Towards Development of Computational Methods for Handwriting Analysis with Focus on Writer Identification, Personality Prediction, and Emotion Detection

Ph.D. Thesis

By

Atta Ur Rahman

CS1904

In partial fulfilment

of the requirements for the degree of

Doctor of Philosophy (Ph.D.) in

Computer Science

Ghulam Ishaq Khan Institute of Engineering Sciences and Technology

Faculty of Computer Science and Engineering

February, 2022

Scholar id-: CS1904 Atta Ur Rahman

The work in this dissertation has been carried out at the Machine Intelligence Research Group (MInG), Faculty of Computer Science and Engineering, **Ghulam Ishaq Khan Institute of Engineering Sciences and Technology (GIKI**), Topi, Pakistan. The research was supported by the **GIK** under the GA-1 Fellowship Program.

Scholar id-: CS1904

List of Publications Extracted from This Work

Accepted/Published

1. A. Rahman, Z. Halim, "A graph-based solution for writer identification from handwritten text," Knowledge and Information Systems, vol. --, pp. --, 2022. [ISSN: 0219-3116, Thomson Reuters JCR 2021, Impact factor 2.822, Springer] - in press [HEC JRS Medallion= Bronze] *Extracted from Chapter 3*

2. A. Rahman, Z. Halim, "Predicting the Big Five Personality Traits from Hand-Written Text Features through Semi-Supervised Generative Adversarial Network," Multimedia Tools and Applications, vol. --, pp. --, 2022. [ISSN: 1380-7501, Thomson Reuters JCR 2021, Impact factor 2.757, Springer, Q2] - in press [HEC JRS Medallion= Bronze] *Extracted from Chapter 5*

3. A. Rahman, Z. Halim, "Identifying Dominant Emotional State Using Handwriting and Drawing Samples by Fusing Features," Applied Intelligence, Vol. 44, No. 03, 2016, pp. 645-664. [ISSN: 0924-669X, Thomson Reuters JCR 2021, Impact factor 5.086, Springer] - in press [HEC JRS Medallion= Bronze] *Extracted from Chapter 6*

4. A. Rahman, A. Tubaishat, F. Al-Obeidat, Z. Halim, M. Tahir and F. Qayum, "Extended ICA and M-CSP with BiLSTM towards improved classification of EEG signals," Soft Computing, vol. --, pp. --, 2022. [ISSN: 1433-7479, Thomson Reuters JCR 2021, Impact factor 3.643, Springer, Q2] - [HEC JRS Medallion= Bronze]. DOI: https://doi.org/10.1007/s00500-022-06847-w. *Extracted from Chapter* 7

Abstract

Handwriting is a familiar and adaptable method for human-machine interaction. However, it varies widely among individuals. Therefore, designing a writer identification system that works effectively for all individuals is challenging. Writer identification models are popular in digital forensics and personal authentication. However, the complex structure, a wide range of writing styles, and dynamic variations in writing have made it a challenging task. Numerous factors, such as age, education, and emotional condition, have an impact on handwriting. Various aspects are explored in forensic practices, including the mechanics of ink accumulation to understand the cultural impact on symbol formation. However, multi-level knowledge is needed to consider when identifying an author from a huge corpus of known samples based on a fragment of handwriting. Different types of methods for recognizing writers have been tested in the past, however, they lack to explore the emotional behavior of writers. This study explores the phenomenon of cognitive functions in connection to handwriting activity and the human characteristics involved in the creation of handwritten characters. Experimental techniques are employed to investigate human factors associated with writing practices and individual handwriting characteristics.

Handwriting attributes such as contours, corners, junctions, and thickness, to name a few are used as key descriptions in biometric detection systems that depict the same illustration as a fingerprint. Several studies exist which interpret handwritten texts as a specific texture image and retrieve textural characteristics for author recognition. It is assumed that to describe handwriting as a texture image is not an optimal choice. In contrast to real images with variable color and depth of knowledge, handwriting images are often black-and-white binary-valued, which offer only response values for strokes while leaving wide blank spaces as a background. As a result, characters' images usually include redundant data as compared to natural images. The structural descriptions of handwritten texts are significantly more intuitive, apparent, and consistent than the textural features. A growing number of researchers have recently focused on structure-based strategies for identifying writers. The structure-based techniques extract characteristics from the contour of handwriting. Graphs are the first option for representing different sized objects in

many pattern recognition tasks. Its capacity to model various components of an object, as well as its potential to investigate relationships, may be utilized to create strong representations. Rather than perceiving handwriting as static images or temporal trajectories, this work proposes that they should be represented as geometric graphs that preserve both spatial and temporal features.

The first section of this dissertation presents a novel graph-based technique for modeling handwriting samples for the task of author identification. Image segmentation is used to distinguish handwritten lines from scanned images of handwriting. A directed weighted graph is constructed for each writing line, which serves as the foundation for analyzing a writer's structural styles. Statistical characteristics are used to extract features from this graph-based description. To enhance accuracy, the ensemble approach is used for classification.

The second section of this dissertation explains the findings of Big Five personality characteristics from a handwritten text sample employing a Semi-supervised Generative Adversarial Network (SGAN). The style in which a text is written reflects an array of meta-information. This study extracts eleven features from handwriting samples and discovers their correlations with personality traits. Furthermore, this study also develops a novel testing corpus that contains Big Five personality ratings and handwritten text of the 173 users.

The third section of this dissertation explains emotion detection from handwriting and drawing samples. The ability to identify emotions via everyday activities such as writing and drawing is beneficial to one's well-being. To understand more about writing and drawing, there is a need to investigate them in the temporal, spectral, and cepstral domains for discovering new insights. Extracting more information will help to improve classification accuracy. This study combines temporal, spectral, and Mel Frequency Cepstral Coefficient (MFCCs) methods to extract features from such activities, and finds its correlation with depression, anxiety, and stress emotional state of the user. Bidirectional Long-Short Term Memory (BiLSTM) network is used to classify the generated features' vectors.

The last section of this dissertation discusses Electroencephalography (EEG) signals classification for stress level detections. The oscillations of EEG signals are

used for classifying human stress. However, these signals have a high temporal resolution and are rapidly distorted with unwanted noise, resulting in a variety of artefacts. This study utilizes an Extended Independent Component Analysis (E-ICA) based approach for artefacts removal. A Multiclass Common Spatial Pattern (M-CSP) based moving window technique is proposed to obtain the most distinguishable time segment of EEG trials.

This work develops the concept of graph-based structural representation of handwritten text. Most cutting-edge techniques for handwriting analysis rely on fixed-scale interpretations, whereas graphs are adjustable in size and allow exhibiting of both local and global handwriting characteristics. They not only describe a varying quantity of information using vertices, but they can also explicitly describe binary relationships between them using edges. A sequence of extracted features is retrieved from handwriting graphs and fed into classifiers for categorization. The proposed framework is tested on various benchmark datasets to validate its performance.

Chapter 1
Introduction

"Handwriting is an imprint of the self on the page"

Dr. Rosemary Sassoon

The science of handwriting analysis is founded on the notion that everyone writes in their own unique style [1]. Every individual's handwriting is unique and distinctive, like their fingerprints. These characteristics of an individual's writing make a distinction between individuals or classes of handwriting [2]. Biometric techniques are divided into two kinds: Physiological biometrics are relying on the direct assessment of various body components such as the iris, retina, fingerprints, and face, etc. Behavioral biometrics are the second category of modalities, and they are predicated on how an individual performs actions such as gait, voice, keystroke, handwriting, and signature, etc [3]. The use of behavioral biometrics is attracting attention and producing encouraging results for the tasks of identification and verification. One of the most effective integrant of behavioral biometrics is handwriting-based writer identification [4]. As a result, two individuals cannot make specifically uniform handwriting, and even a single individual cannot precisely duplicate his own handwriting. The latter is referred to as intra variation, which is a natural divergence observed in individual writing, causes due to many factors [5]. The variation in individuals' handwriting is a distinguishing characteristic, which makes them unique from other writers. Forensic document professionals extract this differentiation extensively with experiments to prove someone's legitimacy. The process of writer identification includes a careful examination of distinguishing features in both the existing writing sample and the sample of an unknown writer [6]. The experts look for various characteristics such as slant, word spacing, letters, uncommon word formats, etc. All such correspondence is taken, and the expert makes a decision based on a thorough examination of a sample.

Handwriting may be used to profile an individual personality through exploring character traits and quantifying moods and emotions. In particular, among several human affective activities, handwriting contains the most information for gaining a