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6496441{&}queryText=authentication+survey{&}refinements[]={&}ranges[]=2012{_}2015{_}Year{&}matchBoo

[16] J. Roth, X. Liu, A. Ross, and D. Metaxas, "Biometric authentication via keystroke sound," *2013 International Conference on Biometrics*, 2013.

[17] M. Fairhurst, Michael C.and Da Costa-Abreu, "Using keystroke dynamics for gender identification in social network environment," in *4th International Conference on Imaging for Crime Detection and Prevention (ICDP 2011)*, 2011, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=6203675$\delimiter"026E30F$nhttp://digital-library.theiet.org/content/conferences/10.1049/ic.2011.0124

[18] S. P. Banerjee and D. Woodard, "Biometric Authentication and Identification Using Keystroke Dynamics: A Survey," *Journal of Pattern Recognition Research*, vol. 7, no. 1, pp. 116–139, 2012. [Online]. Available: http://jprr.org/index.php/jprr/article/view/427$\delimiter"026E30F$nhttp://www.jprr.org/index.php/jprr/article/view/427/167

[19] D. Gunetti and C. Picardi, "Keystroke analysis of free text," *ACM Transactions on Information and System Security*, vol. 8, no. 3, pp. 312–347, 2005. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1085126.1085129

[20] M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang, "Human computing and machine understanding of human behavior: A survey," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4451 LNAI, pp. 47–71, 2007. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-540-72348-6{_}3

[21] E. Maiorana, P. Campisi, N. González-Carballo, and A. Neri, "Keystroke dynamics authentication for mobile phones," *Proceedings of the 2011 ACM*

Symposium on Applied Computing SAC 11, pp. 21–26, 2011. [Online]. Available: http://dl.acm.org/citation.cfm?id=1982190

[22] J.-W. Lee, S.-S. Choi, and B.-R. Moon, "An evolutionary keystroke authentication based on ellipsoidal hypothesis space," *Proceedings of the 9th annual conference on Genetic and evolutionary computation - GECCO '07*, p. 2090, 2007. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1276958.1277365

[23] S. Modi, W. Lafayette, S. J. Elliott, and W. Lafayette, "Keystroke Dynamics Verification Using a Spontaneously Generated Password," in *Carnahan Conferences Security Technology*, 2006, pp. 116–121. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=4105326

[24] R. Giot, M. El-Abed, B. Hemery, and C. Rosenberger, "Unconstrained keystroke dynamics authentication with shared secret," *Computers and Security*, vol. 30, no. 6-7, pp. 427–445, sep 2011. [Online]. Available: http://dx.doi.org/10.1016/j.cose.2011.03.004http://www.sciencedirect.com/science/article/pii/S0167404811000502

[25] P. H. Pisani and A. C. Lorena, "A systematic review on keystroke dynamics," *Journal of the Brazilian Computer Society*, vol. 19, no. 4, pp. 573–587, jul 2013. [Online]. Available: http://link.springer.com/10.1007/s13173-013-0117-7

[26] K. Renaud and A. De Angeli, "My password is here! An investigation into visuo-spatial authentication mechanisms," *Interacting with Computers*, vol. 16, no. 6, pp. 1017–1041, dec 2004. [Online]. Available: http://iwc.oxfordjournals.org/cgi/doi/10.1016/j.intcom.2004.06.012

[27] P. Teh, A. Teoh, C. Tee, and T. Ong, "Keystroke dynamics in password authentication enhancement," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8618–8627, dec 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0957417410006019http://www.sciencedirect.com/science/article/pii/S0957417410006019

[28] T. Chang, "Dynamically generate a long-lived private key based on password keystroke features and neural network," *Information Sciences*, vol. 211, pp. 36–47, nov 2012. [Online]. Available:

http://linkinghub.elsevier.com/retrieve/pii/S002002551200254Xhttp://www.sciencedirect.com/science/article/pii/S002002551200254X

[29] S. Z. Syed Idrus, E. Cherrier, C. Rosenberger, and P. Bours, "Soft biometrics for keystroke dynamics: Profiling individuals while typing passwords," *Computers & Security*, vol. 45, pp. 147–155, sep 2014. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0167404814000893

[30] P. S. Teh, A. B. J. Teoh, and S. Yue, "A survey of keystroke dynamics biometrics," *The Scientific World Journal*, vol. 2013, 2013.

[31] I. Tsimperidis and V. Katos, "Keystroke forensics: are you typing on a desktop or a laptop?" *Proceedings of the 6th Balkan Conference in ...*, pp. 89–94, 2013. [Online]. Available: http://dl.acm.org/citation.cfm?id=2490272

[32] K. Yu, Y. Wang, and T. Tieniu, "Writer identification using dynamic features," *Biometric Authentication*, pp. 512–518, 2004. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-540-25948-0{_}70

[33] M. Tapiador and J. a. Sigüenza, "Writer Identification Method Based on Forensic Knowledge," in *Biometric Authentication*, 2004, pp. 555–561. [Online]. Available: http://link.springer.com/10.1007/978-3-540-25948-0{_}76

[34] C. Villager and J. Dittmann, "Biometrics for User Authentication," in *Encyclopedia of Multimedia.* Boston, MA: Springer US, 2008, pp. 48–55. [Online]. Available: http://www.springerlink.com/index/10.1007/978-0-387-78414-4{_}124

[35] A. K. Jain, F. D. Griess, and S. D. Connell, "On-line signature verification," *Pattern Recognition*, vol. 35, no. 12, pp. 2963–2972, 2002.

[36] V. Pervouchine and G. Leedham, "Extraction and analysis of forensic document examiner features used for writer identification," *Pattern Recognition*, vol. 40, no. 3, pp. 1004–1013, 2007.

[37] K. M. B. Abdl and S. Z. M. Hashim, "Handwriting Identification: A direction review," in *ICSIPA09 - 2009 IEEE International Conference on*

Signal and Image Processing Applications, Conference Proceedings, 2009, pp. 459–463.

[38] S. N. Srihari, S.-H. Cha, H. Arora, and S. Lee, "Individuality of handwriting." *Journal of forensic sciences*, vol. 47, no. 4, pp. 856–872, 2016.

[39] a. Schlapbach and H. Bunke, "Off-Line Handwriting Identification Using HMM Based Recognizers," *Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004)*, vol. 2, pp. 654–658 Vol.2, 2004.

[40] B. Zhang and S. N. Srihari, "Analysis of handwriting individuality using word features," *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, vol. 2003-Janua, pp. 1142–1146, 2003.

[41] S. N. Srihari, S. H. Cha, and S. Lee, "Establishing Handwriting individuality using pattern recognition techniques," *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, vol. 2001-Janua, no. i, pp. 1195–1204, 2001.

[42] H. Said, "Personal identification based on handwriting," *Pattern Recognition*, vol. 33, no. 59825105, pp. 149–160, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320399000060

[43] S. Pal, M. Blumenstein, and U. Pal, "Off-line signature verification systems: a survey," *Proceedings of the International Conference & Workshop on Emerging Trends in Technology*, no. October, pp. 652–657, 2011.

[44] R. Plamondon, R. Plamondon, S. Srihari, and S. Srihari, "On-line and Off-line handwriting recognition: A comprehensive survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 63–84, 2000.

[45] Cong Shen, Xiao-gang Ruan, and Tian-Lu Mao, "Writer identification using Gabor wavelet," *Proceedings of the 4th World Congress on Intelligent Control and Automation (Cat. No.02EX527)*, vol. 3, no. d, pp. 2061–2064, 2002. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1021447

[46] A. Bensefia, A. Nosary, T. Paquet, L. Heutte, U. F. R. Sciences, U. D. Rouen, and F.-M.-s.-a. Cedex, "Writer Identification By Writer ' s Invariants," *Proceedings of the 2002 Frontiers in Handwriting Recognition,*, pp. 274–279, 2002.

[47] C. I. Tomai, B. Zhang, and S. N. Srihari, "Discriminatory power of handwritten words for writer recognition," *Proceedings - International Conference on Pattern Recognition*, vol. 2, pp. 638–641, 2004.

[48] A. Bensefia, T. Paquet, and L. Heutte, "A writer identification and verification system," *Pattern Recognition Letters*, vol. 26, no. 13, pp. 2080–2092, 2005.

[49] A. Montanari and E. Richard, "A statistical model for tensor PCA," *Advances in Neural Information Processing Systems*, vol. 1, no. 1, pp. 1–30, 2014. [Online]. Available: http://arxiv.org/abs/1411.1076

[50] W. Khalifa, A. Salem, and M. Roushdy, "A Survey of EEG Based User Authentication Schemes," *8th International Conference on Informatics and Systems (INFOS)*, pp. 55–60, 2012.

[51] R. Blanco-Gonzalo, R. Sanchez-Reillo, O. Miguel-Hurtado, and J. Liu-Jimenez, "Usability analysis of dynamic signature verification in mobile environments," *Biometrics Special Interest Group (BIOSIG), 2013 International Conference of the*, pp. 1–9, 2013.

[52] L. Lu, F. Claret-Tournier, C. R. Chatwin, R. C. D. Young, and L. Zhongming, "An agent-oriented mobile payment system secured using a biometrics approach," *International Journal of Agent-Oriented Software Engineering*, vol. 3, no. 2-3, pp. 163–187, 2009. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-62349101148{&}partnerID=40{&}md5=e428f1a50c5691a28b0195e410433b0c

[53] J. Angulo and E. Wästlund, "Exploring Touch Screen Biometrics for User Identification on Smart Phones," *Privacy and Identity 2011, IFIP AICT*, pp. 130–143, 2012.

[54] Y. Ge, J. Lu, X. Feng, and D. Yang, "Body-based human age estimation at a distance," *Electronic Proceedings of the 2013 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2013*, pp. 2–5, 2013.

[55] P. H. Hennings-Yeomans, S. Baker, and V. Kumar, "Recognition of Low-Resolution Faces Using Multiple Still Images\nand Multiple Cameras," *Btas*, 2008.

[56] M. Tistarelli, S. Z.LI, and R. Chellappa, "Biometrics at a Distance: Issues, Challenges, and Prospects," in *Handbook of Remote Biometrics*, 2009, vol. 6256, pp. 3–21. [Online]. Available: http://www.springerlink.com/content/w437828845137 65t/

[57] R. Santos and a. Alexandre, "The UBIRIS . v2 : A Database of Visible Wavelength Iris Images Captured," *Analysis*, vol. 32, no. 8, pp. 1529–1535, 2010.

[58] M. Burmester and J. Mulholland, "The advent of trusted computing: implications for digital forensics," *SAC 06 Proceedings of the 2006 ACM symposium on Applied computing*, vol. 1, pp. 283–287, 2006. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-33751050416{&}partnerID=40{&}md5=5b32f6750c1b59e9eabc4c40cea3490e

[59] M. Choraś, "Intelligent computing for automated biometrics, criminal and forensic applications," *3rd International Conference on Intelligent Computing, ICIC 2007*, vol. 4681 LNCS, pp. 1170–1181, 2007. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-38049054412{&}partnerID=40{&}md5=8a3adbac8d0966048077aea2b428126e

[60] A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: A tool for information security," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 125–143, 2006.

[61] N. A. Spaun and R. W. V. Bruegge, "Forensic Identification of People from Images and Video," *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems*, pp. 1–4, 2008.

[62] A. Suman, "Using 3D pose alignment tools in forensic applications of Face Recognition," *BTAS 2008 - IEEE 2nd International Conference on Biometrics: Theory, Applications and Systems*, vol. 00, 2008.

[63] Y. Zhang, C. McCullough, J. R. Sullins, and C. R. Ross, "Human and computer evaluations of face sketches with implications for forensic investigations," *BTAS 2008 - IEEE 2nd International Conference on Biometrics: Theory, Applications and Systems*, 2008.

[64] G. Guo and G. Mu, "Joint Estimation of Age, Gender and Ethnicity: CCA vs PLS," *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, 2013.

[65] H. Han, C. Otto, X. Liu, and A. K. Jain, "Demographic estimation from face images: Human vs. machine performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1148–1161, 2015.

[66] V. Khryashchev, A. Priorov, and A. Ganin, "Gender and age recognition for video analytics solution," *Proceedings - Applied Imagery Pattern Recognition Workshop*, vol. 2015-Febru, no. February, 2015.

[67] J. Lu and Y. P. Tan, "Gait-based human age estimation," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 761–770, 2010.

[68] T. X. Fujisawa, "Subjective age estimation using speech sounds: Comparison with facial images," *2008 IEEE International Conference on Systems, Man and Cybernetics*, pp. 1900–1904, 2008. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4811567

[69] A. Sgroi, K. W. Bowyer, and P. J. Flynn, "The prediction of old and young subjects from iris texture," *Proceedings - 2013 International Conference on Biometrics, ICB 2013*, pp. 1–5, 2013.

[70] Y. Wang, K. Ricanek, C. Chen, and Y. Chang, "Gender classification from infants to seniors," *IEEE 4th International Conference on Biometrics: Theory, Applications and Systems, BTAS 2010*, 2010.

[71] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, 2014.

[72] S. E. Bekhouche, A. Ouafi, A. Benlamoudi, A. Taleb-Ahmed, and A. Hadid, "Facial age estimation and gender classification using multi level local phase quantization," *3rd International Conference on Control, Engineering and Information Technology, CEIT 2015*, pp. 3–6, 2015.

[73] M. H. Bahari and H. Van Hamme, "Speaker age estimation and gender detection based on supervised non-negative matrix factorization," *BioMS*

2011 - 2011 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications, Proceedings, no. 238803, pp. 27–32, 2011.

[74] S. M. Ali, Z. Darbar, and K. N. Junejo, “Age Estimation from Facial Images Using Biometric Ratios and Wrinkle Analysis,” *5th National Symposium on Information Technology: Towards New Smart World (NSITNSW)*, 2015.

[75] V. Saran, S. Kumar, G. Ak, and S. Ahmad, “Differentiation of Handedness of Writer Based on their Strokes and Characteristic Features,” *Journal of forensic Research*, vol. 4, no. 5, pp. 4–6, 2013.

[76] E. Griechisch and E. Bencsik, “Handedness Detection of Online Handwriting based on Horizontal Strokes,” *13th International Confrence on Document Analysis and Recognition - ICDAR’15*, pp. 1272–1277, 2015.

[77] M. Liwicki, A. Schlapbach, P. Loretan, and H. Bunke, “Automatic Detection of Gender and Handedness from On-Line Handwriting,” *Journal of Social Psychology*, pp. 179–183, 2007.

[78] M. Fairhurst, Michael C.and Erbilek and C. Li, “Enhancing the forensic value of handwriting using emotion prediction,” in *2nd International Workshop on Biometrics and Forensics, IWBF 2014*, 2014.

[79] M. C. Fairhurst, C. Li, and M. Erbilek, “Exploiting biometric measurements for prediction of emotional state: A preliminary study for healthcare applications using keystroke analysis,” in *BIOMS 2014 - 2014 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications, Proceedings*, 2014, pp. 74–79. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=6951539

[80] H. Dibekliolu, A. A. Salah, and L. Akarun, “3D facial landmarking under expression, pose, and occlusion variations,” *BTAS 2008 - IEEE 2nd International Conference on Biometrics: Theory, Applications and Systems*, vol. 00, pp. 3–8, 2008.

[81] Y. Fu, G. Guo, and S. Member, “Age Synthesis and Estimation via Faces: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010.

[82] K. Ricanek and B. Barbour, "What are soft biometrics and how can they be used?" *Computer*, vol. 44, no. 9, pp. 106–108, 2011.

[83] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, "A Comparative Study of Local Feature Extraction for Age Estimation," *Int. Conf. Control, Automation, Robotics and Vision*, no. December, pp. 1280–1284, 2010.

[84] Kuang-Yu Chang and Chu-Song Chen, "A Learning Framework for Age Rank Estimation Based on Face Images With Scattering Transform," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 785–798, mar 2015. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7001258

[85] T. Wu, P. Turaga, and R. Chellappa, "Age Estimation and Face Verification Across Aging Using Landmarks," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1780–1788, dec 2012. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6272347

[86] X. Wang, R. Guo, and C. Kambhamettu, "Deeply-learned feature for age estimation," *Proceedings - 2015 IEEE Winter Conference on Applications of Computer Vision, WACV 2015*, pp. 534–541, 2015.

[87] C. Li, Q. Liu, W. Dong, X. Zhu, J. Liu, and H. Lu, "Human Age Estimation Based on Locality and Ordinal Information," *IEEE Transactions on Cybernetics*, vol. 45, no. 11, pp. 2522–2534, 2015.

[88] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, pp. 112–119, 2009.

[89] C. Shan, "Learning local features for age estimation on real-life faces," *Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis - MPVA '10*, p. 23, 2010. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1878039.1878045

[90] C. Li, Q. Liu, J. Liu, and H. Lu, "Learning ordinal discriminative features for age estimation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2570–2577, 2012.

[91] Y. Liang, L. Liu, Y. Xu, Y. Xiang, and B. Zou, "Multi-task GLOH feature selection for human age estimation," *Proceedings - International Conference on Image Processing, ICIP*, pp. 565–568, 2011.

[92] A. Torrisi, G. M. Farinella, G. Puglisi, and S. Battiato, "Selecting discriminative CLBP patterns for age estimation," in *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW).* IEEE, jun 2015, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7169755

[93] P. Turaga, S. Biswas, and R. Chellappa, "The role of geometry in age estimation," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 946–949, 2010.

[94] Z. Li, U. Park, and A. K. Jain, "A discriminative model for age invariant face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3 PART 2, pp. 1028–1037, 2011.

[95] Z. Li, Y. Fu, and T. S. Huang, "A robust framework for multiview age estimation," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops.* IEEE, jun 2010, pp. 9–16. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5543813

[96] G. Guo and X. Wang, "A study on human age estimation under facial expression changes," *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2547–2553, 2012.

[97] K. Luu, K. Ricanek, T. D. Bui, C. Y. Suen, Ieee, K. Ricanek Jr, T. D. Bui, and C. Y. Suen, "Age estimation using active appearance models and support vector machine regression," *IEEE 3rd International Conference on Biometrics: Theory, Applications and Systems*, pp. 314–318, 2009.

[98] H. T. Chang and H. W. Peng, "Facial Image Prediction Using Exemplar-based Algorithm and Non-negative Matrix Factorization," in *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2012, pp. 1–4.

[99] Ashutosh, B. Laskar, S. Kumar, and S. Majumder, "Gene expression programming based age estimation using facial features," *2013 IEEE 2nd*

International Conference on Image Information Processing, IEEE ICIIP 2013, pp. 442–446, 2013.

[100] Y. Piao and M. Kudo, "How Do Facial Expressions Contribute to Age Prediction?" *2013 2nd IAPR Asian Conference on Pattern Recognition*, pp. 882–886, 2013. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6778457

[101] Y. Z. Y. Zhang and D.-Y. Y. D.-Y. Yeung, "Multi-task warped Gaussian process for personalized age estimation," *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 2622–2629, 2010.

[102] J. Lu and Y. P. Tan, "Ordinary preserving manifold analysis for human age and head pose estimation," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 2, pp. 249–258, 2013.

[103] Y.-l. Chen and C.-t. Hsu, "Subspace Learning for Facial Age Estimation Via Pairwise Age Ranking," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 12, pp. 2164–2176, dec 2013. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6637077

[104] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "A Probabilistic Fusion Approach to human age prediction," *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops*, 2008.

[105] S. Kohli, S. Prakash, and P. Gupta, "Age Estimation Using Active Appearance Models and Ensemble of Classifiers with Dissimilarity-Based Classification," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6838 LNCS, pp. 327–334, 2011.

[106] C. Zhang and G. Guo, "Age estimation with expression changes using multiple aging subspaces," *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 1–6, 2013. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6712720

[107] J. Lu and Y.-P. Tan, "Fusing shape and texture information for facial age estimation," in *2011 IEEE International Conference on*

Acoustics, Speech and Signal Processing (ICASSP). IEEE, may 2011, pp. 1477–1480. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5946772

[108] M. Erbilek, M. C. Fairhurst, and M. D. Costa-Abreu, "Improved age prediction from biometric data using multimodal configurations," in *BIOSIG*, 2014.

[109] M. C. Fairhurst and M. Abreu, "An investigation of predictive profiling from handwritten signature data," *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, pp. 1305–1309, 2009.

[110] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi, "Gait-based age estimation using a whole-generation gait database," in *2011 International Joint Conference on Biometrics (IJCB).* IEEE, oct 2011, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6117531

[111] M. Erbilek, M. C. Fairhurst, and M. Abreu, "Age Prediction from Iris Biometrics," in *5th International Conference on Imaging for Crime Detection and Prevention (ICDP 2013).* Institution of Engineering and Technology, 2013, pp. 1.07–1.07. [Online]. Available: http://digital-library.theiet.org/content/conferences/10.1049/ic.2013.0258

[112] G. Dobry, R. M. Hecht, M. Avigal, and Y. Zigel, "Supervector Dimension Reduction for Efficient Speaker Age Estimation Based on the Acoustic Speech Signal," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 1975–1985, sep 2011. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5711646

[113] S. Gautam, "Developmental Pattern Analysis and Age Prediction by Extracting Speech Features and Applying Various Classification Techniques," *International Conference on Computing, Communication and Automation (ICCCA2015)*, pp. 83–87, 2015.

[114] R. Merkel, M. Hildebrandt, and J. Dittmann, "Application of stirtrace benchmarking for the evaluation of latent fingerprint age estimation robustness," in *3rd International Workshop on Biometrics and Forensics*

(IWBF 2015). IEEE, mar 2015, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7110221

[115] D. Cao, C. Chen, M. Piccirilli, D. Adjeroh, T. Bourlai, and A. Ross, "Can facial metrology predict gender?" *2011 International Joint Conference on Biometrics, IJCB 2011*, no. October, 2011.

[116] E. Mäkinen and R. Raisamo, "An experimental comparison of gender classification methods," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1544–1556, 2008.

[117] F. Matta, U. Saeed, C. Mallauran, and J.-L. Dugelay, "Facial gender recognition using multiple sources of visual information," *2008 IEEE 10th Workshop on Multimedia Signal Processing*, pp. 785–790, 2008. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4665181

[118] Fok Hing, Chi Tivive, and A. Bouzerdoum, "A Shunting Inhibitory Convolutional Neural Network for Gender Classification," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 2006, no. August. IEEE, 2006, pp. 421–424. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1699868

[119] M. Toews and T. Arbel, "Detection, Localization, and Sex Classification of Faces from Arbitrary Viewpoints and under Occlusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1567–1581, sep 2009. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4633361

[120] Z.-h. Wang, "Gender classification using selected independent-features based on Genetic Algorithm," *Machine Learning and Cybernetics, 2009*, no. July, pp. 12–15, 2009. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=5212504

[121] A. Bhaskar and Aneesh R.P., "Advanced algorithm for gender prediction with image quality assessment," in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, aug 2015, pp. 1848–1855. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7275887

[122] S. Camalan and G. Sengul, "Gender prediction by using Local Binary Pattern and K Nearest Neighbor and Discriminant Analysis classifications," in *2016 24th Signal Processing and Communication Application Conference (SIU)*. IEEE, may 2016, pp. 2161–2164. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7496201

[123] J. E. Tapia, C. A. Perez, and K. W. Bowyer, "Gender Classification from Iris Images Using Fusion of Uniform Local Binary Patterns," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 8927, pp. 751–763. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84928819309{&}partnerID=tZOtx3y1http://link.springer.com/10.1007/978-3-319-16181-5{_}57

[124] V. Thomas, N. V. Chawla, K. W. Bowyer, and P. J. Flynn, "Learning to predict gender from iris images," *IEEE Conference on Biometrics: Theory, Applications and Systems, BTAS'07*, pp. 0–4, 2007.

[125] A. Hassaine, S. Al Maadeed, J. Aljaam, and A. Jaoua, "ICDAR 2013 Competition on Gender Prediction from Handwriting," in *2013 12th International Conference on Document Analysis and Recognition*. IEEE, aug 2013, pp. 1417–1421. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6628847

[126] N. Bouadjenek, H. Nemmour, and Y. Chibani, "Local descriptors to improve off-line handwriting-based gender prediction," in *2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*. IEEE, aug 2014, pp. 43–47. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7007979

[127] N. Bouadjenek, H. Nemmour, and Y. Chibani, "Age, gender and handedness prediction from handwriting using gradient features," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, aug 2015, pp. 1116–1120. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7333934

[128] N. Bouadjenek, H. Nemmour, and Y. Chibani, "Histogram of Oriented Gradients for writer's gender, handedness and age prediction," in

2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA). IEEE, sep 2015, pp. 1–5. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7276752

[129] M. Merler, L. Cao, and J. R. Smith, "You Are What You Tweet ... Pic ! Gender Prediction Based on Semantic Analysis of Social Media Images," *Proceedings - IEEE International Conference on Multimedia and Expo*, 2015.

[130] M. Crawford and X. Zhu, "Gender Prediction in Random Chat Networks Using Topological Network Structures and Masked Content," *Proceedings - 2015 IEEE 16th International Conference on Information Reuse and Integration, IRI 2015*, pp. 174–181, 2015.

[131] E. Gulsen, H. Gunduz, Z. Cataltepe, and L. Serinol, "Big data feature selection and projection for gender prediction based on user web behaviour," in *2015 23nd Signal Processing and Communications Applications Conference (SIU)*. IEEE, may 2015, pp. 1545–1548. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7130141

[132] F. Metze, J. Ajmera, R. Englert, U. Bub, F. Burkhardt, J. Stegmann, C. Muller, R. Huber, B. Andrassy, J. G. Bauer, and B. Littel, "Comparison of Four Approaches to Age and Gender Recognition for Telephone Applications," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*. IEEE, 2007, pp. IV–1089–IV–1092. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4218294

[133] D. Cao, C. Chen, D. Adjeroh, and A. Ross, "Predicting gender and weight from human metrology using a copula model," *2012 IEEE 5th International Conference on Biometrics: Theory, Applications and Systems, BTAS 2012*, pp. 162–169, 2012.

[134] J. Zhu, S. Liao, D. Yi, Z. Lei, and S. Z. Li, "Multi-label CNN based pedestrian attribute learning for soft biometrics," *Proceedings of 2015 International Conference on Biometrics, ICB 2015*, pp. 535–540, 2015.

[135] R. Chellappa and P. Turaga, "Recent advances in age and height estimation from still images and video," *2011 IEEE International Conference on*

Automatic Face and Gesture Recognition and Workshops, FG 2011, pp. 91–96, 2011.

[136] Y. Ran, G. Rosenbush, and Q. Zheng, "Computational approaches for real-time extraction of soft biometrics," *2008 19th International Conference on Pattern Recognition*, vol. i, pp. 1–4, 2008. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4761878

[137] P. Tome, J. Fierrez, R. Vera-Rodriguez, and M. S. Nixon, "Soft biometrics and their application in person recognition at a distance," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 3, pp. 464–472, 2014.

[138] S. R. Livingstone, A. R. Brown, and M. Rallf, "Influencing the Perceived Emotions of Music with Intent," in *Emotion*, 2005, pp. 161–170. [Online]. Available: http://espace.library.uq.edu.au/view/UQ:102953

[139] H. Nakahara, S. Furuya, T. Masuko, P. R. Francis, and H. Kinoshita, "Performing music can induce greater modulation of emotion-related psychophysiological responses than listening to music." *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, vol. 81, no. 3, pp. 152–8, sep 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21704661

[140] B. Cardoso, T. Romão, and N. Correia, "CAAT - A Discrete Approach to Emotion Assessment," *Proceedings of the 2013 Conference on Human Factors in Computing Systems*, pp. 1047–1052, 2013.

[141] E. Mower, "A framework for automatic human emotion classification using emotion profiles," *Audio, Speech, and . . .*, vol. 19, no. 5, pp. 1057–1070, 2011. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=5585726

[142] R. Adolphs, "Recognizing emotion from facial expressions: psychological and neurological mechanisms," *Behavioral and cognitive neuroscience reviews*, 2002. [Online]. Available: http://bcn.sagepub.com/content/1/1/21.short

[143] X. Li, G. Mori, and H. Zhang, "Expression-invariant face recognition with expression classification," *Third Canadian Conference on Computer*

and Robot Vision, CRV 2006, vol. 2006, 2006. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=1640432

[144] G. a. Tsihrintzis, M. Virvou, E. Alepis, and I.-O. Stathopoulou, "Towards Improving Visual-Facial Emotion Recognition through Use of Complementary Keyboard-Stroke Pattern Information," *Fifth International Conference on Information Technology: New Generations (itng 2008)*, pp. 32–37, apr 2008. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4492451

[145] H. Yan, a.N. Poo, and M. Ang, "Adaptive discriminative metric learning for facial expression recognition," *IET Biometrics*, vol. 1, no. 3, pp. 160–167, sep 2012. [Online]. Available: http://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2012.0006

[146] B. Schuller, M. Lang, and G. Rigoll, "Multimodal emotion recognition in audiovisual communication," *Multimedia and Expo, 2002. ...*, pp. 745–748, 2002. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=1035889

[147] Z. Ciota, "Emotion Recognition on the Basis of Human Speech," *2005 18th International Conference on Applied Electromagnetics and Communications*, pp. 1–4, 2005. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1613522

[148] P. Zimmermann and S. Guttormsen, "Affective computinga rationale for measuring mood with mouse and keyboard," *International journal of ...*, vol. 9, no. 4, pp. 539–551, 2003. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/14675525http://www.tandfonline.com/doi/abs/10.1080/10803548.2003.11076589

[149] E. Alepis, M. Virvou, and K. Kabassi, "Affective Student Modeling Based on Microphone and Keyboard User Actions," *Advanced Learning Technologies, 2006. Sixth International Conference on*, pp. 139–141, 2006. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=1652388

[150] L. H.-R. L.-L. W.-J. Jin, "Emotion recognition based on pressure sensor keyboards," *2008 IEEE International Conference on Multimedia*

and Expo, pp. 1089–1092, jun 2008. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4607628

[151] O. C. Santos, S. Salmeron-majadas, and J. G. Boticario, "Emotions Detection from Math Exercises by Combining Several Data Sources," in *Artificial Intelligence in Education*, 2013, pp. 742–745.

[152] A. Kolakowska, "A review of emotion recognition methods based on keystroke dynamics and mouse movements," in *2013 6th International Conference on Human System Interactions, HSI 2013*, 2013, pp. 548–555. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=6577879

[153] R. Bixler and S. D'Mello, "Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits," in *Proceedings of the 2013 international conference on Intelligent user interfaces*, 2013, pp. 225–233. [Online]. Available: http://dl.acm.org/citation.cfm?id=2449426

[154] W.-H. Tsui, P. Lee, and T.-C. Hsiao, "The effect of emotion on keystroke: An experimental study using facial feedback hypothesis." *Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, vol. 2013, pp. 2870–3, jan 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24110326

[155] K. Bakhtiyari and H. Husain, "Fuzzy Model in Human Emotions Recognition," *arXiv preprint arXiv:1407.1474*, pp. 77–82, 2014. [Online]. Available: http://arxiv.org/abs/1407.1474

[156] S. Mutalib and R. Ramli, "Towards emotional control recognition through handwriting using fuzzy inference," *Information Technology, 2008. ITSim 2008. International Symposium on*, vol. 2, pp. 1–5, 2008. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=4631735

[157] S. Kedar, D. Bormane, A. Dhadwal, S. Alone, and R. Agarwal, "Automatic Emotion Recognition through Handwriting Analysis: A Review," *2015 International Conference on Computing Communication Control and Automation*, pp. 811–816, 2015. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7155960

[158] B. P. Shan and M. Madheswaran, "Extraction of fetal biometrics using class separable shape sensitive approach for gestational age estimation," *2009 ICCTD - International Conference on Computer Technology and Development*, vol. 2, pp. 376–380, 2009.

[159] L. K. Wee, L. M. Yun, T. L. See, and E. Supriyanto, "Comparison studies of 2D and 3D ultrasound biparietal diameter for gestational age estimation," in *2011 2nd International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering*, vol. 265, no. November. IEEE, nov 2011, pp. 163–167. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6108616

[160] Z. Imaduddin, M. A. Akbar, H. A. Tawakal, I. P. Satwika, and Y. B. Saroyo, "Automatic detection and measurement of fetal biometrics to determine the Gestational Age," in *2015 3rd International Conference on Information and Communication Technology (ICoICT)*. IEEE, may 2015, pp. 608–612. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7231495

[161] R. Tranter, D. Bell, P. Gutting, C. Harmer, D. Healy, and I. M. Anderson, "The effect of serotonergic and noradrenergic antidepressants on face emotion processing in depressed patients." *Journal of affective disorders*, vol. 118, no. 1-3, pp. 87–93, nov 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19250683

[162] C. Bidet-Ildei, P. Pollak, S. Kandel, V. Fraix, and J.-P. Orliaguet, "Handwriting in patients with Parkinson disease: effect of L-dopa and stimulation of the sub-thalamic nucleus on motor anticipation." *Human movement science*, vol. 30, no. 4, pp. 783–91, aug 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21339012

[163] C. O'Reilly and R. Plamondon, "Looking for the brain stroke signature," *Pattern Recognition (ICPR)*, no. Icpr, pp. 1811–1814, 2012. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=6460504

[164] R. Mandryk, K. Inkpen, and T. Calvert, "Using psychophysiological techniques to measure user experience with entertainment technologies," *Behaviour & information ...*, vol. 25, no. 2, pp. 141–158, 2006.

[Online]. Available: http://www.informaworld.com/openurl?genre=article{&}doi=10.1080/01449290500331156{&}magic=crossrefhttp://www.tandfonline.com/doi/abs/10.1080/01449290500331156

[165] A. Kapoor, W. Burleson, and R. Picard, "Automatic prediction of frustration," *International journal of human*, vol. 65, no. 8, pp. 724–736, aug 2007. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1071581907000377http://www.sciencedirect.com/science/article/pii/S1071581907000377

[166] R. Mandryk and M. Atkins, "A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies," *International Journal of Human-Computer Studies*, vol. 65, no. 4, pp. 329–347, apr 2007. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1071581906001881

[167] J. Hong, J. Ramos, and A. Dey, "Understanding physiological responses to stressors during physical activity," *Proceedings of the 2012 ACM conference*, pp. 270–279, 2012. [Online]. Available: http://dl.acm.org/citation.cfm?id=2370260

[168] L. Vizer, "Detecting cognitive and physical stress through typing behavior," *CHI'09 Extended Abstracts on Human Factors*, p. 3113, 2009. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1520340.1520440http://dl.acm.org/citation.cfm?id=1520440

[169] Y. Gu, S.-L. Tan, K.-J. Wong, M.-H. R. Ho, and L. Qu, "A biometric signature based system for improved emotion recognition using physiological responses from multiple subjects," *2010 8th IEEE International Conference on Industrial Informatics*, pp. 61–66, jul 2010. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5549464

[170] A. Liapis and X. Michalis, "The physiological measurements as a critical indicator in users' experience evaluation," *Proceedings of the 17th Panhellenic Conference on Informatics - PCI '13*, p. 258, 2013. [Online]. Available: http://dl.acm.org/citation.cfm?id=2491845.2491883

[171] J. Sykes and S. Brown, "Affective gaming: measuring emotion through the gamepad," *CHI 03 CHI 03 extended abstracts on Human factors in*

computing systems, vol. Ft. Lauder, pp. 732–733, 2003. [Online]. Available: http://portal.acm.org/citation.cfm?id=765957

[172] Y. Gao, N. Bianchi-Berthouze, and H. Meng, "What Does Touch Tell Us about Emotions in Touchscreen-Based Gameplay?" *ACM Transactions on Computer-Human Interaction*, vol. 19, no. 4, pp. 1–30, dec 2012. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2395131.2395138

[173] C. Epp, M. Lippold, and R. L. Mandryk, "Identifying emotional states using keystroke dynamics," *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, p. 715, 2011. [Online]. Available: http://dl.acm.org/citation.cfm?doid=1978942.1979046

[174] R. Likert, "A technique for the measurement of attitudes." *Archives of Psychology*, vol. 22, no. 140, p. 55, 1932.

[175] L. M. Vizer, L. Zhou, and A. Sears, "Automated stress detection using keystroke and linguistic features: An exploratory study," *International Journal of Human Computer Studies*, vol. 67, no. 10, pp. 870–886, oct 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1071581909000937

[176] J. Hernandez and P. Paredes, "Under pressure: sensing stress of computer users," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,CHI '14*, 2014, pp. 51–60. [Online]. Available: http://dl.acm.org/citation.cfm?id=2557165

[177] S. Leon and A. Nikov, "Intelligent Emotion Oriented Ecommerce System," *Journal Recent Advances In Artificial Intelligence. . . .* , pp. 202–207, 2008. [Online]. Available: http://www.wseas.us/e-library/conferences/2010/Cambridge/AIKED/AIKED-32.pdf

[178] A. Kaklauskas, M. Krutinis, and M. Seniut, "Biometric Mouse Intelligent System for Student's Emotional and Examination Process Analysis," *2009 Ninth IEEE International Conference on Advanced Learning Technologies*, pp. 189–193, jul 2009. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5194199

[179] A. Xu, J. Biehl, E. Rieffel, T. Turner, and W. V. Melle, "Learning how to feel again: Towards affective workplace presence and communication

technologies," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 839–848, 2012. [Online]. Available: http://dl.acm.org/citation.cfm?id=2208524

[180] K. Herdem, "Reactions: Twitter based mobile application for awareness of friends' emotions," *Proceedings of the 2012 ACM Conference on . . .*, pp. 796–797, 2012. [Online]. Available: http://dl.acm.org/citation.cfm?id=2370397

[181] J. Mostow, R. Maxion, and J. Cohn, "HCC: Medium: Using keystroke timing to detect students' mental states," pp. 1–31, 2011. [Online]. Available: http://www.cs.cmu.edu/afs/cs.cmu.edu/project/listen/dmg/Listen/Proposals/2011NSFHCCkeystroke/2011NSFKeystrokeProjectDescription.pdf

[182] M. C. D. C. Abreu and M. C. Fairhurst, "Analyzing the benefits of a novel multiagent approach in a multimodal biometrics identification task," *IEEE Systems Journal*, vol. 3, no. 4, pp. 410–417, 2009.

[183] M. C. Da Costa Abreu and M. C. Fairhurst, "Enhancing identity prediction using a novel approach to combining hard- and soft-biometric information," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 41, no. 5, pp. 599–607, 2011.

[184] K. Killourhy and R. Maxion, "The effect of clock resolution on keystroke dynamics," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5230 LNCS, 2008, pp. 331–350.

[185] K. M. B. Abdl and S. Z. M. Hashim, "Swarm-based feature selection for handwriting identification," *Journal of Computer Science*, vol. 6, no. 1, pp. 80–86, 2010.

[186] S. N. Srihari and K. R. Bandi, "Writer Demographic Classification Using Bagging and Boosting," *Proceedings of the 12th Biennial Conference of the International Graphonomics Society*, no. D, pp. 133–137, 2005.

[187] M. Liwicki, A. Schlapbach, and H. Bunke, "Automatic gender detection using on-line and off-line information," *Pattern Analysis and Applications*, vol. 14, no. 1, pp. 87–92, 2011.

[188] J. Chapran, M. Fairhurst, R. Guest, and C. Ujam, "Task-related population characteristics in handwriting analysis," *Computer Vision, IET*, vol. 2, no. 2, pp. 75–87, 2008. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=4543868

[189] M. Abreu and M. C. Fairhurst, "Improving identity prediction in signature-based unimodal systems using soft biometrics," in *Biometric {ID} Management and Multimodal Communication*, 2009, pp. 348–356. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-04391-8{_}45

[190] M. Erbilek and M. C. Fairhurst, "Framework for managing ageing effects in signature biometrics," *IET Biometrics*, vol. 1, no. 2, p. 136, 2012.

[191] M. Erbilek and M. C. Fairhurst, "A methodological framework for investigating age factors on the performance of biometric systems," *Proceedings of the on Multimedia and security (MM&Sec '12). ACM, New York, NY, USA*, pp. 115–122, 2013. [Online]. Available: http://kar.kent.ac.uk/35790/

[192] A. K. Jain, S. C. Dass, and K. Nandakumar, "Soft Biometric Traits for Personal Recognition Systems," *LNCS*, vol. 3072, pp. 731–738, 2004.

[193] L. Constantine and H. Hajj, "A survey of ground-truth in emotion data annotation," *2012 IEEE International Conference on Pervasive Computing and Communications Workshops, PERCOM Workshops 2012*, no. March, pp. 697–702, 2012.

[194] I.-O. Stathopoulou and G. Tsihrintzis, "An improved neural-network-based face detection and facial expression classification system," *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, vol. 1, pp. 666–671, 2004.

[195] I. Shafran, M. Riley, and M. Mohri, "Voice signatures," in *Automatic Speech Recognition and Understanding, 2003. ASRU '03. 2003 IEEE Workshop*, 2003, pp. 31–36. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=1318399

[196] Y. Kobayashi, "The EMOSIGN-analyzing the emotion signature in human motion," *Systems, Man and Cybernetics, 2007. ISIC.*

IEEE ..., no. 3, pp. 1171–1176, 2007. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=4413929

[197] M. C. Fairhurst, M. Erbilek, and C. Li, "Study of automatic prediction of emotion from handwriting samples," *IET Biometrics*, vol. 4, no. 2, pp. 90–97, jun 2015. [Online]. Available: http://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2014.0097

[198] R. Mergl, P. Tigges, a. Schröter, H. J. Möller, and U. Hegerl, "Digitized analysis of handwriting and drawing movements in healthy subjects: methods, results and perspectives." *Journal of neuroscience methods*, vol. 90, no. 2, pp. 157–69, aug 1999. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10513598

[199] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching in fixed dimensions," *Proceedings of the 5th ACM-SIAM Sympos. Discrete Algorithms*, vol. 1, no. 212, pp. 573–582, 1994.

[200] J. Furnkranz and G. Widmer, "Incremental Reduced Error Pruning," *International Conference on Machine Learning*, pp. 70–77, 1994. [Online]. Available: http://www.predictionworks.com/Assets/reference/incr-reduced-error.pdf

[201] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and other kernel based learning methods.* New York, NY, USA: Cambridge University Press, 2000, vol. 22, no. 2.

[202] M. C. Fairhurst and M. Erbilek, "Towards the development of user tools for knowledge acquisition in digital document analysis," *Journal of e-Learning and Knowledge Society*, vol. 10, no. 2, pp. 35–52, 2014.

[203] D. Impedovo and G. Pirlo, "Automatic signature verification: The state of the art," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 38, no. 5, pp. 609–635, 2008.

[204] W. H. Tsui, P. Lee, and T. C. Hsiao, "The effect of emotion on keystroke: an experimental study using facial feedback hypothesis," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, vol. 2013, pp. 2870–2873, 2013. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=6610139

[205] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wróbel, "Emotion Recognition and its Application in Software Engineering," *6th International Conference on Human System Interaction*, pp. 532–539, 2013. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=6577877$\delimiter"026E30F$nhttp://kio.eti.pg.gda.pl/publications/files/hsi13{_}emorg{_}preprint.pdf

[206] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statistics Surveys*, vol. 4, pp. 40–79, 2010. [Online]. Available: http://projecteuclid.org/euclid.ssu/1268143839

[207] M. C. C. Fairhurst and M. Da Costa-Abreu, "Exploring emotion prediction from biometric-based keystroke dynamics data using multiagent systems," in *6th International Conference on Imaging for Crime Prevention and Detection (ICDP-15).* London: Institution of Engineering and Technology, 2015, pp. 14 (6 .)–14 (6 .). [Online]. Available: http://digital-library.theiet.org/content/conferences/10.1049/ic.2015.0112

[208] T. G. Dietterich, "Ensemble methods in machine learning," *Multiple Classifier Systems*, vol. 1857, pp. 1–15, 2000.

[209] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Information Fusion*, vol. 16, no. 1, pp. 3–17, 2014.

[210] S. Haykin, "Neural networks-A comprehensive foundation," pp. pp107–116, 1994.

[211] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review*, vol. 65, no. 6, pp. 386–408, 1958.

[212] M. Buhmann, "Radial basis functions," *Acta Numerica 2000*, vol. 9, no. January 2000, pp. 1–38, 2000. [Online]. Available: http://journals.cambridge.org/abstract{_}S0962492900000015

[213] J. Kittler and F. M. Alkoot, "Sum versus vote fusion in multiple classifier systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 1, pp. 110–115, 2003.