Keyboard Usage Recognition: A Study in Pattern Mining and Prediction in the Context of Impersonation

Thesis submitted in accordance with the requirements of
the University of Liverpool for the degree of Doctor in Philosophy by

Abdullah Alshehri

May 2018

Dedication

To my dearest parents, and to my beloved wife ...

Acknowledgements

My sincere and most profound thanks go first to my primary supervisor, Professor. Frans Coenen. I would like to take this opportunity to convey my gratitude for his support, guidance, and encouragement throughout the time of pursuing my PhD. I would like also to express my warm thanks to my second supervisor, Dr. Danushka Bollegala who has provided me with insightful comments at various stages of my research. Special thanks are also due to my advisors, Dr. Floriana Grasso and Dr. Keith Dures for their constant feedback and constructive assessment of my research.

My profound gratitude is also expressed to my parents, sibling and all my family who have extended to me their moral and emotional support so that I can follow my dreams. I would especially like to express a heartfelt thank you to my wife for standing beside me during this challenging period.

I am also very grateful to my friends and colleagues within the "Data Mining and Machine Learning" (DMML) group at the Department of Computer Science who were always helpful and provided me with their assistance whenever necessary. Not forgetting all staff members of the Department of Computer Science who also facilitated the work of my PhD since day one.

And finally, last but by no means least, I gratefully acknowledge the generous funding received from the Government of Saudi Arabia, especially the Saudi Arabian Cultural Bureau in London, that has enabled me to undertake my PhD.

Abstract

The research presented in this thesis is directed at an investigation into the use of keystroke dynamics (typing patterns) for the purpose of impersonation detection, especially in the context of online assessments. More specifically, the aim was to research the nature of time series analysis approaches for the purpose of continuous user authentication. The research question to be answered was *"Is it possible to continuously authenticate individuals, according to their keyboard usage patterns; and if so what are the most appropriate mechanisms for achieving this?"*.

The main contribution of the thesis is a collection of three time series analysis approaches to continuous user authentication using keystroke dynamics: (i) Once-only Keystroke Continuous Authentication (OKCA), (ii) Iterative Keystroke Continuous Authentication (IKCA) and (iii) Keystroke Continuous Authentication based Spectral Analysis (KCASA). The OKCA approach was a benchmark, proof-of-concept, approach applicable in the static (as opposed to the continuous) context, and directed at establishing the veracity of the time series approach. The IKCA system was the first of two proposed continuous iterative authentication approaches. The IKCA approach was founded on the OKCA approach. A particular novel aspect of the operation of the IKCA approach was that it used the concept of a bespoke similarity threshold. The KCASA approach was then an improvement on the IKCA approach that operated in the spectral domain rather than the temporal domain used in the case of the OKCA, and IKCA approaches. Two spectral transformations were considered: (i) the Discrete Fourier Transform (DFT) and (ii) the Discrete Wavelet Transform (DWT). All three of the proposed approaches used Dynamic Time Warping (DTW) as the time series similarity determination mechanism because this offered advantages over the more standard Euclidean distance similarity measurement.

The systems were evaluated using a dataset collated by the author, and two further datasets taken from the literature. Both Univariate and Multivariate Keystroke Time Series (U-KTS and M-KTS) were considered. The evaluation was conducted to compare the operation of the proposed approaches and to compare the operation of the proposed approaches with the established feature vector-based approach from the literature. All the proposed time series-based approaches were found to be more accurate than the

feature vector-based approach. The most accurate of the three proposed time series-based approaches was found to be the KCASA approach. More specifically, KCASA with DWT coupled with M-KTS. However, DFT was found to be more efficient in terms of run-time complixity.

Contents

Illustrations

List of Figures

List of Tables

List of Algorithms

Notations

The following notations and abbreviations are found throughout this thesis:

CS	Cosine Similarity
DTW	Dynamic Time Warping
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
ED	Euclidean Distance
$\mathbf{F}^t$	Flight time
FAR	False Acceptance Rate
FFT	Fast Fourier Transform
FMR	False Match Rate
FNMR	False Non-Match Rate
FRR	False Acceptance Rate
FVR	Feature Vector Representation
IKCA	Iterative Keystroke Continuous Authentication
KCA	Keystroke Continuous Authentication
KCASA	Keystroke Continuous Authentication based Spectral Analysis
KD	Key Down
KH	Key Hold
KSA	Keystroke Static Authentication
KTS	Keystroke Time Series
KU	Key Up
M-KTS	Multivariate-Keystroke Time Series
MOOCs	Massive Open Online Courses
MRR	Mean Reciprocal Rank
OKCA	Once-only Keystroke Continuous Authentication
U-KTS	Univariate-Keystroke Time Series
WBKTR	Web-Based Keystroke Timestamp Recorder
WD	Wrapping Distance
WP	Wrapping Path

Glossary

Cosine Similarity. A method for finding similarity between two non-zero vectors; it is extensively used in the context of information retrieval, data mining and pattern recognition.

Discrete Fourier Transform. A mechanism for transforming time series data from the temporal domain to the frequency domain by representing the time stream as a linear combination of sinusoidal coefficients.

Discrete Wavelet Transform. An alternative mechanism to the Discrete Fourier Transform for transforming time series data by considering the time span over which different frequencies (coefficients) are present in a time series.

Dynamic Time Warping. A time series similarity checking mechanism that considers the optimal alignment between two time series.

Flight Time. The time between consecutive key presses and releases.

Key-Hold Time. The length of time between a key press and a key release.

Keyboard Event. A point in a point series which is parametrised as a tuple of the form $\langle \mathcal{T}_i, \mathcal{X}_i \rangle$, where $\mathcal{T}$ is an identifying index and $\mathcal{X}$ is one or more keystroke dynamics.

Keystroke Time Series. An ordered discrete sequence of keyboard events.

Multivariate-Keystroke Time Series (M-KTS). A keystroke time series that comprises multiple keyboard events.

Shapelet. A subsequence of a keystroke time series.

Univariate-Keystroke Time Series (U-KTS). A keystroke time series that comprises a single keyboard event.

User Typing Template. A set of shapelets that are known to belong to a particular user.

Chapter 1

Introduction

1.1 Overview

Recent decades have witnessed a massive increase in digital learning, or simply eLearning, and online education methods. Digital learning, whatever form this might take, refers to internet facilitated education, as opposed to traditional face-to-face classroom style education. Many universities and educational institutions now provide online programmes where students can learn outside of the traditional, nine to five, bricks and mortar, setting. For instance, in 2017, the University of Liverpool had more than 10,000 students enrolled in over 30 postgraduate programmes, all delivered fully online [118, 119]. Another example of the prevalence of eLearning is the emergence of the Massive Open Online Courses (MOOCs) phenomena [35]. The term MOOCs was first coined by George Siemens, from the School of Computing and Information Systems at Athabasca University, and Stephen Downes, from Canada's National Research Council, in 2008 [41, 60, 123]. MOOCs are also web-based teaching platforms, but are different to other eLearning approaches in that they are typically not bound by formal regulations or schedules; thus students can register, learn at their own pace and withdraw, openly and freely [23]. The number of students enrolled on MOOCs has been increasing year on year. Figure 1.1 shows a graph of the number of students, year-on-year from 2011, registered with some well known MOOC operators, specifically Coursera [32], Udacity [153] and edX [43] . The increase in student numbers can be clearly observed.

Many eLearning and MOOC providers give students an option, at an additional cost, to register for an award, although many MOOC users take courses solely out of interest, rather than with the aim of acquiring some kind of certification. Nevertheless, where students do want some sort of certification the provider has to have systems in place that allow them to confirm that students have met the learning outcomes of the course they are taking. This is typically done through some form of remote assessment, and thus the provider must correctly authenticate the student's identity before granting any form of certificate (certificates should clearly not be awarded to imposters).

The increasing prevalence of eLearning has meant that user (student) authentication has become a significant issue [48, 97, 105]. Today, the vast majority of eLearning

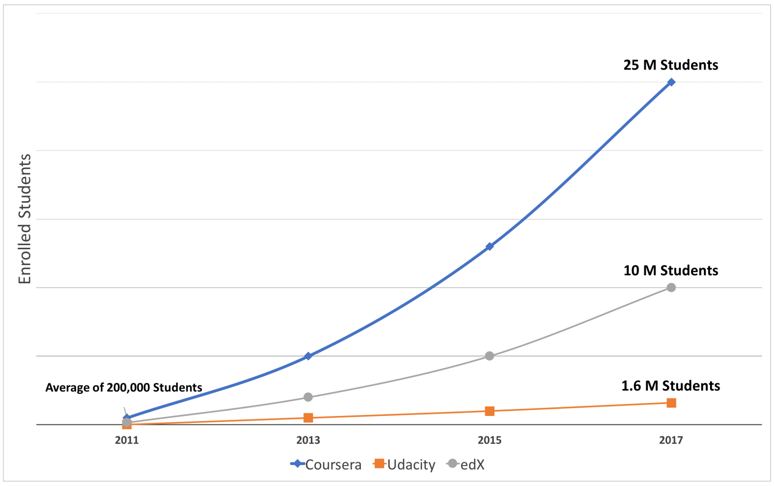


FIGURE 1.1: Well known MOOC providers and the number of students enrolled from 2011 to 2017. The numbers have been taken from the official websites of each provider (Coursera, Udacity and edX) respectively [32, 43, 153].

and MOOC systems depend on (traditional) *log-in* credentials, such as passwords and usernames, for authentication. However, this means that the identity of students is only authenticated at the start of an eLearning assessment [105] and even then usernames and passwords are open to misuse. The utilisation of this form of authentication is obviously inadequate with respect to what is known as "insider attacks". The form of insider attack most relevant to eLearning is impersonation, where an imposter poses as the real user when performing some remote assessment. Consequently, a major issue with respect to the certification of individuals taking courses using eLearning systems and MOOCs is how to continuously confirm that a student taking an assessment is who they say they are. This means, not only that students need to be authenticated at the start of each eLearning assessment, but throughout the course of the assessment; continuous surveillance of student identity is thus required. The work presented in this thesis is motivated by this requirement and seeks to find a suitable solution to the continuous authentication problem in the context of online assessments.

Generally speaking, the process of authentication, with respect to any computerised system, can be categorised (at least at a high level) as:

1. **Token-Based:** The authentication is accomplished using stored credentials previously assigned to the user, such as: a given password, answers to a security question, or magnetic identity cards. This type of authentication process is also sometimes described as *one-off validation* [141], because it is only done once, for example on entering a building or logging in to a computer system.

2. **Biometric-Based:** In contrast to the above, the biometric-based authentication process operates using some feature that is inherent to the user [135]. Biometrics can be classified into following:

(a) **Physiological Biometrics:** Physiological biometrics are the organic characteristics of an individual. Well known examples include: (i) iris recognition [166], (ii) face recognition [126] and iii) fingerprint recognition [98].

(b) **Behavioural Biometrics:** Behavioural biometrics are concerned with the manner in which individuals perform certain tasks. Examples include: (i) keystroke dynamics (typing patterns) [108], (ii) mouse movement usage [4], (iii) voice recognition [65], (iv) handwriting recognition [157] and (v) gait (walking style) recognition [100].

A further distinction between physiological and behavioural biometrics is that the first is typically also used for one-of-verification as in the case of token-based authentication.

In the context of token-based authentication mechanisms, user credentials (password and username) are the most common method used for authentication; this is especially the case in the context of computerised systems. However, because of their one-off nature, password and username are inappropriate for continuous user authentication as required for the purpose of monitoring students taking online assessments. Also, passwords can be easily violated by (say) theft, hacking and "shoulder surfing". Thus, token-based knowledge, using (say) passwords and usernames, is unsuited to addressing the security concerns regarding eLearning and online education assessment methods; more robust methods are required.

Biometrics can produce strong authentication solutions [9]. As already noted, biometric mechanisms authenticate users based on personal traits rather than something they know, or carry around with them; and are thus considered more reliable. Typically, the use of physiological biometrics requires specialised equipment to operate, such as iris, face or fingerprint recognition devices. However, in the context of the eLearning domain, it seems unreasonable to expect online students to purchase such equipment for authentication reasons. Moreover, physiological biometrics are impractical for continuous authentication in that students need to re-conduct the biometric authentication periodically. In contrast, behavioural biometrics, seem well suited to continuous user authentication in the eLearning context, because they do not require dedicated devices which in turn make their deployment relatively straightforward. The most obvious behavioural biometrics of this type is keystroke dynamics (typing patterns). The work presented in this thesis is therefore directed at continuous user authentication, in the context of eLearning environments, using keystroke dynamics.

The intuition behind the use of keystroke dynamics, as will be demonstrated in the next section, is that individuals use keyboards in different manners regardless of what they are typing. Thus such "typing rhythms", captured using keystroke dynamics, can be effectively used to authenticate keyboard users. Although the usage of keystroke dynamics for user authentication, and user identification, is not new (there are some previous reports in the literature) the idea of viewing keystroke dynamics in terms of a data stream and using this for continuous user authentication as proposed in this thesis is, to the best knowledge of the author, entirely new.

1.2 Motivations

From the foregoing, the main motivation for the work presented in this thesis is the need for effective, in terms of accuracy and cost, mechanism to support continuous user authentication in the context of online learning environments. The fundamental idea presented in this thesis is, as already noted above, to use keystroke dynamics for this purpose. The research presented in this thesis was founded on the observation that the way in which individuals use keyboards is distinctive and that keyboard usage displays a pattern that can be detected and associated with individual users. Thus, the suggestion here is that impersonation with respect to online tests and exams can be detected using some kind of keyboard usage recognition software in a way that is not provided by other forms of authentication mechanisms. Furthermore, using keystroke dynamics, the identity of the user can be continuously monitored. The intuition is that typing patterns can be continuously extracted regardless of the text being typed, or the nature of a particular keyboard layout, and that these patterns can be used for continuous authentication purposes. In summary, keystroke dynamics, as a means of continuous authentication, provides the following advantages over other types of authentication:

- Usage of keystroke dynamics provides a continuous authentication mechanism as typing patterns can be continuously extracted and monitored throughout an entire online session.

- It is difficult to mimic the typing style of other users; thus keystroke dynamics are difficult to circumvent [151].

- The usage of keystroke dynamics for authentication is low-cost [71] in that it does not require specialist equipment as in the case of other forms of biometric authentication, such as iris and fingerprint recognition.

- Keystroke dynamics are a transparent and a nonintrusive biometric technology; keystroke authentication systems can operate in the background without the knowledge of the user [34].

To date, most work on authentication systems founded on keystroke dynamics has concentrated on providing an extra level of security with respect to user credentials, such as users typing in passwords and usernames, or pin numbers (see for example [13, 21, 22, 72, 83, 110, 120, 149]). This form of authentication is also known as *static authentication*. Some keystroke dynamic systems have also been proposed that operate using typing patterns extracted from free text, although they have received little attention in the literature (some examples can be found at [3, 38, 54, 107, 142]); these are also known as *continuous authentication* systems (continuous because the text being typed is continuous as opposed to static). For ease of presentation, in this thesis, the concept of static authentication will be referred as to KSA (**K**eystroke **S**tatic **A**uthentication). Similarly, the concept of continuous authentication will be referred as to KCA (**K**eystroke **C**ontinuous **A**utehntication).

The most common existing mechanism for representing keystroke dynamics, regardless of whether KSA or KCA is being considered, is the feature vector representation, where individual feature vectors describe individual typing profiles [18, 70, 108, 167]. Keystroke dynamic feature vectors typically comprise statistical values representing keystroke timing data such as the average and standard deviation of hold times [117], or the interval time of selected frequently occurring keystrokes [38]. Authentication is then conducted by comparing the similarity between stored feature vectors representing reference profiles which are known to belong to a specific user, and a previously unseen profile that is claimed to belong to a particular user. The feature vector approach has met with some success. However, there are some critical disadvantages concerning the usage of the feature vector approach with respect to keystroke continuous authentication. These disadvantages can be summarized as follows:

1. The feature vectors tend to be large; this is especially the case for continuous authentication where a large number of features are required for effective authentication to take place. This also can raise effeciency concerns.

2. The feature vector values are either typing pattern abstractions (for example average hold times) or only a subset of the available dynamics (for example only frequently occurring n-graphs). Thus it can be argued that more accurate authentication can be achieved if more of the data is considered without adversely affect the efficiency.

From the preceding discussion, the work presented in this thesis, therefore explored other mechanisms, taking into account the disadvantages mentioned above, whereby keystroke continuous authentication could be conducted. To this end, usage of time series analysis techniques was proposed as these could take into consideration all keystrokes, rather than some subset as in the case of feature vector based approaches. More specifically, the idea proposed in this thesis is to view keystrokes in terms of press-and-release temporal events such that a series of successive events can be recorded. The intuition was that the time series analysis paradigm could be more readily used to identify suspicious typing behaviour from free text dynamically, and consequently provide an effective and efficient solution to the continuous authentication problem in the context of eLearning.

1.3 Research Question and Related Issues

Given the above motivation, the research question, at which the research presented in this thesis is directed, is as follows:

"Is it possible to continuously authenticate individuals, according to their keyboard usage patterns; and if so what are the most appropriate mechanisms for achieving this?"

The provision of an answer to this research question entailed the resolution of a number of subsidiary questions as follows:

1. How can keyboard usage patterns best be represented in a way that avoids the disadvantages associated with the feature vector representations used to date?
2. Given a collection of patterns, what is the most appropriate mechanism whereby a new pattern can be compared with an existing pattern, for the purpose of user authentication?
3. Given solutions to 1 and 2 above, how do we go about evaluating whether a good solution has been discovered (or not)?
4. What is the most effective process whereby continuous, real-time, authentication can be conducted in the context of online assessment and more generally?
5. In the context of 4 above, how should the process deal with "away from keyboard" events?
6. Can we recognise typing patterns in a way that avoids the knowledge of the text that is being typed by the user so as to avoid data privacy concerns?

1.4 Research Methodology

The high level proposed research methodology, designed to provide an answer to the above research question and associated issues, was to generate an appropriate training set and to then experiment with a number of authentication mechanisms. A commitment had already been made to adopt a time series analysis based approach because there was very little evidence in the research literature suggesting that this had been attempted previously.

To collect training data the idea was to develop an online data collection facility where users were invited to type text in response to questions, the idea being to simulate a "discussion question" style online assessment. The view was that it would be necessary to collect a number of samples from each individual (responding to different questions). Each sample would be anonymised, but each would be allocated an ID number. In this manner keystroke data could be collected. More specifically, the time between consecutive key presses (the flight time), and the key-hold time were extracted. Either would provide a point series for further analysis. Alternatively, both could be used. Figure 1.2 shows a typical keystroke time series, the vertical axis, in this case, represents flight time while the horizontal axis represents keystrokes (keyboard keys being pressed). A commitment was also made to use existing keystroke dynamics data sets from the literature.

The idea was to commence with an initial, "proof of concept", system and test this using the collected time series so as to provide an initial insight into whether it was possible to distinguish between typing patterns represented as time series (the expectancy

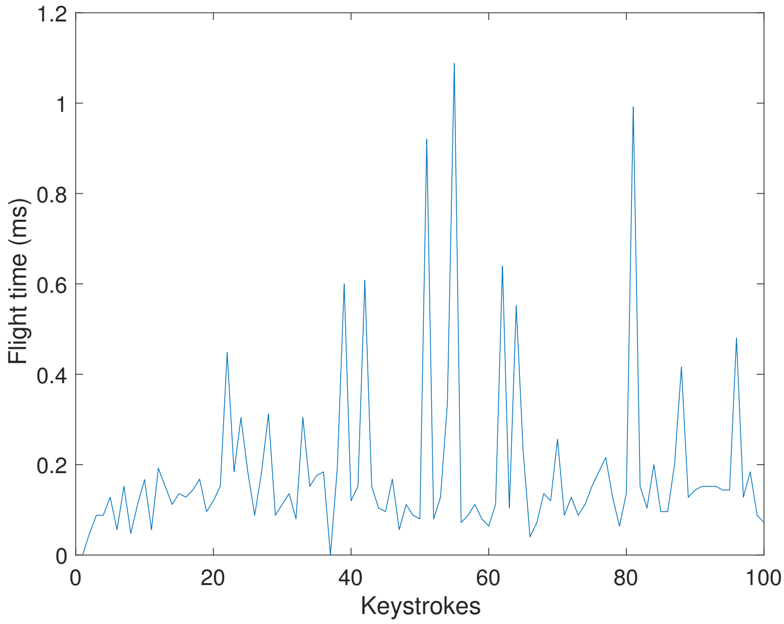


FIGURE 1.2: Typical keystroke time series.

was that this would be the case) and also how much sample text might be required to allow patterns to be distinguished. Once established this benchmark system could then be elaborated on. One idea was to transpose point series from the time domain to the frequency domain and determine the effectiveness of keystroke authentication in this case. To this end, the Discrete Fourier Transform (DFT) and the Discrete Wavelet Transform (DWT) were adopted. Also, a further representation to be considered was using multivariate keystroke time series, instead of univariate keystroke time series, in both the time domain and the frequency domain.

To perform continuous authentication monitoring, the fundamental idea was to experiment with the use of a sliding window, although the nature of this window was unclear at the commencement of the research. By doing so, it was suggested that the proposed windowing approach is able to handle changes in the typing pattern of a subject as typing processes. It was also clear, from the beginning, that whatever techniques were developed the comparison of point series would be required. The most straightforward mechanism for doing this would be to use the average Euclidean Distance (ED) between the points in two given keystroke time series (assuming they were of the same length). However, from a brief exploration of the time series to be considered it was clear that this would be unlikely to be successful because of the way that features in the keystroke time series would be offset against one another. Instead, an approach known as Dynamic Time Warping (DTW) [20] was adopted, where the linearity of time series (possibly of different lengths) was "warped" so that the sequences were aligned, and phase shifting between pairs of time series could be taken into consideration. Thus DTW was considered to be a more reliable form similarity measurement in the case of keystroke time series, as proposed in this thesis, than ED measurement.

The final phase of the research methodology was to evaluate the proposed keystroke continuous authentication mechanisms. Thus, comparisons were undertaken with respect to the feature vector representation, as previously proposed for keystroke continuous authentication. The evaluation was performed to assess the effectiveness, in terms of accuracy and efficiency (run-time complexity), for the proposed methods.

1.5 Contributions

The work presented in this thesis makes a number of contributions with respect to the Computer Science research community, these are summarised as follows:

1. Enhancement of the accuracy of current pattern detection mechanisms for keystroke continuous authentication; reported keyboard dynamics continuous authentication accuracy, to date, has been generally poor [151].

2. Confirmation that keystroke dynamics can be effectively encapsulated in the form of a discrete time series representation.

3. A system for real-time/continuous keystroke authentication that is independent of the keyboard layout, and therefore generic, as required with respect to eLearning and MOOC systems; some of the reported work to date has only been directed at controlled environments using prescribed keyboards,

4. A keystroke modeling technique that is a privacy-preserving in that knowledge of which keys are being pressed is not required.

5. A keyboard authentication mechanism that, although intended for use with respect to online assessment, has general applicability. For example, it may equally well be used to detect certain human conditions, such as detecting keyboard user emotions as described in [132].

6. A keystroke online data capture tool, namely the Web-Based Timestamp Keystroke Recorder (WBTKR), which facilitates the collection of keystroke data sets.

7. A keystroke dynamics data set, collected from real users in an uncontrolled environment, available for public use.

8. Eight competing techniques directed at keystroke dynamics authentication as identified in Figure 1.3:

   i Static authentication using a univariate keystroke time series representation with flight time (a benchmark algorithm).

   ii Static authentication using a multivariate keystroke time series representation with flight time and key-hold time (a benchmark algorithm).

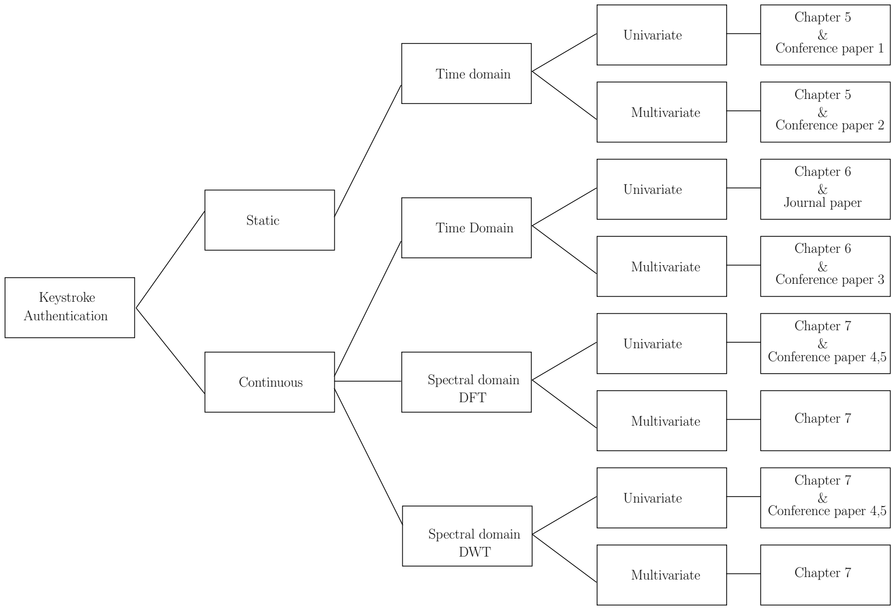


FIGURE 1.3: Categorization of the main keystroke analysis techniques proposed in this thesis.

iii Continuous authentication using a univariate keystroke time series representation with flight time.

iv Continuous authentication using a multivariate keystroke time series representation with flight time and hold time.

v Continuous authentication using DFT spectral (transformed) univariate keystroke time series with flight time.

vi Continuous authentication using DFT spectral (transformed) multivariate keystroke time series with flight time and hold time.

vii Continuous authentication using DWT spectral (transformed) univariate keystroke time series with flight time.

viii Continuous authentication using DWT spectral (transformed) multivariate keystroke time series with flight time and hold time.

The techniques are summarised in diagrammatic form in Figure 1.3. Note that in the figure the chapter numbers in the thesis where the techniques are described are included, as are, where applicable, references to where the techniques have been published.

1.6 Published Work

As noted in Figure 1.3 a number of the contributions presented in this thesis have been published previously in the form of peer-reviewed articles as follows:

Journal papers

1. *Abdullah Alshehri, Frans Coenen and Danushka Bollegala (2018): Iterative Keystroke Continuous Authentication: A Time Series Based Approach. German Journal on Artificial Intelligence (KI- Künstliche Intelligenz), ISSN 1610-1987, DOI 10.1007 /s13218-018-0526-z, pp 1-13. Springer.*

   The paper presents a mechanism for real-time iterative keystroke continuous authentication using sequences of keystroke dynamics and univariate time series. The idea was to continuously monitor the time series for unusual typing behaviour in consecutive time series subsequences obtained using a sliding window and a predefined similarity threshold. The work presented in this paper is included in Chapter 6, where the concept of continuous authentication is introduced. The evaluation conducted in the paper utilised three data sets, the same data sets as used in this thesis as discussed in Chapter 3, and compared the proposed method with a feature vector based approach from the literature. The results from this evaluation, including additional detail, are also considered in Chapter 7.

Conference papers

1. *Abdullah Alshehri, Frans Coenen and Danushka Bollegala (2016): Towards Keystroke Continuous Authentication Using Time Series Analytics, Thirty-sixth SGAI International Conference on Artificial Intelligence (SGAI 2016), pp. 325-339. Springer International Publishing.* **Winner of best student paper prize.**

   This was the first paper resulting from the work presented in this thesis. In this paper, what was described as a novel approach to representing typing behaviour from an arbitrary text, typed in the context of heterogeneous environments, using time series analytics (as conceived of in this thesis), was first introduced. More specifically the idea presented in this paper was to conceptualise the keystroke process in terms of ongoing sequential streams to which authentication processes could be applied. Keystrokes were encapsulated in terms of press-and-release temporal events such that a series of successive events could be recorded. Each keystroke was defined in terms of a pair $P = (t, k)$, where t was a time stamp or temporal identifier of some form; and k was a dimensional keystroke attribute (such as the flight time between keys or key-hold length). The approach was applied in the context of static authentication to determine the effectiveness of the proposed representation. In the context of this thesis, the univariate (single) keystroke time series reported on in this paper is discussed in Chapter 5.

2. *Abdullah Alshehri, Frans Coenen and Danushka Bollegala (2017): Keyboard Usage Authentication using Time Series Analysis, 18th International Conference on Big Data Analytics and Knowledge Discovery (DaWaK 2016), pp.239-252. Springer International Publishing.*

   This paper built on the earlier conference paper listed above by considering keystroke dynamics time series from the multivariate (3D) perspective. The intuition was that be using 3D time series, where both flight time and hold time, are considered more effective authentication could be conducted. The proposed method was evaluated by comparing its operation with that of the univariate method proposed in the previous conference paper. From the evaluation, it was found that the use of the multivariate representation gave better results in terms of authentication accuracy, but not with regard to efficiency. The approach presented in this paper was also considered with respect to static authentication. The concept of multivariate keystroke time series as proposed in this paper is also discussed in Chapters 4 and 5 of the thesis.

3. *Abdullah Alshehri, Frans Coenen and Danushka Bollegala: Accurate Continuous and Non-intrusive User Authentication with Multivariate Keystroke Streaming, 9th International Conference on Knowledge Discovery and Information Retrieval (KDIR 2017.), pp 61-70. SciTePress.* **Short-listed for KDIR 2017 best student paper prize.**

   The content of this paper was an extension of the work presented in the above journal paper; however, the idea of multivariate keystroke time series was used for continuous authentication, as opposed to static authentication. The idea was to use sequences of keystroke timing features in the form of a multivariate time series representation by incorporating flight time and key-hold time features. The process of iterative authentication was then conducted by continuously extracting subsequences from keystroke streams (which are representative of typing patterns), and to monitor the changes in typing behaviour as these subsequences were received. The hypothesis investigated in this paper was the idea that when incorporating more than one keystroke timing feature, in a keyboard time series representation, the quality of the user authentication can be improved significantly. Some of the content of this paper was used with respect to the work presented in Chapter 6.

4. *Abdullah Alshehri, Frans Coenen and Danushka Bollegala (2018): Spectral Analysis of Keystroke Streams: Towards Effective Real-Time and Continuous User Authentication, 4th International Conference on Information Systems Security and Privacy (ICISSP 2018), pp 62-73. SciTePress*

   This paper proposes a mechanism for enhancing the efficiency of the proposed keystroke dynamics time series representation by transforming the raw keystroke time series into some other form. The paper introduced the Keystroke Continuous Authentication based Spectral Analysis (KCASA) technique. The proposed model was motivated by the idea that transforming such time series from the temporal domain to the spectral (frequency) domain might lead to faster, and more accurate, patterns extraction and usage. Two types of the spectral transform are considered in this study: (i) the Discrete Fourier Transformation (DFT) and (ii) the Discrete Wavelet Transform (DWT). Chapters 5 and 6 use material from this paper.

5. *Abdullah Alshehri, Frans Coenen and Danushka Bollegala (2017): Spectral Keyboard Streams: Towards Effective and Continuous Authentication, 12th International Workshop on Spatial and Spatiotemporal Data Mining at the IEEE International Conference on Data Mining (ICDM 2017), pp 242-249. IEEE*

   This paper presented further work on transforming keystroke time series from the time domain to the spectral domain. The paper considered the nature of spectral transformations for keystroke time series and how these might be used to discriminate typing patterns. An evaluation of the proposed approach using a spectral transformation was included in the paper. The content presented in this paper has contributed to the material presented in Chapter 7.

1.7 Thesis Outline

The rest of this thesis is organized as follows:

Chapter 2: Background and Related Work. This chapter presents the background to the work presented in this thesis. The chapter commences with a review of the relevant work with respect to the field of keystroke dynamics in the context of static and continuous authentication. This is followed by a review of the relevant previous work regarding the adopted time series representation and the analysis methods that are pertinent to the work presented in this thesis.

Chapter 3: Evaluation Data Acquisition. Chapter introduces the three data sets used to support the research work presented in this thesis. The first of these was acquired using the Web-Based Keystroke Timestamp Recorder (WBKTR) data collection tool developed by the author, whilst the other two data sets were obtained from the literature. This chapter provides detail concerning each of these data sets and the WBKTR tool.

Chapter 4: Keystroke Time Series Fundamentals and Preliminaries. The chapter provides material associated with the fundamentals and preliminaries of the keystroke time series representation. The chapter also introduces a formal description of Univariate-Keystroke Time Series (U-KTS) and Multivariate-Keystroke Time Series (M-KTS), the main representations in which keystroke streams can be realised. Dynamic Time Warping (DTW), the adopted method for similarity checking with respect to the U-KTS and M-KTS representations, is detailed in this chapter.

Chapter 5: Once-only Keystroke Continuous Authentication. The chapter considers the first keystroke continuous authentication approach presented in this thesis, namely: Once-only Keystroke Continuous Authentication (OKCA). The OKCA approach served as a proof of concept system to establish the utility of the proposed keystroke time series representation with respect to extracting typing patterns from free text.

Chapter 6: Iterative Keystroke Continuous Authentication in Time Series. The chapter introduces the Iterative Keystroke Continuous Authentication (IKCA) approach. The chapter covers different models of the IKCA approach. Furthermore, an extensive evaluation concerning the performance of the proposed IKCA techniques is presented.

Chapter 7: Keystroke Continuous Authentication based spectral Analysis. In this chapter, an alternative approach to continuous keystroke authentication, namely Keystroke Continuous Authentication based Spectral Analysis (KCASA), is introduced. The approach translates the keystroke time series from the temporal domain to the spectral domain. KCASA approach is fully described and detailed in the chapter, including comparisons with the proposed IKCA approach presented in the preceding chapter.

Chapter 8: Conclusion and Future Work. The chapter concludes the work presented in this thesis by summarising the main findings regarding the research question and the associated subsidiary questions. Furthermore, the chapter provides some recommendations and discussion for possible directions for future work.

1.8 Summary

This chapter has introduced the main content of the work presented in this thesis. The introduction has provided the motivation, research question, methodology and contributions of the thesis. Furthermore, the published articles resulted from this work have been briefly reviewed. The following chapter presents the background and the previous related work concerning the material described in subsequent chapters of this thesis.

Chapter 2

Background and Related Work

2.1 Introduction

As noted in Chapter 1, the work conducted in this thesis intersects across three established research areas: (i) user authentication, (ii) keystroke dynamics and (iii) time series analysis. This chapter will thus provide the necessary background concerning each of these research areas with respect to the work presented later in this thesis. The remainder of this chapter is hence constructed as follows. Section 2.2 introduces the fundamentals and methods of user authentication in computerised systems. It also provides an overview of the following topics: (i) static versus continuous authentication, (ii) the difference between authentication and identification and (iii) biometrics-based authentication. In Section 2.3, a comprehensive review of keystroke dynamics is presented. The section includes material on the state-of-the-art of the techniques and methods that have been used to implement typing pattern recognition systems. Section 2.4 provides a general background to time series analysis. The section includes consideration of: (i) time series representation methods, (ii) the process of transforming time series (streaming) data to different domains and (iii) techniques for finding the similarity between time series data. The chapter is concluded with a summary in Section 2.5.

2.2 Authentication Fundamentals

Authentication has become an essential task in the era of computerised systems. Broadly speaking, authentication is defined as the process of verifying the identity of an entity [147], where the entity, in this context, can be either a machine or a human (user). Thus, from the literature, the "state-of-the-art" can be divided into two categories: (i) machine authentication and (ii) user authentication.

The first category describes the process of assuring the validity of a computer's identity when a communication or transaction task is to be performed between two networked computers. This kind of authentication is concerned with the security network protocols used to establish a secure channel for communication. An example of a well-established protocol of this type is the Secure Socket Layer (SSL) protocol [134, 162].

The second category is concerned with authenticating the identity of individuals, more specifically to validate the correctness of a claimed individual's identity [121]. Typically, the individual's identity is verified, using some authentication method, to grant the individual access to some system or location.

The work presented in this thesis is directed at user authentication, this is, therefore, the focus of the rest of this section. In Sub-section 2.2.1 user authentication is considered in more detail. Sub-section 2.2.2 then considers user authentication in the context of token-based knowledge while Sub-section 2.2.3 considers user authentication in the context of biometrics.

2.2.1 User Authentication

User authentication, for the purpose of allowing access to computer systems, has been an issue of concern since the introduction of remote access to computer systems and shared computing facilities. One commonly encountered example, with an established history, is the Automated Teller Machine (ATM). A more recent example, as noted in the introduction to this thesis, is where students are performing online assessments and examinations. A well-known security concern in this later context is impersonation [26, 87]. Whatever the case, user authentication can be conducted in a static or a continuous context, this is discussed further in Sub-section 2.2.1.2 below. Furthermore, there is a distinction between user authentication and user identification; this is thus discussed further in Sub-section 2.2.1.1. For completeness, an overview of the process of static/continuous user authentication/identification is presented in Figure 2.1.

2.2.1.1 User Authentication and User Identification

The work presented in this thesis is directed at user authentication. The term is frequently considered to be synonymous with user identification. However, in this thesis, they are not considered to be the same thing. User authentication refers to the process of determining the validity of a user's claimed identity [152]. The goal is to verify that the user is who they say they are. For example, in the case of a banking scenario, we wish to verify a user's identity using an entered pin number which (it is claimed) belongs to the user. On the other hand, the goal of user identification is to determine who the user is [96, 152], typically with respect to a bank of user profiles (references). For example, in an access control situation we might wish to apply iris recognition to identify the individuals who have access rights, if the obtained iris pattern is not included in a pre-registered database, access can be denied. From a practical perspective user authentication, as defined here, is less resource intensive; only a limited number of comparisons need to be made, while in the case of user identification an entire database has to be searched. Although the focus of this thesis is user authentication (continuous user authentication), user identification is considered in Chapter 5.

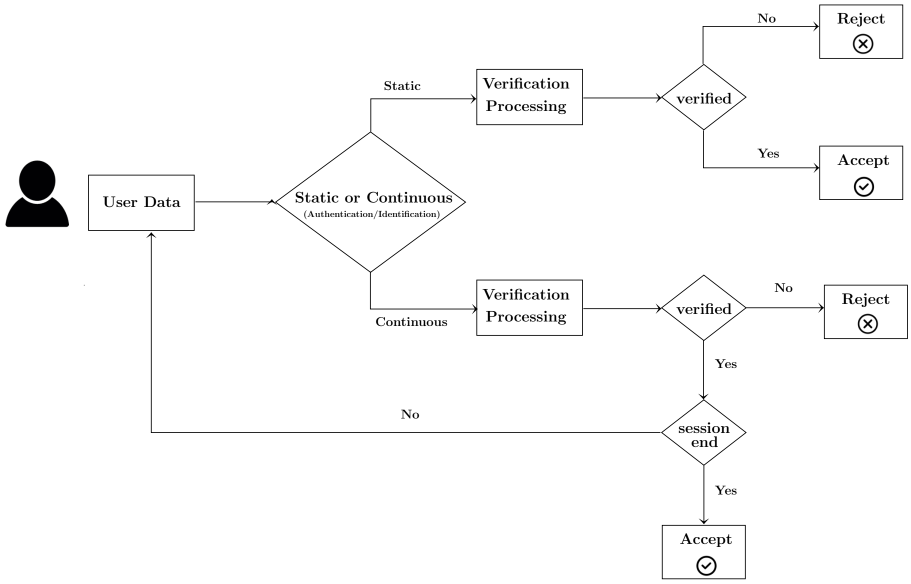


FIGURE 2.1: Data flow diagram illustrating the authentication process.

2.2.1.2 Static Authentication versus Continuous Authentication

User authentication can be conducted in a static or continuous context. Static user authentication is where the user authentication is conducted once only, typically at the beginning of a session, for example when a user requests access to an online banking service or when a user wishes to enter a restricted access location (as in the case of passport control at airports). In the context of internet-facilitated working, static authentication typically takes place when a user first connects ("logs-in"). Whereas continuous authentication is when authentication continues to take place throughout a session. Continuous authentication is clearly a requirement in the context of the online assessments that feature in distance learning (eLearning) environments. Although the central theme of this thesis is directed at continuous authentication (using keystroke dynamics), static authentication is also considered.

2.2.2 User Authentication using Token-Based Knowledge

As already noted in Chapter 1, user authentication can be conducted using either *token-based* or *biometric-based* methods. Token-based methods are discussed further in this sub-section while biometric based methods are discussed further in the following sub-section (Sub-section 2.2.3).

Token-based methods are typically concerned with mechanisms that operate using knowledge that the user knows or possesses, such as user credentials (password and username), PIN codes and magnetic cards [140]. In practice, token-based methods, especially those founded on user credentials, are the most popular form of user authentication method. Note also that token-based methods lend themselves to static authentication, especially remote static user authentication, as in the case of eLearning platforms and online banking systems. The reason that token-based methods are more popular than biometric-based methods is because of their simplicity of usage. Furthermore, token-based methods are non-intrusive; there is no need for particular devices or sophisticated technology.

However, token-based methods feature a significant disadvantage in that they can easily be violated, for instance by (say) hacking and theft. Thus the knowledge used does not guarantee the user is who they say they are, they may be an impersonator. Moreover, the knowledge used can have been obtained by fraudulent means, or have been given by the user to the impersonator. For example in the context of online assessment a student can initially be authenticated by providing appropriate credentials at the start of a session, thus static authentication, even though they are not who they say they are. This is a well-known form of impersonation attack.

2.2.3 User Authentication using Biometrics

Biometrics are dependent on inherent physiological or behavioural attributes from the user [68], such as fingerprints, face recognition, typing patterns, and handwritten signatures. Figure 2.2 shows example of a hierarchical classification of the most common biometric methods.

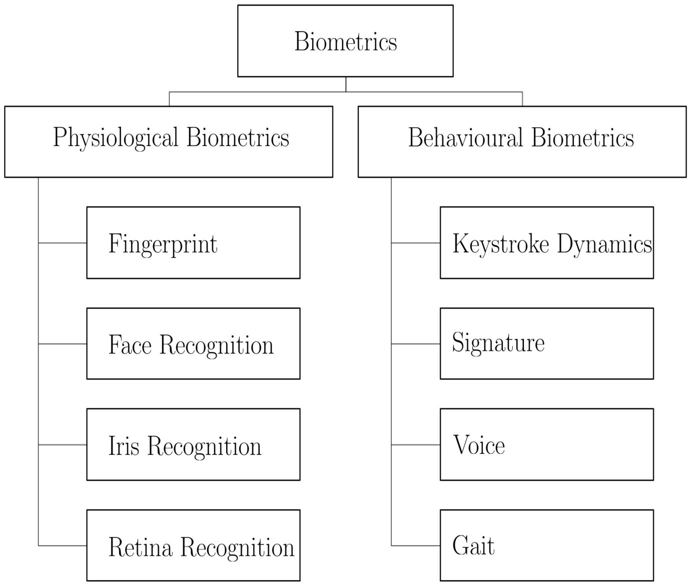


FIGURE 2.2: Example of a hierarchical classification of the most common physiological and behavioural biometric methods.

A biometric authentication mechanism can be defined as a pattern recognition system in which biometric data is first collected from users, after each pattern, features are extracted from the collected biometric data to construct a distinctive template, which can then be compared, for authentication purposes, against previously stored templates in a database [65]. More specifically, biometric user authentication, in general, comprises a two-stage process [36, 39, 154], namely: (i) enrollment and (ii) verification (Figure 2.3). The first is concerned with the creation of an enrollment database, and the second with using that database for authentication or identification purposes. During the enrollment

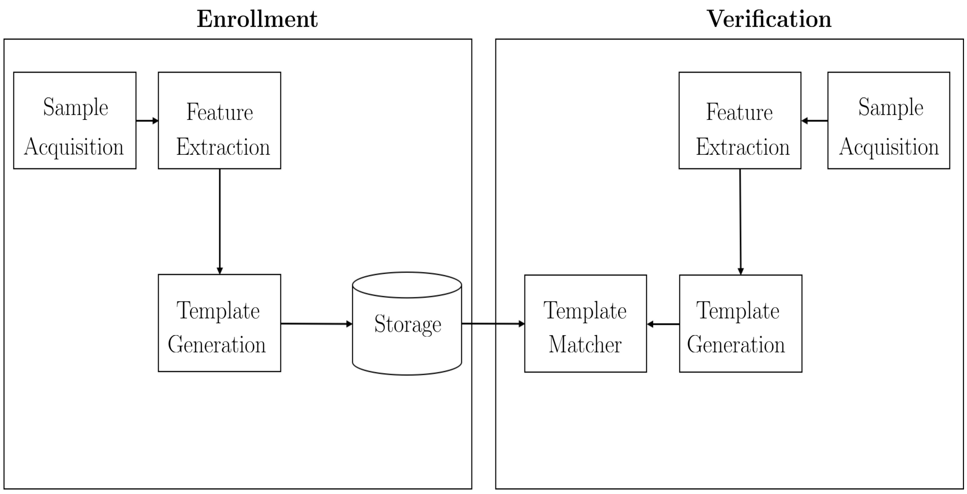


FIGURE 2.3: The schematic of the process of biometric user authentication [148].

stage, biometric data is acquired from real users using devices/sensors, such as fingerprint scanners, cameras, keyboards, voice recording and so on. This data is usually known as the *biometric sample* [39]. Once the biometric data is obtained, features are extracted from this data to represent a distinctive template for each user. All the collected user templates are then stored in a database, referred to as a *biometric database* [64].

During verification (authentication), a newly generated template is compared against the stored templates in the database. The comparison is made so as to authenticate the user. Typically, in the context of user authentication, the primary interests of this thesis, the biometric validation operates by comparing the generated biometric template with the claimed users template(s) stored in the biometric database. Biometrics are unique to individual users and thus provide a safeguard against impersonation [165].

In other words, biometric authentication systems operate as a pattern recognition based binary classification. The classifier is trained using both templates that are known to belong to the user and those that are known not to belong to the user. The resulting classifier can then be used to classify previously unseen templates as $True$ (the user is who they say they are) or $False$ (the user is not who they say they are). The operation of the classifier can be measured in terms of the number of: (i) true accepts (true positives), (ii) true rejects (true negatives), (iii) false accepts (false positives) and iv) false rejects (false negatives). The last two outcomes (false accept and false reject), in the context of biometric user authentication, are referred to as *biometric system errors* [68].

The metric most commonly used for measuring the effectiveness of a biometric user authentication system is the *validity rate*. This is discussed further in the next sub-section.

2.2.3.1 Validity Rate

The validity of a biometric system is measured in terms of the number of false accepts (false positives) and the number of false rejects (false negatives). These values are used to calculate the False Match Rate (FMR), and a False Non-Match Rate (FNMR) [101]:

$$FMR = \frac{false\ positives}{n} \tag{2.1}$$

$$FNMR = \frac{false\ negatives}{n} \tag{2.2}$$

where n is the total number of comparisons such that:

$$n = true\ positives + true\ negatives + false\ positives + false\ negatives$$

Note that the terms False Acceptance Rate (FAR) and False Rejection Rate (FRR) are interchangeably used in the literature to refer to FMR and FNMR respectively [137]. However, the terms FMR and FNMR are used in this thesis. The lower the FMR and FNMR values the more robust a biometric user authentication system is said to be.

2.2.3.2 Biometrics Application Domain

Biometrics, in general, provide strong authentication solutions compared to token-based knowledge methods [66, 67, 128]. Unlike token-based knowledge methods, they are also well suited to continuous authentication [40]. However, which biometrics to use in which situation is very much application dependent. There is no optimal biometric that meets the security requirements of all applications [68]. In the case of the online assessment continuous authentication domain is of interest with respect to this thesis, physiological biometrics may not be the ideal solution because they need a dedicated device which makes its deployment infeasible by the end user. However, typing patterns (keystroke dynamics) seem like an ideal behavioural biometric for online assessment continuous authentication, and this was thus selected for further investigation in the context of the work presented in this thesis. The next section gives a comprehensive review of keystroke dynamics and the techniques, from the literature, used to recognize typing patterns for the user's authentication.

2.3 Typing Pattern and Keystroke Dynamics Approach

The idea of using typing patterns for user authentication can be traced back to the introduction of telegraphy in the 19th century [5, 15, 74, 124] when it was observed that telegraph operators (typing Morse code) had individual typing patterns which could be used to identify individual operators [89]. During the Second World War, there was a mechanism called "`Fist of the Sender`" which was used to identify telegraph operators based on the pace and syncopation of signal taps [39, 103, 111].

Since the development of the computer keyboard, typing patterns have received a considerable amount of attention in the literature as a behavioural biometric which can be used to authenticate individuals. The fundamental approach is to use the detailed timing information of press-and-release movements, known as keystroke dynamics [50]. As indicated in Section 1.2, keystroke dynamics have considerable advantages over other forms of biometrics. The most significant advantage is that keystroke dynamics are an ideal means for providing continuous authentication without requesting specialised equipment, this, in turn, makes them well suited to the continuous authentication of users taking online assessments (such as online exams). The following sub-section considers the distinction between static and continuous keystroke authentication.

2.3.1 Static Versus Continuous Keystroke Authentication

Keystroke dynamics have been studied, as a biometric technology, for different authentication purposes. From the literature (as also noted in Chapter 1), keystroke dynamics have been applied to work in the context of both *static* and *continuous* authentication.

Static authentication using keystroke dynamics, as the name implies, is directed at static (predefined) text such as passwords, usernames, and pin numbers. The goal is to verify the identity of the claimed user at the initiation of a session. This is the most common form of keystroke dynamics biometric authentication. Thus, typing patterns are learned from a fixed text and associated with users for authentication purposes. Examples of applications where static keystroke authentication is used include: log-in strengthening [16, 21, 56], access code verification at physical access points [88, 145], and password sharing prevention [99, 106].

Whereas static keystroke authentication is directed at fixed texts, continuous keystroke authentication is direct at continuous (free) text. Thus, typing patterns are learned from an arbitrary set of texts associated with a specific user and then used to continuously authenticate (monitor) the user. This type of authentication is required with respect to applications where the aim is to monitor a remote user's identity in a continuous manner. It should also be noted here that the term *dynamic authentication* is also used in the literature to describe the process of continuous authentication using keystroke dynamics as described here (see for example [108]).

From the previous, it is conjectured in this thesis that continuous authentication, founded on keystroke dynamics, can be successfully employed for user authentication in the context of online assessments such as those encountered in eLearning and MOOC platforms. However, in this thesis, the terms *"once-only"* and *"iterative"* are also used to describe the nature of the authentication. The term once only is used to refer to "one-time-only" authentication applied once typing has been completed, regardless of whether we are considering keystroke static authentication (fixed text) or keystroke continuous authentication (free text). The term iterative then refers to a process where the authentication is repeatedly conducted whilst typing is taking place, preferably in real-time, as required when monitoring online assessments and examinations. It does not make sense to apply iterative authentication in the context of keystroke static authentication; thus the term iterative authentication is assumed to apply to keystroke continuous authentication only. In Chapter 5, the concept of iterative keystroke continuous authentication will be described in further details. From the literature, however, most existing work on keystroke continuous authentication has been directed at once only authentication, authentication conducted when typing has been completed.

The following subsections provide further details concerning previous work directed at keystroke dynamics systems, with a focus on keystroke continuous authentication as this is a central theme of the work presented in this thesis.

2.3.2 Keystroke User Authentication Workflow

The process of using keystroke dynamics, as an authentication biometric, is similar as to that used with respect to other forms of biometrics. The process commences by collecting keystroke timing information (samples) from typed text, and this is then processed further to extract features and collected keystroke timing information. The extracted features are then used to generate a typing template for each user which can then be used for authentication purposes. Figure 2.4 illustrates the overall keystroke dynamics user authentication workflow. Note that the process maps into the general procedure for biometric verification given in Figure 2.3.

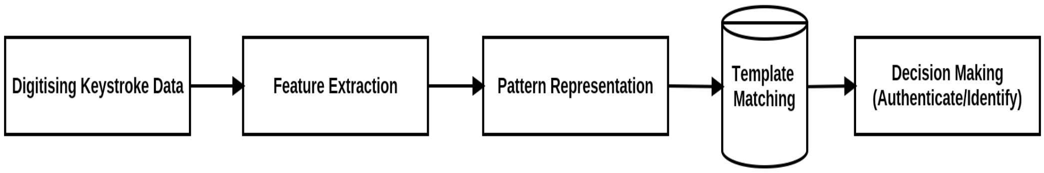


FIGURE 2.4: Keystroke dynamics user authentication workflow.

2.3.3 Keystroke Timing Features

Approaches founded on keystroke dynamics are founded on the temporal patterns generated as a result of sequences of key presses and releases. The features used with respect to keystroke authentication systems are what are known as *timing features*. For each keystroke press, there are four timing values that can be obtained: (i) Key-down time $\mathcal{KD}^t$, (ii) Key-up time $\mathcal{KU}^t$, (iii) Key-hold time $\mathcal{KH}^t$ and (iv) flight-time $\mathcal{F}^t$. For any keystroke i, $\mathcal{KH}^t_i$ can be calculated as $\mathcal{KH}^t_i = \mathcal{KU}^t_i - \mathcal{KD}^t_i$. The value for $\mathcal{F}^t_i$ can then be obtained from $\mathcal{F}^t_i = \mathcal{KD}^t_i - \mathcal{KU}^t_{i-1}$. Since $\mathcal{KH}^t$ and $\mathcal{F}^t$ incorporate other values, these are considered the most important keystroke features. Figure 2.5 illustrates the relationship between the four keystroke timing features.

The nature of the main keystroke features ($\mathcal{F}^t$ and $\mathcal{KH}^t$) can be summarised as follows:

1. **Flight Time**: The time between consecutive key presses and releases.

2. **Key-Hold Time**: The length of time between a key press and a key release.

Note that in the context of flight time, the sequences of key presses are typically referred to as n-graphs, where n is the number of keys pressed; where $n = 1$ it is called

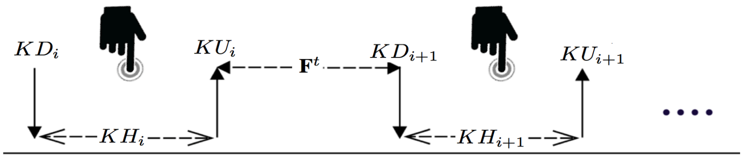


FIGURE 2.5: The basic keystroke timing features.

a monograph, where $n = 2$ it is called a digraph, where $n = 3$ it is a trigraph, and so on. Furthermore, in the literature, flight time has been given various diverse names such as: (i) latency time [72, 106, 136] (thus digraph and trigraph latency is also referred to, see for example [74]), (ii) elapsed time [4, 144] and (iii) transition time [82, 158]. Similarly, key hold time has also been given different names in the literature, such as: (i) dwell time [150, 172] and (ii) duration time [19, 44]. Nevertheless, in this thesis, the terms key-hold time and flight-time are used.

The fundamental idea of keystroke dynamics is to utilise these features to construct a distinctive typing profile (also referred to as a signature or reference profile, the term template is also used) associated with individual users for authentication purposes [50]. The next subsection thus introduces a further overview concerning typing profile construction.

2.3.4 Typing Profile Construction

To date, the most common mechanism used to build individual typing profiles (templates) is by constructing feature vectors based on keystroke timing information. Examples can be found in [14, 55, 72, 139]. Such feature vectors usually comprise statistical values, for example, the average and standard deviation of flight or hold times [58, 109].

In the context of keystroke static authentication, the template is constructed from a fixed text; users are typically requested to type the fixed text several times [6, 52, 53, 83]. Table 2.1 shows examples from the literature concerning fixed text template construction in terms of: (i) the number of participants, (ii) fixed text used and (iii) the number of repetitions.

In the context of keystroke continuous authentication, however, the construction of templates is more challenging. This is because, by definition, the text to be considered is unstructured; we do not know in advance the expected sequence of key presses. This, in turn, means that typing templates need to be more generic, and consequently more sophisticated. A common mechanism for defining feature vector templates, in the context of keystroke continuous authentication, is by identifying statistical details concerning the most frequently occurring sequences of keystrokes, n-graphs (see for example the studies at [38, 107]). For instance, if `ea` and `th` are the most frequently occurring digraphs in typing samples, the typing template can be constructed using the means and standard deviations of the flight times for these digraphs. Consequently, within the verification stage, given a typing sample received by the user, we can search for

Table 2.1: Statistics concerning the construction of templates, in the static authentication context, taken from examples presented in the literature.

Study	Giot *et al* [53]	Killourhy and Maxion [83]	Allen [6]	Giot *et al* [52]
# of users	100	51	104	133
Fixed text	"greyc laboratory"	".tie5Roanl"	"pr7q1z" "Jeffery Allen" "drizzle"	"greyc laboratory"
Repetition	6	50	89-504	5-107

templates whose statistical similarity (matching score) falls within a global predefined threshold; whenever the similarity between two typing samples falls below the threshold, it will be recognised as a real sample, otherwise it will be rejected. This mechanism is the prominent approach used for continuous authentication. However, a criticism that can be directed at this mechanism is that the typing profile might not feature the same frequently occurring n-graphs as the samples to be authenticated, which in turn can lead to poor verification rates. A suggested solution is to increase the number of training n-graphs considered to cover all possibilities; however, this means that the user needs to be asked to provide a lot of samples. An obvious question is how many n-graphs do we require to ensure that a typing template is sufficiently robust? Whatever the answer, the number of n-graphs, and hence the number of required samples, is significant. Further discussion regarding iterative keystroke continuous authentication is presented in the following sub-section.

2.3.5 Keystroke Continuous Authentication

Much existing work on keystroke continuous authentication (and keystroke static authentication) is directed at one time only authentication; there has been very little work directed at iterative keystroke continuous authentication. Some exceptions in the literature, where iterative keystroke continuous authentication has been considered, can be found in [107], [38], [54] and [3].

In [107] a training set was used to generate a feature vector represented typing template repository which was then used to "identify" users (see also [109]). The feature vectors were constructed by computing the flight time means of all digraphs that were featured in the training set. The iterative user identification was then conducted by repeatedly generating "test" feature vectors for a given user, one every minute, and comparing with the stored templates. If a statistically similar match was found this was considered to be an indication of the typer's (user's) identity. For evaluation purposes a kNN approach was used where $k = 1$. Experiments were also conducted using a number of different methods for comparing feature vectors: Euclidean distance, Mahalanobis, probability, and weighted probability. The disadvantage of the approach, however, was the size of the feature vectors (a great number of digraphs were required) and the number of stored templates. To minimise the search complexity, the authors proposed a clustering mechanism, so only the most relevant cluster had to be searched in detail. However, this then meant that re-clustering was required every time a new user was added. The overall reported accuracy, in the context of user identification and iterative KCA, was 23%; not, it is argued here, a good result.

In [38] digraph latency (flight time) was used for the construction of feature vectors. Each feature vector was generated by considering the first 500 digraphs and trigraphs in the input typing sample, and the most frequently occurring 2000 keywords in the English language and determining the associated latency (flight) times. Valid latency times had to be within the range 10ms to 750ms. The mean and Standard Deviation (SD) of each digraph, trigraph and keyword were calculated and n-graphs with SD values in the top and bottom 10% pruned so as to remove n-graphs that had large SD's. During authentication, potential imposter samples were compared with a stored template and an "alert criterion" adjusted accordingly. A deviation (threshold) value was then used to identify imposters. For evaluation of the process, a simulated environment was used. The metric used to measure the performance of the system was the False Match Rate (FMR). Experiments were conducted using digraphs, trigraphs, and keywords; independently and in combination. Best results were obtained using digraphs. The reason trigraphs and keywords did not work well was that the trigraphs did not appear as frequently as digraphs, and many keywords did not appear at all.

The study presented in [54][1] is one of the most promising studies that deal with continuous authentication. In this study, a feature vector representation was again used, however, with respect to the latency values (flight times) associated with the entire shared set of n-graphs included in the evaluation data set. The similarity between two typing samples was determined as follows. The latency times of all shared n-graphs in the two samples were extracted, and ordered (in ascending order of latency time) in two arrays. The difference between the order numbering of each n-graph in each array, the counterpart distance d, was then computed and summed to give a *degree of disorder* value[2]. This was then used as a similarity measure, the smaller the degree of disorder the more similar the two typing samples. The process is illustrated in Figure 2.6 (taken from [54]). The figure shows five digraphs, `ic-he-th-ti-ca`, that feature across two typing samples E_1 and E_2. The digraphs are ordered according to latency time. The associated counterpart distances are then $\{2, 0, 2, 3, 1\}$ respectively. The similarity, sim, between the two samples can then be computed as:

$$sim(E1, E2) = \frac{\sum_{i=1}^{n} d_i}{(n^2 - 1)/2}$$

where n is the number of shared digraphs (five in the example). In the study, the authentication process was evaluated by comparing a new sample to a collection of samples belonging to each enrolled users. For each user, an average degree of disorder value was obtained and the user with the lowest average value selected. This was a

[1]This study has received considerable attention in the literature in terms of the citation.

[2]An idea inspired by Spearman's rank correlation coefficient.

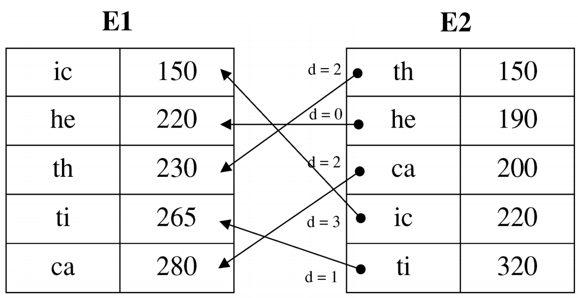


FIGURE 2.6: An example showing how similarity is computed using "degree of disorder" as proposed in [54] (the figure has been taken from the original study [54]).

computationally expensive process given an enrolment database of any size. In the reported evaluation, 600 reference templates were considered (generated from 40 users, each supplying 15 samples one of which was used as a previously unseen sample); the time taken for a single match was thus substantial. Moreover, each typing sample comprised between $700 - 900$ of keystrokes, so the average template size of each user consisted of $11,200$ $(14 \times 800 = 11200)$ keystrokes. Another disadvantage of the approach was the need for a large number of n-graphs so that accurate authentication results could be obtained.

In [3] a mechanism was proposed to mitigate against the expense (in terms of time and size) of pattern extraction and template construction for iterative keystroke continuous authentication. The idea was to use an Artificial Neural Network (ANN) to predict missing n-graphs based on the available data that subjects provided. A feature vector representation was again used. The features used were key-down time and average digraph and monograph flight time. The ANN classifier was then used to build a prediction model with which to conduct user authentication. This mechanism worked reasonably well in a controlled experimental setting; typing of the same text using the same keyboard layout in an allocated environment. However, this is not the situation that will be encountered in the context of real-world authentication tasks (such as in the case of online assessment user authentication), where iterative keystroke continuous authentication is expected to operate.

From the above, it can be observed that most existing iterative keystroke continuous authentication studies have been directed at the usage of quantitative statistical measures to represent keystroke dynamics which have then been encapsulated in a feature vector format. However, feature vector representation may not be the most appropriate representation for iterative keystroke continuous authentication due to the disadvantages detailed above (also summarised in Section 1.2). The central drawback, in this context, is that feature vector based approaches depend on how the feature vectors are constructed. In other words, feature vector based typing patterns are typically constructed by calculating some statistical information of the most frequently occurring n-graphs in a set of typing samples (training data) belonging to a collection of authors where the pairings between individual authors and samples were known. Thus the training data might not feature the same frequently occurring n-graphs as the samples to be authenticated. Of course, a suggested solution to this criticism is to increase the amount of training data available, but this clearly has limitations in terms of the increased size of the vectors that can consequently adversely affect the performance of authentication.

Intuitively, keystroke data forms a natural temporal sequence made up of press-and-release keyboard actions (events). This series of events can provide desirable discriminative typing patterns. Therefore, in this thesis, it is conjectured that representing all keystroke features (events), rather than those associated with specific digraphs, as a time series can lead to a better understating of typing patterns. To the best knowledge of the author, there is no prior work in the literature that has considered the concept of a time series representation for iterative keystroke continuous authentication. Time series analysis is therefore considered in further detail in the following section.

2.4 Time Series Analysis

In this section, an overview of time series analysis, and time series similarity measures, will be presented. The section is structured as follows. In Subsection 2.4.1 methods for representing time series data are considered. This is followed, in Subsection 2.4.2, with consideration of the techniques and mechanisms that can be adopted for measuring the similarity between time series. Of particular note in this context is Dynamic Time Warping (DTW), a well known time series comparison mechanism that was adopted with respect to the proposed time series based continuous authentication mechanisms presented later in this thesis. This is therefore discussed in further detail in Subsection 2.4.3.

2.4.1 Time Series Representation

A time series is an accumulation of observations which occur in a sequential (chronological) manner. Time series data is ubiquitous, it appears across many domains, including science, medicine, and finance [76, 133]. The prevalence of time series data has attracted researchers and practitioners to develop a great variety of methods to mine and analyse time series data. Examples can be found in the context of stock market data [84], anomaly detection in data streams [80], mining shapelets in medical images [46] and the analysis of electrocardiograms (ECGs) [63].

The majority of existing studies directed at time series assume that the time dimension is discrete or can be discretised [164]. Without loss of generality, the same assumption has been made with respect to the work presented in this thesis. Generally speaking, a time series T can be described as a sequence of pairs such that:

$$T = [(t_1, p_1), (t_2, p_2, \ldots, (t_i, p_i), \ldots, (t_n, p_n)]$$

where p_i is a data point in a d-dimensional data space, and t_i is a time stamp associated with data point p_i [164]. Note that given two-time series with the same sampling rate, the time stamp t_i can be omitted; in other words the time series can be considered as a simple sequence of data points:

$$T = [p_1, p_2, \ldots, p_i, \ldots, p_n]$$

The sequence T is then called the raw representation of the time series (in anticipation of further processing). Note that n refers to the number of points in T, or simply the length of the time series T.

One of the primary concerns of time series analysis is how they can best be represented so that they can be processed with respect to some end goal. For short time series this is not usually an issue, but for long time series processing can be expensive [164]. One solution is to recast the time series of interest into a form which will allow efficient processing while at the same time maintaining effectiveness.

In the literature, there are many methods for representing time series other than in their raw (default) format, these including: (i) sampling methods [10, 170], (ii) Piecewise Aggregate Approximation (PAA) [77, 78], (iii) symbolic methods [8, 92, 93, 112, 122, 168] and (iv) spectral (transformed) methods [2, 17, 73, 94, 131].

With respect to the work presented in this thesis two categories of representation are considered: (i) time domain representations and (ii) transformed representations. The former is concerned with representing keystroke timing features in their raw format, the default representation, and then using a sliding window approach to detect typing patterns. Although this is the default time series representation, as far as the author is aware, this representation has not been considered previously concerning continuous (free text) iterative keystroke user authentication. The second considered representation involves transforming time series from the temporal domain to the spectral domain.

The usage of the raw data representation is an obvious start point for user authentication using keystroke dynamics as it allows for a straightforward representation. With respect to the transformed representation two well-known methods from the literature were adopted: (i) Discrete Fourier Transform and (ii) Discrete Wavelet Transform. These are therefore considered in further detail in the following two subsections.

2.4.1.1 Discrete Fourier Transform

The Discrete Fourier Transform (DFT) has been widely adopted with respect to time series data of all kinds (see for example [2, 159]). Typically, DFT is used to transform time series data from the temporal domain to the frequency domain. The idea is that this will then allow comparisons of time series more efficiently (than if a transformation had not been conducted) without losing any salient information. The transformation is performed by representing the time stream as a linear combination of sinusoidal coefficients. The similarity between the transformed coefficients for any pair of corresponding signals can then be computed for time series comparison purposes.

Given a time series $T = \{p_1, p_2, \ldots, p_n\}$, where n is the length of the time series, The DFT transform compresses T into a linear set of sinusoidal functions with amplitudes p, q and phase w:

$$T = \sum_{i=1}^{n} (p_i Cos(2\pi w_k p_i) + q_i Sin(2\pi w_i p_i)) \tag{2.3}$$

The time complexity for transforming T is $\mathcal{O}(n \ \mathtt{log} \ n)$ using the radix 2 DFT algorithm [31, 69].

Further detail concerning the proposed iterative user authentication process, using the DFT, is given in Chapter 5.

2.4.1.2 Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) is an alternative form of time series representation, to the DFT described above, that considers the time span over which different frequencies are present in a time series. DWT is sometimes claimed to provide a better transformation than DFT in that it retains more information [28]. DWT can be applied to time series according to different scales, orthogonal [57] and nonorthogonal [49]. With respect to the work proposed in this thesis, an orthogonal scale was used for the DWT; more specifically the well known Haar transform was adopted [57] as described in [28].

Fundamentally, a Haar wavelet is simply a sequence of functions which together form a wavelet comprised of a series of square shapes. The Haar transform is considered to be the simplest form of DWT; however, it has been shown to offer advantages with respect to time series analysis where the time series feature suddenly changes. The transformation is usually described in terms of Equation 7.3 where x is a time series point. The time complexity for transforming a time series T, using the Haar transform is $\mathcal{O}(n)$.

$$\phi(x) = \begin{cases} 1 & \text{if} \quad 0 < t < \frac{1}{2} \\ -1 & \text{if} \quad \frac{1}{2} < t < 1 \\ 0 & \quad \text{otherwise} \end{cases} \tag{2.4}$$

Chapter 7 presents further detail concerning the usage of the Haar DWT transformation in the context of the proposed user authentication mechanisms considered in this thesis.

2.4.2 Similarity Measures for Time Series Data

Time series similarity is a fundamental task with respect to time series analysis. The similarity is ideally considered in the context of an entire time series; exact matches are rare [47]. In other words, the optimal matching of two corresponding time series is conducted in an approximate manner. Figure 2.7 illustrates the intuition of conducting time series similarity whereby, given a set of time series $T = \{T_1, T_2, T_3, T_4\}$ and a

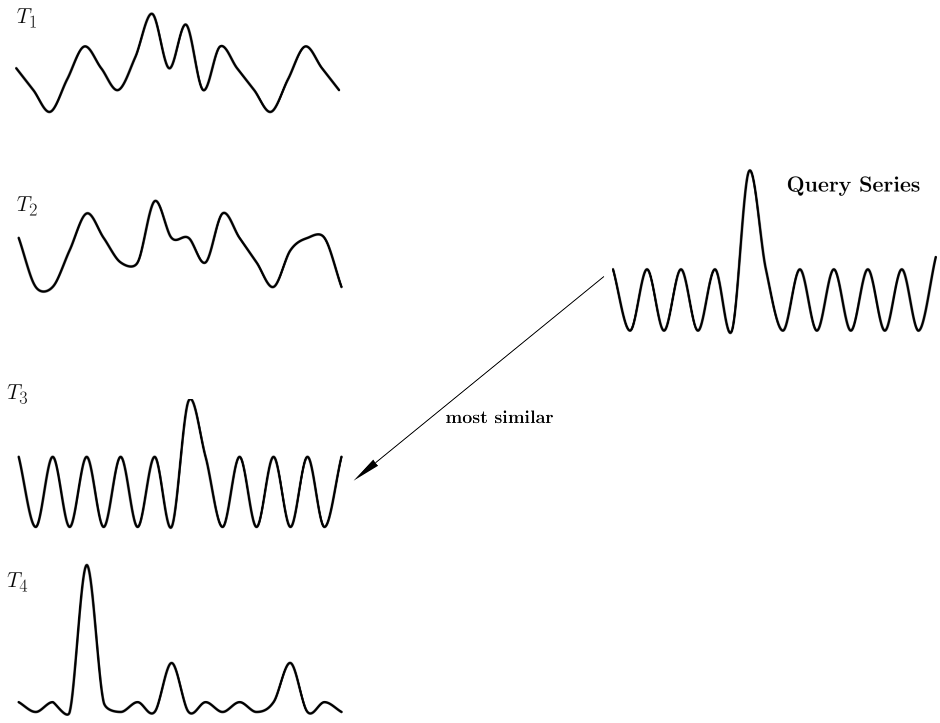


FIGURE 2.7: Time series similarity.

query series, its similarity is considered in terms of the most similar, "closest fit", time series in T.

The different approaches to finding the similarity between two time series can be categorised in terms of: (i) finding the similarity in time (correlation-based), (ii) finding the similarity in shape (shape-based) and (iii) finding the similarity in change (autocorrelation-based) [79, 95, 104]. Which is the most appropriate is application dependent. For example, if the goal is to predict the stock market trend in the next quarter of the financial year, it is intuitive to work with an autocorrelation-based time series similarity approach. With respect to the work presented in this thesis shape-based similarity is used for reasons explained in Chapters 4 and 5.

From the literature, two categories of distance measure, for determining the similarity between two-time series, can be identified: (i) Lock-step Measures, and (ii) Elastic Measures. Each is considered in further details in Sub-sections 2.4.2.1 and 2.4.2.2 below.

An alternative categorisation of time series similarity measurement techniques is in terms of Entire sequence matching and Subsequence matching [2]. Thus, Sub-sections 2.4.2.3 and 2.4.2.4 consider these categories in further detail.

2.4.2.1 Lock-step Measures

Lock-step measures, also known as *one-to-one* comparison point measures, are calculated as follows. Given two time series sequences $\mathcal{S}_1 = \{p_1, p_2, \ldots, p_i\}$ and $\mathcal{S}_2 = \{q_1, q_2, \ldots, q_j\}$, where i is the length of $\mathcal{S}_1$, j is the length of $\mathcal{S}_2$ and $(i = j)$, a lock-step similarity function Sim will compute the distance between each point i in $\mathcal{S}_1$ with the corresponding point j in $\mathcal{S}_2$. The simplest method to calculate this kind of distance measure is by using Euclidean distance:

$$Sim(\mathcal{S}_1, \mathcal{S}_2) = \sqrt{\sum_{i=1}^{|\mathcal{S}|}(p_i - q_i)^2}$$

An alternative might be Manhattan distance [29] or L_p norms [30]. Figure 2.8 illustrates the Lock-step time series similarity measurement approach.

An advantage of Lock-step methods is that they can linearly measure similarity, so the complexity of finding the similarity between two-time series is $\mathcal{O}(n)$, where n is the length of the time series. Furthermore, they are parameter-free and thus have generic applicability [164]. However, Lock-step methods are susceptible to time shifting in the time series as the mapping between the points in the two time series is one-to-one. In other words, local time-shifting, where the shape of the series can imply similarity, cannot be captured by using these kinds of methods. A further disadvantage is that the two time series to be compared have to be of the same length.

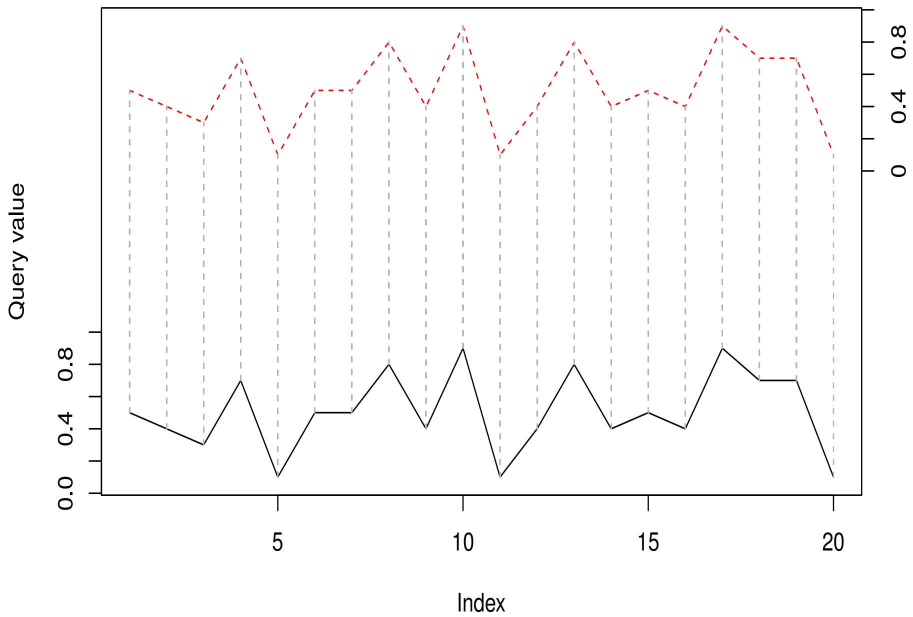


FIGURE 2.8: Lock-step time series similarity measurement.

2.4.2.2 Elastic Measures

Elastic Measures allow for more flexible similarity measurement, between two given time series, than in the case of Lock-step measurement. These methods operate by stretching or compressing the given time series so as to provide a superior similarity measurement. Typically, the comparison between time series points is computed as a *one-to-many* mapping. Figure 2.9 shows a conceptualisation of the Elastic measure approach. From the figure, it can be observed that similar parts (sub-sequences) of the two-time series are shifted so as to capture similarity in shapes between the parts. Thus, Elastic measures can be said to be best suited to shape-based similarity mechanisms; the category adopted with respect to the work presented in this thesis.

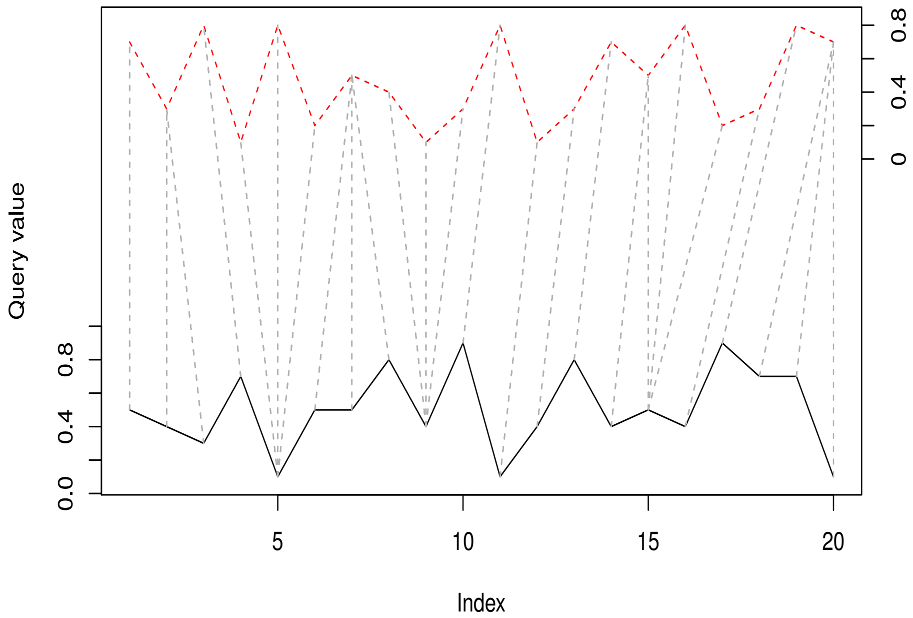


FIGURE 2.9: A conceptualisation of the Elastic measure approach.

However, Elastic measures are quadratic in nature, thus the complexity of finding the similarity between two time series is given by $\mathcal{O}(n)^2$; assuming the two time series have the same length n, although Elastic Measures can work well in the context of time series with different lengths (unlike in the case of Lock-step measures). A well-known Elastic measure is the *Warping Distance* measure obtained using Dynamic Time Warping (DTW); this is the approach adopted with respect to the work presented in this thesis. Thus, further details concerning DTW are given Sub-section 2.4.3.

2.4.2.3 Entire Sequence Matching

Entire sequence matching is directed at determining time series similarity by considering the two given time series in their entirety. However, the utilization of the entire time series in its raw format may be computationally expensive. Therefore, it has been suggested that the best way to proceed is to first transform the given time series into a different domain, such as DFT or DWT, so the dimensionality of time series can be reduced. The most frequently used mechanism for measuring similarity, in this context, is the Euclidean distance measure. The advantage offered is simplicity and hence efficiency in comparison with alternative methods such as elastic methods. However, an issue with the Euclidean distance measure is that it does not take into account offsets (shifts) in the data whereas elastic methods can address this [75, 102, 125]. For example, in the case of ECG time series the characteristics of the time series are frequently unaligned (offset).

2.4.2.4 Subsequence Matching

Using Subsequence matching, a short query sequence is compared with a much longer time series. The idea is to determine whether the query sequence is contained somewhere within the larger sequence. In [45] a subsequence matching time series technique was proposed whereby a time series is mapped into a small set of multidimensional rectangles in a feature space. These rectangles can then be indexed using well known spatial methods such as R-trees. In other words, a sliding window is used to pass over the time series and rectangles extracted. Intuitively, this kind of mechanism allows for fast extraction of patterns in time series data; a similar approach has been adopted with regards to the work presented in this thesis.

2.4.3 Dynamic Time Warping

The Dynamic Time Warping (DTW) time series similarity checking mechanism is a well-established method, which has been used effectively to find the similarity between pairs of point (time) series. It has been adopted in many domains such as speech recognition [130, 171], time series data mining [1, 11, 37], pattern recognition [20, 164], motion and music data analysis [114], medicine [25, 127], robotics [155], biometrics [7, 12, 86, 115], user interface analysis [85, 129, 156], the study of historical manuscript data [61] and biology [81]. A great advantage of DTW is that it serves to warp the linearity of sequences (even of different lengths) so that any phase shifting can be taken into consideration. Thus it can be usefully adopted to find similarity in shape between two corresponding keyboard dynamics time series.

In more detail, the central idea of DWT is to compare two time series by finding the optimal alignment between them in which the similarity can be identified accordingly. More specifically, the operation of DTW can best be described by considering two time series (say) $T_1 = \{a_1, a_2, \ldots, a_i, \ldots, a_x\}$, of length $x \in \mathbb{N}$, and $T_2 = \{b_1, b_2, \ldots, b_j, \ldots, b_y\}$, of length $x \in \mathbb{N}$. A matrix of size $(x) \times (y)$, i.e. $\boldsymbol{M} \in \mathbb{R}^{x \times y}$, is then constructed whereby the value held at each cell $m_{i,j} \in \boldsymbol{M}$ is a local cost measure (c) from point $a_i \in T_1$ to point $b_j \in T_2$. Thus, the matrix, $\boldsymbol{M}$, is utilised to determine the optimal alignment by finding the optimal *Warping Path-WP*, $(\mathbb{P})$, from cell $m\langle 0, 0\rangle$ to cell $m\langle x, y\rangle$. In this context, a warping path, $\mathbb{P} = \{k_1, \ldots, k_n\}$ where $k_n = c(a_n, b_n) \in [1 : x] \times [1 : y]$, needs to satisfy the following constraints [113]:

- **Bounding Constraint**: As indicated earlier, the warping path needs to start from the location $\langle 1, 1\rangle$ to the location $\langle x, y\rangle$ in the matrix $\boldsymbol{M} \in \mathbb{R}^{x\times y}$, such that $k_1 = c(a_1, b_1)$ and $k_n = c(a_x, b_y)$ where $k_1, k_n \in \mathbb{P}$.

- **Monotonicity Constraint**: The elements of the warping path are required to be ordered according to time such that $a_1 \leq a_2 \leq \cdots \leq a_n$ and $b_1 \leq b_2 \cdots \leq b_n$.

- **Continuity Constraint**: The warping path moves in a forward manner to the neighbouring location such that given $k_n = m\langle i, j\rangle$ the follow on location is either $m\langle i + 1, j\rangle$, $m\langle i, j + 1\rangle$ or $m\langle i + 1, j + 1\rangle$.

For further explanation concerning these constraints Figure 2.10 gives examples of warping paths extracted from two time series $T_1 = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and $T_2 = \{1, 2, 3, 4, 5, 6, 7\}$ (the example has been taken from [113]). The figure shows four subfigures as follows: (**a**) shows the warping path that satisfies all constrains, (**b**) shows the warping path where the first constraint is discarded, (**c**) depicts the warping path where the second constraint is avoided and (**d**) illustrates the warping path where the third constraint is ignored.

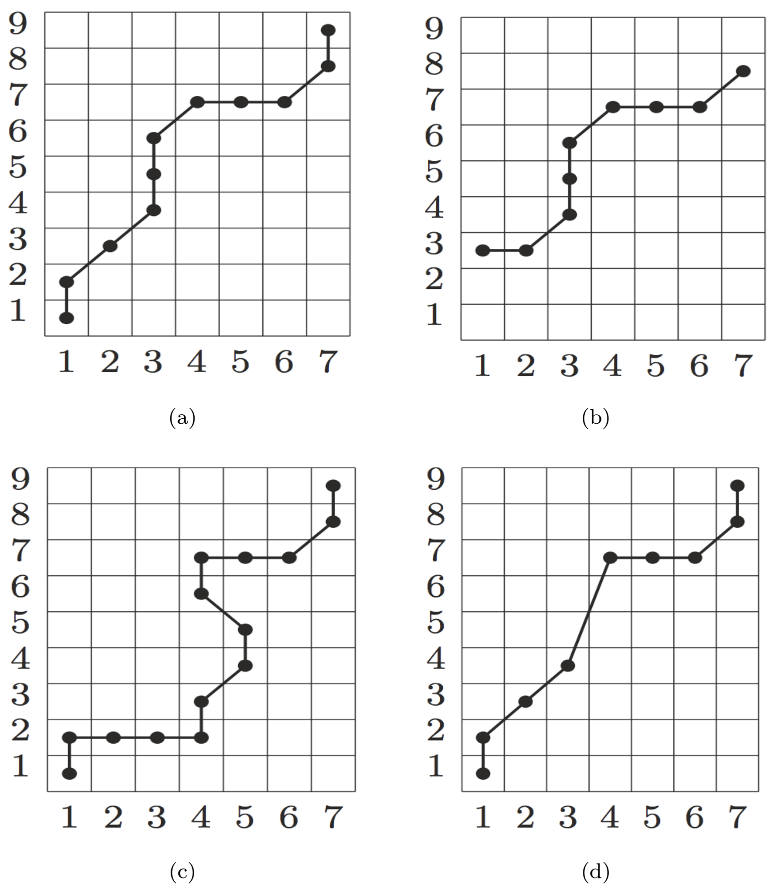


Figure 2.10: Example of warping path constrains in DTW [113].

The warping path is then used to determine a minimum *Warping Distance-WD*, (Θ), which is consequently used as a similarity measure. The warping distance, Θ, is the accumulated sum of the values associated with $\mathbb{P}$ such that:

$$\Theta = \sum_{n=1}^{|\mathbb{P}|} k_n \in \mathbb{P} \tag{2.5}$$

It is worth noting here that the DTW concept has similarities with Levenshtein Distance calculation [90], also referred to as Edit Distance, used for measuring the similarity between two strings. However, in this case, the values used are the number of deletions, insertions or substitutions required to transform the first string into the other.

2.5 Summary

This chapter has presented the background to the principal areas of research covered in this thesis. Three research areas were identified as being pertinent to the presented work: (i) user authentication, (ii) keystroke dynamics and (iii) time series analysis. Each of which was considered in further detail and its relevance highlighted with respect to the central research question of the thesis.

The chapter commenced with a discussion of the concept of user authentication. The distinctions between user authentication and user identification, and static and continuous authentication, was highlighted. A number of authentication schemes were considered categorised as being either Token-based or Biometric-based.

Keystroke dynamics were considered next and discussed in detail, The state-of-the-art was reviewed and current developments highlighted. The chapter also reviewed the techniques most appropriate to continuous (free text) authentication and for the extraction and representation of typing patterns. It was noted that keystroke dynamics user authentication presented in the thesis can be categorised as iterative continuous user authentication.

The chapter was concluded with a discussion of time series analysis techniques, relevant because the mechanisms proposed in the thesis consider keystroke dynamics in terms of time series. The discussion including consideration of the methods that may be adopted for representing time series and time series similarity measures. The well-known Dynamic Time Warping (DTW) time series similarity checking methods, utilised extensively with respect to work presented later in the thesis, was presented in detail.

The following chapter introduces the evaluation data sets utilised with respect to the evaluations reported on later in the thesis, including the process whereby the data was acquired in each case.

Chapter 3

Evaluation Data Acquisition

3.1 Introduction

A fundamental precursor to any form of machine learning is the acquisition of data sets from which some desired model can be learned and evaluated. This chapter presents an extensive overview of the data sets used in this thesis.

Generally speaking, for many machine learning application well-documented benchmark training/test data sets are available. The University of California, Irving (UCI) machine learning repository [91] is a commonly used source for benchmark data sets. However, in the case of keystroke user authentication there are no widely recognised benchmark data sets and very few publicly available data sets. Two data sets that are publicly available, although not sufficiently widely used to merit the adjective "benchmark", are: (i) the GP (Gunetti and Picardi) data set from The University of Torino, Italy [54] and (ii) the VHHS (Vural, Huang, Hou and Schuckers) data set from Clarkson University, USA [161].

The GP data set comprised keystroke data from 31 subjects where the VHHS data set comprised keystroke data from 39 subjects. Although, both data sets could be usefully employed to evaluate the work presented in the thesis it was felt that this would not be sufficient to support extensive evaluation of the techniques proposed in this thesis. Consequently, it was decided to collate a third evaluation data set, the University of Liverpool, ACB (Alshehri, Coenen and Bollegala) data set, so as to provide for a more extensive evaluation. The ACB data set comprised 30 records. This was constructed using a bespoke tool developed by the author namely "Web-Based Timestamp Keystroke Recorder (WBKTR)".

The rest of this chapter is organised as follows. Section 3.2 presents a discussion of the factors to be taken into consideration when generating a keystroke dynamics data set. Section 3.3 then presents the WBKTR tool used to collect the ACB data set and details concerning the make up of this data set. The nature of the GP and VHHS data sets is then discussed in the following two sections, Section 3.4 and Section 3.5. Note that Sections 3.3, 3.4 and 3.5 are structured in an identical manner. Section 3.6 then provides a summary, for reference purposes, of the three evaluation data sets, whilst Section 3.7 concludes this chapter.

3.2 Principles of Keystroke Dynamics Evaluation Data Collection

As noted earlier, in its most basic form, keystroke dynamics data is founded on the timestamp of each key press and release. This section presents a discussion of the factors to be taken into consideration when wishing to generate a keystroke dynamics data set. However, the overriding consideration is the application domain in which the data set is to be used. In the case of the work presented in this thesis, this is subject authentication in the context of subjects typing continuous text in a manner analogous to subjects completing online written assessments.

In [15] two factors are identified that need to be taken into consideration when gathering keystroke dynamics data: (i) the nature of the text to be typed, and (ii) the environment in which the typing is to take place.

The first is concerned with whether the typing samples should be structured (fixed/predetermined) or unstructured (free/arbitrary) text. In the context of the structured text, the text can be obtained as being either short or long. Short texts in this context are things like subject credentials (password and username), while long texts include phrases that are copied from a book (see for example [50]) or from a speech (see for example [161]). In contrast, unstructured text is any text of an arbitrary nature. With respect to the work presented in this thesis unstructured evaluation training/test data was required.

The second factor is concerned with whether the typing samples are collected within a controlled or uncontrolled environment. In a controlled environment, subjects are required to use a specified type of computer with a specified operating system and keyboard. Typically, subjects are required to provide typing samples, regardless of whether a structured or unstructured text is to be collected, in a lab session. An uncontrolled environment is the opposite of a controlled environment; subjects are not obliged to provide typing samples under specific settings, they are free to use whatever computer, operating system and keyboard they have at hand. The advantage offered by samples collected using a controlled environment is the avoidance of noise that might be introduced from using a variety of computers, operating systems, and keyboards; for example, traditional free-standing PC keyboards and built-in laptop keyboards have different layouts. However, an argument against the used of controlled environments is that for many authentication applications, such as the online assessment application domain considered in this thesis, data collected in this way does not reflect the true life data to which any generated model can be expected to be applied. Thus, concerning the work presented in this thesis training/test data collected in an uncontrolled environment was considered to be the most desirable.

Regardless of whether structured or unstructured text is collected, and whether this is collected in a controlled or uncontrolled environment, data can be collected from subjects using either: (i) plug-in software which subjects are requested to download on to their computers where it operates in the background to collect keystroke data, or (ii) a web-based interface which subjects are requested to access and provide typing samples by entering text into a web form.

The first method is privacy intrusive in that some sensitive data, such as passwords, can be recorded, so subjects are typically not willing to run such software. Of course, the subjects can be advised to switch off the plug-in data collector while typing some sensitive data, but this seems undesirable. The second method is more secure in that the information typed is not collected at all. Consequently, concerning the collection of the ACB data set, the second mechanism was adopted.

3.3 The University of Liverpool ACB Data Set

This section presents details concerning the collection and make up of the ACB data set. Note that the ACB data set comprised unstructured text obtained in the context of an uncontrolled environment. The section comprises three subsections. The first, Sub-section 3.3.1, is concerned with the WBKTR tool used to collect the data set; the second, Sub-section 3.3.2, with adopted data collection process (using the WBKTR tool). Sub-section 3.3.3 then gives a statistical analysis of the ACB data set.

3.3.1 Data Collection Mechanism (ACB Data Set)

The ACB data was collected using the web-based interface approach (see above). More specifically using a bespoke keystroke dynamics data collection tool, the Web-Based Keystroke Timestamp Recorder (WBKTR) tool developed in JavaScript, and embedded into a HTML "front end"[1]. Figure 3.1 shows a screenshot of the WBKTR interface. Subjects were asked to provide answers (in English) to three discussion style questions; thus three typing samples could be collected from each subject. The idea was to mimic the situation where students are conducting online assessments; a situation where no constraints would be imposed regarding specific keyboards or operating systems. In other words, the idea was to allow subjects to type in the same manner as they would within an online learning environment that can be rendered using a variety of platforms, browsers and so on.

JavaScript was used to create the WBKTR tool because of its robust, cross-platform, operating characteristics. This also offered the advantage that no third-party plug-ins were required to enable the WBKTR tool. Another advantage of using JavaScript was that it avoided any adverse effect that might result from network delay when passing keystroke timing data to the "home" server, which might have affected the accuracy of recorded times. In other words, the script function works at the end user station to

[1]The interface can be found at http://cgi.csc.liv.ac.uk/ hsaalshe/WBKTR3.html

FIGURE 3.1: A Screenshot of the WBKTR interface.

record timestamp data (in milliseconds) within the current limitations of the accuracy of the end users' computer clock, not within the server time accuracy. Note also that the ability to paste text was disabled so the subjects could not cut and paste text from other sources while they were providing answers. The client-side scripting was done using JQuery, the most commonly used JavaScript open-source library.

Using the WBKTR tool three timing values were collected as follows:

1. $\mathcal{KD}^t$: key-down timestamp recording when a key is pressed.
2. $\mathcal{KU}^t$: key-up timestamp recording when a key is released.
3. $\mathcal{KH}^t$: key-hold time calculated as $\mathcal{KU}_i^t - \mathcal{KD}_i^t$.

Figure 3.2 shows some samples of the recorded keystroke data. From this data, the interval time between a keypress (say) k_i and a following keypress (say) k_{i+1}, or simply the *flight time* $\mathcal{F}^t$, can also be extracted such that:

$$\mathcal{F}_i^t = \mathcal{KD}_i^t - \mathcal{KU}_{i-1}^t$$

While typing each sample, a PHP script was used to transfer the keystroke time series data, using the JSON open-standard file format, and AJAX, to a local server. Note that the characters associated with key presses were not recorded. This was for two reasons: (i) because the thesis was focused on keystroke dynamics represented as time series, character (ASCII) codes do not lend themselves to this format; and (ii)because, as noted in the motivation for the work presented in this thesis (see Chapter 1), data confidentiality was considered to be important; therefore, recording what was actually being typed was considered inappropriate.

3.3.2 Data Collection Process (ACB Data Set)

The WBKTR tool described in the foregoing sub-section was used to collect data using volunteers, mostly students and staff at the University of Liverpool (staff volunteers

Index of keystroke	***KD***	***KU***	***KH***
1	1431628620594	1431628620714	120
2	1431628620722	1431628620754	32
.....	.....	.....	.....
4	1431628621610	1431628621698	88
.....	.....	.....	.....
86	1431628651457	1431628651585	128
87	1431628651618	1431628651761	143
.....	.....	.....	.....

FIGURE 3.2: Samples of the recorded keystroke timing features

included online instructors teaching online programmes). Subjects were asked to type at least 100 words in response to three discussion question (with no maximum limitation) so that an adequate number of keystrokes could be collected (for convenience the WBKTR environment included a scripting function to count the number of words per question). In this manner, it was conjectured, the minimum number of keystrokes collected per subject would not be less than 1200 (calculated by assuming that the average word length would be 4 keystrokes, thus $4*300*3 = 1200$). In the event of samples comprised of less than 1200 keystrokes, these would be discarded.

In total, data from 30 subjects was collected. The identity of each subject was anonymised; each subject was given a randomly selected ID number which could not be traced back to the subject. No additional information concerning subjects was recorded, such as gender and/or age. This was a deliberate decision so as to minimise the resource required by subjects providing the data. Also because this information was not needed as the focus of interest was subject authentication; the work presented in this thesis is not directed at drawing any conclusions about the nature of keyboard usage behaviour in the context of (say) age or gender.

3.3.3 Quantitative Interpretation (ACB Data Set)

This sub-section provides a brief statistical analysis of the collated ACB data set so as to provide an overview of the data. For the analysis, each subject's record was viewed as a single keystroke time series comprised of the three constituent time series.

Inspection of the data set demonstrated that the maximum keystroke time series length, in terms of the number of keystroke events (data points or key presses), was 4581; and the minimum 2206. Thus the average length of a keystroke time series, for the entire data set, was 3589 with a Standard Deviation (SD) of 529. The total number of keystroke events was 107677. Figure 3.3 shows the distribution of keystroke time series lengths over the entire data set.

As noted in Chapters 1 and 2, the main features used to represent keystroke time series, in the context of this thesis, are flight time ($\mathcal{F}^t$) and key-hold time ($\mathcal{KH}^t$). Thus,

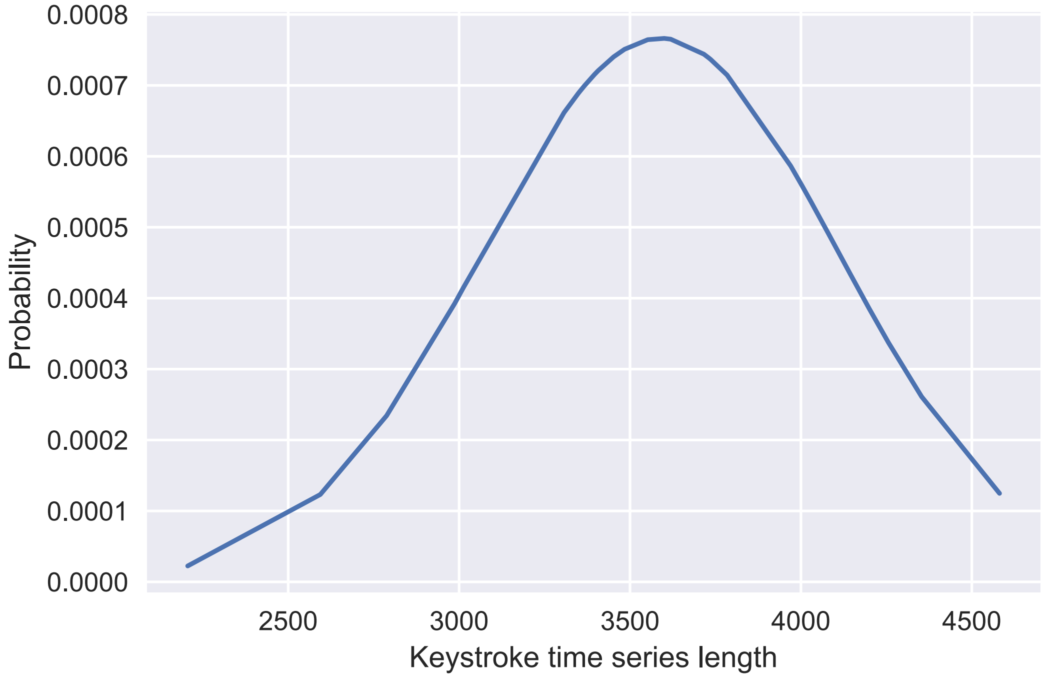


FIGURE 3.3: The distribution of keystroke time series size in The University of Liverpool Data Set.

further statistical information, concerning the average flight time and key-hold time, $Avg(\mathcal{F}^t)$ and $Avg(\mathcal{KH}^t)$ respectively, together with the associated standard deviations, $SD(\mathcal{F}^t)$ and $SD(\mathcal{KH}^t)$ respectively, is given in Table 3.1.

3.4 The University of Torino GP Data Set

This section presents an overview of the GP data set also used for evaluation purposes, in addition to the ACB data set described above, as described later in this thesis. The GP data set was generated by researchers at the University of Torino, in Italy, and is described in [54]. This data set was gathered from subjects who provided unstructured text typing samples, written in Italian, typing within an uncontrolled environment. As in the case of the previous section, the section is divided into three subsections, Sub-section 3.4.1 considers the adopted methodology used to collect the GP data set, Sub-section 3.4.2 the process and Sub-section 3.4.3 some statistical analysis of the data set.

3.4.1 Data Collection Mechanism (GP Data Set)

Typing samples, as in the case of the author's ACB data set, were collected using an HTML interface which featured a text form into which subjects could type. According to [54], the subjects could type anything they liked; suggestions included daily activities, favourite movies, holidays and so on. The subjects could access the HTML interface

TABLE 3.1: Average (Avg.) and Standard Deviation (SD) for the flight time $\mathcal{F}^t$ (Sec) and hold time $\mathcal{KH}^t$ (Sec) for each subject (record) in the ACB University of Liverpool data set.

Subject	$Avg(\mathcal{F}^t)$	$SD(\mathcal{F}^t)$	$Avg(\mathcal{KH}^t)$	$SD(\mathcal{KH}^t)$
sub1	0.204	0.257	0.373	0.820
sub2	0.268	0.259	0.085	0.034
sub3	0.280	0.374	0.069	0.036
sub4	0.321	0.336	0.080	0.028
sub5	0.290	0.280	0.068	0.045
sub6	0.316	0.373	0.092	0.083
sub7	0.238	0.277	0.063	0.027
sub8	0.194	0.259	0.072	0.023
sub9	0.442	0.392	0.104	0.046
sub10	0.346	0.396	0.082	0.037
sub11	0.242	0.314	0.072	0.072
sub12	0.526	0.485	0.117	0.055
sub13	0.348	0.323	0.098	0.048
sub14	0.382	0.393	0.093	0.025
sub15	0.371	0.405	0.084	0.033
sub16	0.326	0.364	0.056	0.044
sub17	0.519	0.416	0.137	0.074
sub18	0.297	0.367	0.112	0.041
sub19	0.202	0.230	0.444	0.142
sub20	0.290	0.340	0.982	0.782
sub21	0.284	0.421	1.532	0.115
sub22	0.334	0.451	0.089	0.039
sub23	0.316	0.408	0.094	0.041
sub24	0.198	0.250	0.077	0.025
sub25	0.316	0.408	0.094	0.041
sub26	0.547	1.293	0.067	0.028
sub27	0.419	0.925	0.095	0.043
sub28	0.594	0.200	0.073	0.045
sub29	0.683	1.097	0.092	0.049
sub30	0.772	0.119	0.081	0.053

once per day (thus one session a day), so could provide one typing sample per day. The subjects were also free to use whatever keyboard and operating system they had available. During each session, the key down timestamp ($\mathcal{KD}^t$) was recorded (in milliseconds) together with the ASCII value for each keystroke. However, in this thesis, because the focus is on keystroke dynamics as time series, the ASCII data was not used. Unfortunately, the timestamp for key releases ($\mathcal{KU}^t$) was not recorded so information concerning key hold time, ($\mathcal{KH}^t$), could not be calculated. Thus, in the data set flight time ($\mathcal{F}^t$) was derived as the time between two consecutive key down events, such that: $\mathcal{F}^t_i = \mathcal{KD}_i - \mathcal{KD}_{i-1}$. The timing data of keystroke dynamics was recorded using JavaScript, as also used with respect to the WBKTR tool described above, whilst the subject was filling the form. When the form was completed, the subject pressed a

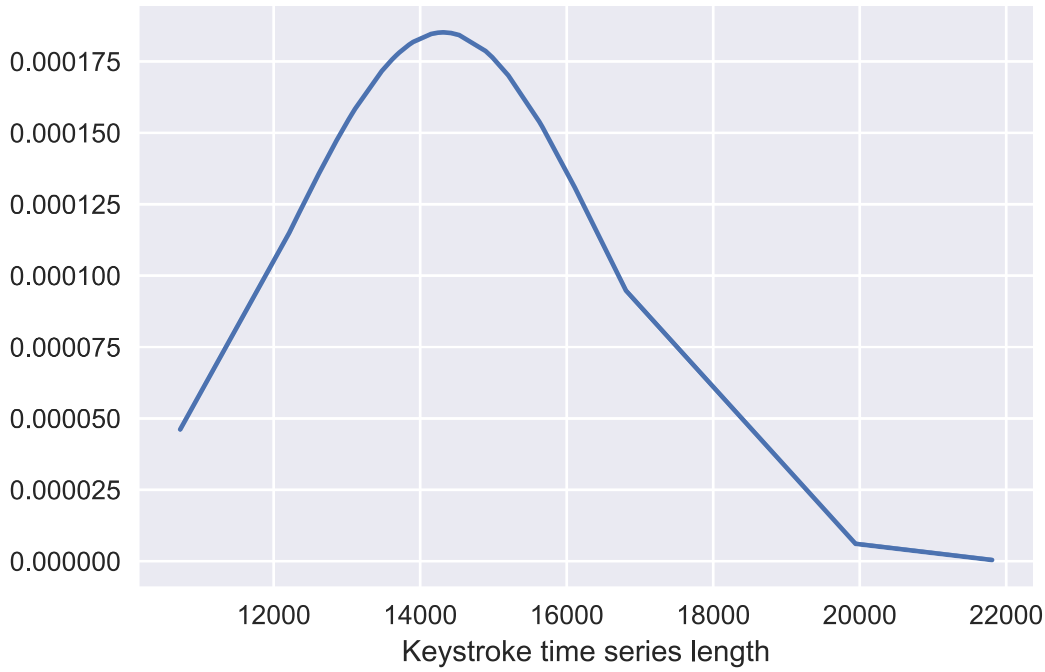


FIGURE 3.4: The distribution of keystroke time series size, per record, for the GP University of Torino Data Set.

submit button, at which point the keystroke data was transferred to a local server.

3.4.2 Data Collection Process (GP Data Set)

The GP data set was collected using two groups of subjects. The first group acted as legal users whilst the second group acted as imposter users. However, only the first group, real subjects, was used with respect to the evaluation presented later in this thesis. The reason was to maintain consistency with the other data sets used in this thesis where no imposter data was recorded.

The total number of real subjects was 40; however, the public version of the data set has only 31 subject records. The recruited subjects were postgraduate students and staff within the Department of Computer Science at the University of Torino; they were all native Italian speakers.

Each subject had to provide 15 typing samples; each was produced on a different day over a period of six months. The majority of the subjects provided typing samples on an irregular basis; some provided samples every two and three weeks; in another case, there was a lapse of two months between some samples. Note that the subjects provided samples in their own time using whatever keyboard and operating system they had at hand.

3.4.3 Quantitative Interpretation (GP Data Set)

According to [54], the HTML text form used to collect the GP data set was 56 characters wide and 12 lines long, thus space for a maximum of 672 ($56 \times 12 = 672$) characters. However, subjects could exceed this limit; the average length of each sample was 800 keystrokes (in both groups).

Overall, in the context of the real subject data, the maximum keystroke time series length was 21806 characters (keyboard events), whilst the minimum length was 10724 characters. Figure 3.4 shows the distribution of keystroke time series lengths per real subject record in the GP data set. The average and SD of the keystroke time series size were 14315 and 2189 respectively; the total length of keystroke time series, in the GP data set, was 443778. Concerning the $\mathcal{F}^t$ feature, Table 3.2 provides some statistical information, again in terms of the average and SD of the feature values.

3.5 Clarkson University VHHS Data Set

The VHHS data set was gathered by the authors of [161], at Clarkson University, USA. This data set comprised both structured (fixed) and unstructured (free) text samples, but all were obtained under laboratory conditions (thus a controlled environment). For the evaluation reported on in this thesis, only the unstructured (free) text samples were considered so that meaningful comparisons could be made with the results obtained using the other data sets. Thus the following subsections are concentrated on giving detail of the unstructured text data collection. As in the case of the descriptions of the ACB and GP data sets (Sections 3.3 and 3.4 above) this section is also divided into three sub-sections. Sub-section 3.5.1 presents the adopted collection mechanism, Sub-section 3.5.2 the data collection process and Sub-section 3.5.3 some statistics concerning the collated VHHS data set.

3.5.1 Data Collection Mechanism (VHHS Data Set)

For the VHHS data set the keystroke data was collected, using a controlled environment, in the form of timing information, using a client web browser (as in the case of the ACB and GP data sets). The timestamps for keystroke presses and releases were recorded in milliseconds. Typing samples, in the context of the unstructured text, were collected based on writing opinions in response to the set of questions listed in Figure 3.5 which gives a screenshot of the relevant part of the Clarkson keystroke data collection website. Note that the first six questions are "survey" questions whilst the last two ask subjects to provide a written description of the scene in two given pictures of the form shown in Figure 3.6 [2]. Note that in Figure 3.5, the numbering starts at four because, as already noted, we are interested in using unstructured text, the preceding numbering refers to structured text thus is not included here. Subjects were asked to write text into eight

[2]The second picture is not provided in [161].

TABLE 3.2: Average (Avg) and Standard Deviation (SD) for the flight time $\mathcal{F}^t$ (Sec) for each subject (record) in the GP University of Torino data set.

Subject	$Avg(\mathcal{F}^t)$	$SD(\mathcal{F}^t)$
sub1	0.294	0.798
sub2	0.231	0.395
sub3	0.333	1.039
sub4	0.278	0.984
sub5	0.252	0.552
sub6	0.416	0.801
sub7	0.251	0.547
sub8	0.222	0.581
sub9	0.240	0.634
sub10	0.326	0.781
sub11	0.176	0.492
sub12	0.259	0.421
sub13	0.451	1.381
sub14	0.443	1.022
sub15	0.324	0.933
sub16	0.234	0.378
sub17	0.309	0.660
sub18	0.207	0.532
sub19	0.225	0.517
sub20	0.236	0.653
sub21	0.286	0.432
sub22	0.272	0.807
sub23	0.234	1.032
sub24	0.376	1.461
sub25	0.237	0.396
sub26	0.321	0.707
sub27	0.353	0.950
sub28	0.267	0.511
sub29	0.358	1.553
sub30	0.211	0.292
sub31	0.354	1.453

web-based forms one per question, hence the first six were devoted to providing responses to the six survey questions, whilst the last two were dedicated to writing descriptions for two given pictures.

For the controlled environment used to collect the VHHS data the Explorer 8 web browser and desktop (PC) computers were used. The timestamp values for $\mathcal{KD}^t$ and $\mathcal{KU}^t$ were sequentially recorded for each keystroke; values for the $\mathcal{F}^t$ feature could thus subsequently be derived. The recorded keystroke data, per subject, was transferred, using a PHP script, to a local server in the form of a plain text file.

Freetext questions

4. Describe why you chose to study/work at Clarkson University. Are you satisfied with your decision to come to Clarkson University? In your experience what do you think is the most important quality in a university? Would you recommend Clarkson University to high school graduates/employees and why?

5. What activities do you like to do in your spare time? Do you like art, music or theatre? Describe the type of outdoor activities you like to do. What activities do you like to do in your current city Potsdam NY? Do you think there is a good balance of indoor and outdoor activities here in your current city Potsdam NY? If you could change anything about your current city Potsdam NY to increase social activities, what would you do?

6. Describe what you like and dislike about your current city Potsdam NY. Do you like the buildings and the lecture rooms? How do you find campus dining? Do you think the school has enough sporting facilities? What other recreational activities could be developed? How do you find the education? What do you think of the professors? What do you think about the classrooms and laboratories?

7. Describe your favorite vacation spot. Where is it located? What does it look like? Describe the environment. How crowded is it? What is the region surrounding look like? How often do you have the chance to go there? If you had the chance to go abroad for three months which countries would you like to visit? Why would you be interested in visiting them?

8. Describe your favorite class in college or high school and why you like it. Describe your least favorite class in college or high school and how the instructor can make it more interesting. Do you like interaction in classes? What do you think about labs? Do you find them useful in learning the concepts of a class? Would more homeworks and quizzes help you learn better? If there was not any grading involved would you enjoy the class better? Would you also learn better?

9. Describe your future career plans. Which sector are you planning to work in? Where do you see yourself in five years from now? Do you think you are close to your goals? What would make achieving your goals easier? Do you think Clarkson University prepares you well for achieving your goals? Do you think achieving your dreams is more about luck or hard work?

10. Describe what you see in this picture. What does the scene look like ? What are the people doing?

11. Describe what you see in this picture. What does the scene look like ? What are the people doing?

FIGURE 3.5: A screenshot from the Clarkson University keystroke data collection web site listing the questions to which subjects were asked to respond [161].

3.5.2 Data Collection Process (VHHS Data Set)

For the purpose of collecting the VHHS data set, a total of 39 subjects were recruited and asked to attend lab sessions at Clarkson University. The subjects were employees and students of Clarkson University. Although a description of the data set is given in [161] further detail concerning the subjects, such as identity, age and gender, is not provided.

The data collection was generated as follows. The subjects were required to attend two lab sessions on two separate days and write responses to the posed questions (see above). Subjects were asked to type at least 500 keystrokes in response to each question, thus a minimum of 4000 keystrokes per subject per session (8000 across both sessions). The subjects were allowed to attend the two sessions over a period of eleven months (between August 2011 and June 2012). According to [161], the average time span between sessions was one to two months, the minimum time span was a couple of days and the maximum four or five months.

Figure 3.6: A screenshot from the Clarkson University keystroke data collection web site showing one of the pictures which subjects were asked to describe [161].

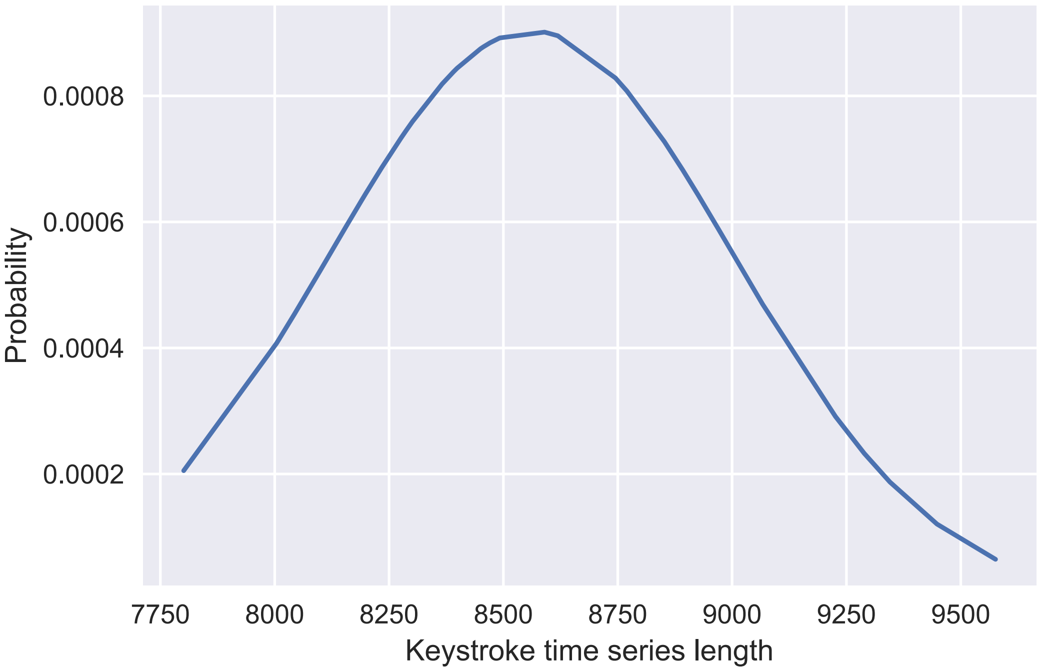


Figure 3.7: The distribution of keystroke time series size, per record, for the VHHS Clarkson University Data Set.

3.5.3 Quantitative Interpretation of (VHHS Data Set)

This sub-section provides a statistical analysis of the VHHS data set. For the analysis, the keystroke data, per subject, was concatenated into a single keystroke time series. The total number of keystroke events was 331128, with an average of 8490 and an SD of 712. The maximum and minimum keystroke time series lengths, for a single subject, were 9576 and 5034 respectively. Note that, as described in the previous sub-section, subjects were requested to type at least 500 characters per question response. Thus the minimum keystroke should not be less than 8000 across both sessions, whilst the minimum value was found to be 5034. The later was simply (according to [161]) because one subject had not completed all questions in one of the typing session. Whatever the case, the distribution of the size of the keystroke time series in the VHHS data set is shown in Figure 3.7. Further statistical information, with respects to the values of $\mathcal{F}^t$ and $\mathcal{KH}^t$ per subject, is given in Table 3.3.

3.6 Data Summarisation

In the previous three sections details concerning the three data sets used with respect to the evaluations reported later in this thesis were presented. As noted in the previous three sections, each data set had different characteristics and properties. An overall summarisation for these data sets is thus provided in this section so as to provide a simple reference which can be used with respect to the material reported later in this thesis. The summarisation is presented in Table 3.4; the table lists the characteristics and properties of each data set. More specifically, the table gives the number of subjects (# Sub.), the kind of text entry (Txt.), the environment setting (Env.), the language used to type samples (Lang.), the number of sample for each subject (# Sam.), the duration over which the samples were collected (Dur.) and the average (Avg.) and standard deviation (SD) of the keystroke time series with respect to each entire data set.

3.7 Summary

In this chapter, the evaluation data sets, which have been utilised for evaluating the keystroke continuous authentication proposed in this thesis, have been introduced. Typically, collecting keystroke data, for the purpose of biometric authentication, is a non-trivial task, the nature of which depends on the nature of the biometric authentication to be conducted, static or continuous. The factors to be taken into consideration when planning to collect keystroke data for the evaluation authentication mechanisms was therefore first discussed. It was noted that, in the context of subject (user) authentication directed at those completing online assessments, it was most appropriate to collect unstructured text within uncontrolled environments as this best simulated the process of online assessment. In total three data sets were described: (i) The University of

TABLE 3.3: Average (Avg) and Standard Deviation (SD) for the flight time $\mathcal{F}^t$ (Sec) and hold time $\mathcal{KH}^t$ (Sec) for each subject (record) in the VHHS Clarkson University data set.

Subject	$Avg(\mathcal{F}^t)$	$SD(\mathcal{F}^t)$	$Avg(\mathcal{KH}^t)$	$SD(\mathcal{KH}^t)$
sub1	0.112	0.241	0.560	7.962
sub2	0.094	0.237	0.112	0.353
sub3	0.103	0.246	0.144	0.456
sub4	0.149	0.316	0.174	0.551
sub5	0.127	0.257	0.753	0.238
sub6	0.163	0.300	0.215	0.681
sub7	0.124	0.293	0.149	0.472
sub8	0.179	0.324	0.197	0.623
sub9	0.126	0.211	0.172	0.544
sub10	0.137	0.290	0.181	0.571
sub11	0.132	0.238	0.159	0.503
sub12	0.180	0.369	0.189	0.596
sub13	0.139	0.266	0.302	0.953
sub14	0.247	0.417	0.459	0.462
sub15	0.144	0.304	0.244	0.770
sub16	0.166	0.257	0.261	0.825
sub17	0.147	0.269	0.232	0.732
sub18	0.160	0.264	0.544	0.172
sub19	0.130	0.267	0.295	0.932
sub20	0.141	0.283	0.794	0.251
sub21	0.086	0.192	0.267	0.844
sub22	0.165	0.305	0.255	0.806
sub23	0.140	0.268	0.319	0.101
sub24	0.165	0.323	0.353	0.111
sub25	0.103	0.240	0.257	0.814
sub26	0.192	0.316	0.141	0.446
sub27	0.149	0.236	0.115	0.363
sub28	0.125	0.258	0.230	0.728
sub29	0.181	0.320	0.338	0.107
sub30	0.112	0.277	0.167	0.528
sub31	0.137	0.254	0.123	0.388
sub32	0.160	0.298	0.683	0.217
sub33	0.133	0.261	0.165	0.523
sub34	0.228	0.393	0.245	0.773
sub35	0.176	0.322	0.274	0.866
sub36	0.147	0.294	0.224	0.707
sub37	0.103	0.230	0.681	0.215
sub38	0.123	0.230	0.208	0.658
sub39	0.160	0.330	0.253	0.800

Liverpool ACB data set collated by the author, (ii) The University of Torino GP data set, obtained from [54] and (iii) The Clarkson University VHHS data set, obtained from [161]. Each was described in the chapter in terms of the mechanism and process used to collect the data, and in terms of a statistical analysis of the content of each data

Table 3.4: Summary of data sets.

Data set	# Sub.	Txt.	Env.	Lang.	# Sam.	Dur.	Avg.	SD
ACB	30	Unstr.	Uncont.	English	3	1 Day	3589	529
GP	31	Unstr.	Uncont.	Italian	15	15 Days -6 mths	14315	2189
VHHS	39	Unstr.	Cont.	English	12	2 Days -11 mths	8490	712

Key: Unstr. = *Understructured*, Uncont. = *Uncontrolled*, Cont. = *Controlled*, mths = *months*.

set. The following chapter presents some preliminaries and formulations concerning the proposed keystroke continuous authentication mechanisms presented later in the thesis.

Chapter 4

Keystroke Time Series Fundamentals and Preliminaries

4.1 Introduction

In Chapter 2 it was emphasised that the prominent methods for recognising typing patterns, in the context of continuous authentication, have focused on the use of quantitative statistical measures to represent n-graph timing information which in turn was encapsulated in a feature vector format (see for instance the work presented in [3, 38, 54, 107, 109]). However, n-graph representations only utilise a subset of the data that is potentially available. The claim made in this thesis is that more sophisticated methods, methods that use a greater proportion of the available data than used with respect to the n-graph approaches from the literature, may be more effective in the context of keystroke continuous authentication. The proposed mechanisms presented in this thesis are thus directed at using all available data other than the nature of the keypress; the latter for reasons of privacy preservation, as discussed earlier in the thesis. More specifically the methods proposed in this thesis consider keystroke dynamics in terms of a discrete data stream, a *time series*, comprised of a sequence of press-and-release temporal events. The distinction, between the proposed time series based method and previous work, can be said to be that the proposed method uses what might be termed "global features", whilst the previous methods use what might be termed "local features". The anticipation was that the collated time series would feature keystroke dynamic patterns unique to individual users. In other words, the shapes (fluctuations within the keystroke time series) would define continuous typing patterns that in turn can be used for the purpose of keystroke continuous authentication.

This chapter presents a number of items concerned with the fundamentals of keystroke time series analysis for continuous authentication, items required as a precursor to the work presented later in this thesis. The chapter commences, Section 4.2, with a formal description of the adopted Keystroke Time Series (KTS) representation and then goes on to describe the adopted data cleaning process in Section 4.3. Two types of KTS are considered in this thesis: (i) Univariate-Keystroke Time Series (U-KTS), and (ii) Multivariate-Keystroke Time Series (M-KTS). The distinction between U-KTS and M-KTS is discussed in Section 4.4. A significant task associated with any form of time series analysis is the comparison of time series, especially to determine the similarity between time series. The approach to keystroke time series similarity calculation adopted throughout this thesis was the Dynamic Time Warping (DTW) approach introduced previously in Chapter 2. Because of its significance with respect to the work presented in this thesis, this is described in further detail, with respect to both U-KTS and M-KTS, in Section 4.5. The chapter is concluded with a summary in Section 4.6.

4.2 Keystroke Time Series (KTS) Representation

Keyboard usage is typically undertaken in a sequential manner key-press by key-press; even when using the shift key to (say) type an uppercase letter the shift key is pressed before the character key. The nature of this typing process can be recorded as a sequence of real-valued data points, each with its own time stamp. In other words, the process of typing can be viewed as a sequence of keystroke events (key-presses), thus a time series. More formally, as noted in Section 2.3.3 (Chapter 2), for each key-press, there are two primary real-valued timing data entities which can be recorded: (i) the Key-down timestamp $\mathcal{KD}^t$ and (ii) Key-up timestamp $\mathcal{KU}^t$. These values can then be used to determine the value for two secondary data entities: (i) flight time $\mathcal{F}^t$, and (ii) key-hold time $\mathcal{KH}^t$. For any keystroke event i, the flight time $\mathcal{F}^t_i$ value is calculated using:

$$\mathcal{F}^t_i = \mathcal{KD}^t_i - \mathcal{KU}^t_{i-1} \tag{4.1}$$

whilst the value for $\mathcal{KH}^t_i$ is calculated using:

$$\mathcal{KH}^t_i = \mathcal{KU}^t_i - \mathcal{KD}^t_i \tag{4.2}$$

It is thus these secondary data entities, $\mathcal{F}^t$ and $\mathcal{KH}^t$, that are utilised with respect to the keystroke time series representation used in this thesis.

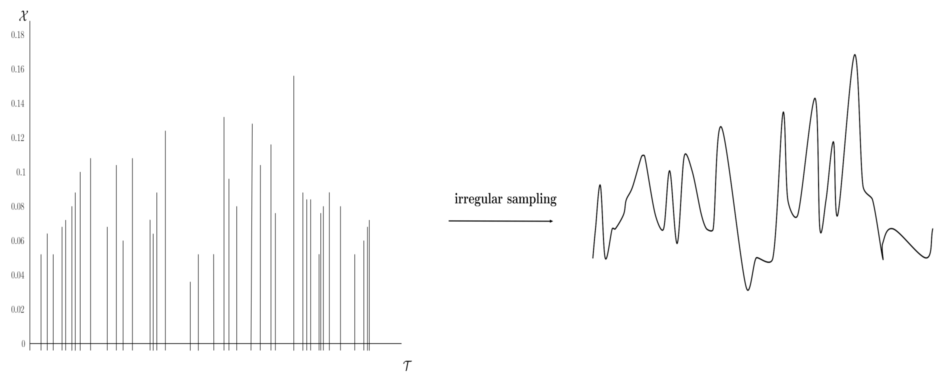
In more detail, if $\mathcal{X}$is used to indicate a keystroke event, either $\mathcal{F}^t$ or $\mathcal{KH}^t$, occurring at a time $\mathcal{T}$, the generic form of a keystroke time series will then be $\{\rho_1, \rho_2, \ldots\}$ where each point ρ_i is comprised of a tuple of the form of $\langle \mathcal{T}, \mathcal{X} \rangle$, such that $\rho_i = \{\langle \mathcal{T}_i, \mathcal{X}_i \rangle\}$. Thus:

$$\{\langle \mathcal{T}_1, \mathcal{X}_1 \rangle, \{\langle \mathcal{T}_2, \mathcal{X}_2 \rangle, \ldots, \{\langle \mathcal{T}_n, \mathcal{X}_n \rangle\}$$

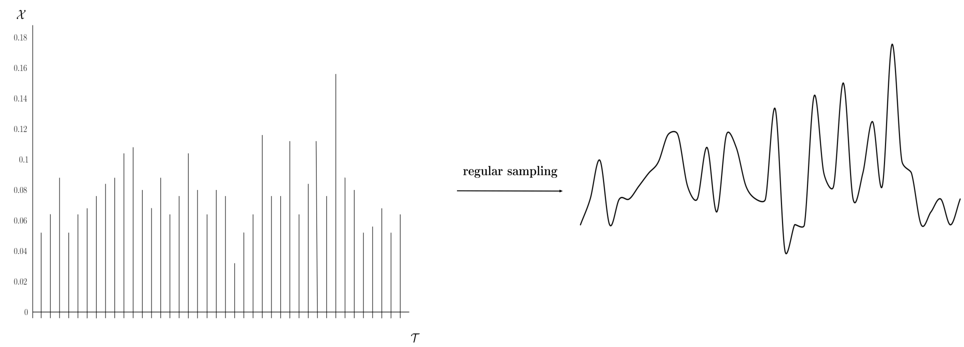
As such, a time series can be viewed as a 2D plot with time $\mathcal{T}$along the horizontal axis and the keystroke event value $\mathcal{X}$represented by the vertical axis. However, the "ticks" along the time series horizontal axis will typically then be irregularly spaced, which is seen as undesirable because it will tend to hinder the time series analysis, especially where there is a significant pause in typing activity. Consequently, in this thesis, the horizontal axis is considered to represent keystroke index numbers, such that each keystroke index $\mathcal{KN} = \{1, \ldots, \infty\}$. By doing so, the obtained series will be regularly spaced, and unaffected by changes in typing pace. Also, because the indexing can be inferred from the ordering of points in the time series, the keystroke time series can be simply conceptualised as a series of points such that:

$$\{\mathcal{X}_1 \mathcal{X}_2, \ldots, \mathcal{X}_n\}$$

Figure 4.1 illustrates the difference between using actual time stamps and keystroke indexes, as bar charts and as point series. Figure 4.1**(a)** gives a time series where the horizontal axis represents time stamps, note that the ticks are irregularly spaced; whilst Figure 4.1**(b)** gives the same time series where the horizontal axis records the keystroke index, note that the ticks are regularly spaced.



(a)



(b)

FIGURE 4.1: Examples of keystroke time series illustrating the distinction between using time stamps, Figure **(a)**, and keystroke event indexes, Figure **(b)**.

Given the foregoing the following definitions should be noted:

Definition 4.1. *A keystroke time series stream* ($\mathcal{K}_{ts}$) *is an ordered discrete sequence of keyboard events* $\{\rho_1, \rho_2, \ldots, \rho_n\}$ *where* $n \in \mathbb{N}$ *is the length of the series.*

Definition 4.2. *A point, keystroke event,* $\rho_i \in \mathcal{K}_{ts}$*, is parametrised as a tuple of the form* $\langle \mathcal{T}_i, \mathcal{X}_i \rangle$*, where* $\mathcal{T}$ *is an identifying index and* $\mathcal{X}$ *is one or more keystroke dynamics.*

Given a keystroke time series $\mathcal{K}_{ts_i}$ of length n, it can be divided in to $\frac{n}{l}$ subsequences where $l \in \mathbb{N}$ and $1 < l \leq n$ is the length of the derived subsequences.

Definition 4.3. *A keystroke time series subsequence (s), of length l, is a subsequence of $\mathcal{K}_{ts}$ that starts at the point ρ_i within $\mathcal{K}_{ts}$ and ends at point ρ_{i+l-1}, thus:*

$$s = \{\rho_i, \rho_{i+1}, \ldots, \rho_{i+l-1}\}$$

A subsequence (s) of $\mathcal{K}_{ts}$ can then be formally indicated using the notation:

$$s \preceq \mathcal{K}_{ts} \rightarrow \forall \rho_i \in s, \exists \rho_j \in \mathcal{K}_{ts} | \rho_i \equiv \rho_j$$

The significance is that the proposed mechanisms, with respect to keystroke continuous authentication, operate using such subsequences. The proposed idea is to monitor the keyboard data stream and then periodically collate the points that have arrived so far and package these into a subsequence that can then be analysed for the purpose of continuous user authentication.

4.3 Data Cleaning

An issue with KTS represented using $\mathcal{F}^t$ values is that these capture significant pauses in keyboard activity and, on occasion, "away from keyboard" events. It was found that such pauses could adversely affect the extraction of typing patterns from KTS. The problem is illustrated in Figure 4.2 where a time series, indicated using a black-dotted line, is shown featuring 300 keystrokes and $\mathcal{F}^t$ values where some of the values are significantly higher than the rest. The keystroke time series stream has been taken from the ACB data set presented in Chapter 3. From the figure, it can be seen that there is significant fluctuation in the amplitude of the curve, fluctuation which was found to impede the effectiveness of any time series analysis applied. To address this issue, it was decided to apply some data cleaning to the keystroke time series stream as it arrived so that data with abnormally high $\mathcal{F}^t$ values, in other words "noise" or "outlier" values, could be removed.

To this end, a threshold, φ, for acceptable values of $\mathcal{F}^t$ was defined. The idea was to use this threshold, not to remove points from time series subsequences, but to reduce the associated $\mathcal{F}^t$ value to the value of φ where $\mathcal{F}^t > \varphi$. Returning to Figure 4.2 the red time series indicates the same time series as the black-dotted time series but with a φ threshold of 2 seconds applied. Note that an alternative solution for handling noise values in this context is by removing them; however, removing some values can discard useful information for keystroke time series so this was not considered for the proposed approaches.

In the context of the continuous authentication approaches proposed in this thesis, the above was applied to each keystroke time series subsequence s as it was collated. The pseudo code presented in Algorithm 1 describes the data cleaning process. The

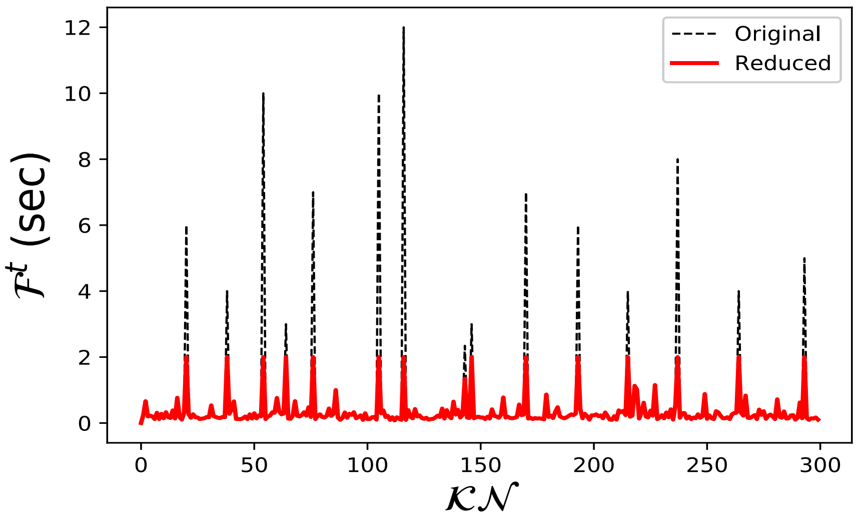


FIGURE 4.2: The effect of applying a threshold φ to a keystroke time series stream $\mathcal{K}_{ts}$ so as to limit the $\mathcal{F}^t$ values, $\varphi = 2$ (sec).

inputs are: a subsequence s where $s \subseteq \mathcal{K}_{ts}$ and the points represent $\mathcal{F}^t$ values, and a φ value. The output is a subsequence $\widehat{s}$ with $\mathcal{F}^t$ values greater than φ reduced to the value of φ.

The question that remains is what the value of φ should be. This is considered in further detail in Chapter 6 where the results from a series of experiments are reported on using a range of values for φ from 0.75 to 2.00 seconds increasing in steps of 0.25 seconds ($\{0.75, 1.00, 1.25, 1.50, 1.75, 2.00\}$).

It should also be noted that key-hold time, $\mathcal{KH}^t$, is normally no longer than 1 second. Inspection of the data sets used in this thesis indicated that the highest recorded value of $\mathcal{KH}^t$ was 0.950 millisecond. Consequently, it was felt that no threshold needed to be applied as in the case of $\mathcal{F}^t$. However, the same solution as proposed in the case of flight time can be proposed for key-hold noise data whenever it occurs in keystroke data.

Algorithm 1 Reducing Outlier Values of $\mathcal{F}^t$.

Input: $s \leftarrow$ subsequence of $\mathcal{K}_{ts}$, $\varphi \leftarrow \mathcal{F}^t$ limit.
Output: $\widehat{s} \leftarrow$ subsequence with reduced $\mathcal{F}^t$.

$s = (\rho_1, \rho_2, \ldots, \rho_i, \ldots, \rho_l)$
$l \leftarrow$ length of s
for $i = 1$ to $i = l$ **do**
  $\rho_i \leftarrow \langle \mathcal{F}^t_i \rangle$ ▷ Return $\mathcal{F}^t$ value from ρ (a tuple point).
  if $\rho_i > \varphi$: **then**
    $\rho_i == \varphi$
    Update(s)
  end if
end for
Return $\widehat{s}$

4.4 Univariate and Multivariate Keystroke Streaming

The keystroke time series considered in this thesis, as noted above, comprises $\mathcal{F}^t$ and $\mathcal{KH}^t$ values. Consequently, the keystroke time series data can be considered in either *Univariate* or *Multivariate* form. With respect to this thesis, both Univariate-Keystroke Time Series (U-KTS) and Multivariate-Keystroke Time Series (M-KTS) were considered. Thus, concerning the U-KTS form we have:

$$\{\langle \mathcal{T}_1, \mathcal{F}_1^t \rangle, \langle \mathcal{T}_2, \mathcal{F}_2^t \rangle, \ldots, , \langle \mathcal{T}_n, \mathcal{F}_n^t \rangle\}$$

or:

$$\{\langle \mathcal{T}_1, \mathcal{KH}_1^t \rangle, \langle \mathcal{T}_2, \mathcal{KH}_2^t \rangle, \ldots, \langle \mathcal{T}_n, \mathcal{KH}_n^t \rangle\}$$

Because $\mathcal{T}_i$ is represented as an index, this index can be inferred from the ordering of the points in the time series; and therefore, the above notation can be simplified to give:

$$\{\mathcal{F}_1^t, \mathcal{F}_2^t, \ldots, \mathcal{F}_n^t\}$$

and:

$$\{\mathcal{KH}_1^t, \mathcal{KH}_2^t, \ldots, \mathcal{KH}_n^t\}$$

Figures 4.3 and 4.4 give examples of U-KTS subsequences of length $n = 300$. Figures 4.3 shows four subsequences where the vertical axis records $\mathcal{F}^t$ values, whilst Figure 4.4 shows four subsequences where the vertical axis records $\mathcal{KH}^t$ values. The subsequences were randomly selected from the ACB data set used for evaluation purposes as reported on later in this thesis, and first introduced in Chapter 3, two subsequences per subject (a and b). From the figures, it can be seen that there are clear similarities in the keystroke subsequences associated with the same subjects (despite the subsequences being related to different texts), and clear dissimilarities in the keystroke subsequences associated with different subjects. The uniqueness of the individual typing patterns clearly demonstrates that distinct typing profiles (templates) exist with respect to individual users.

The argument against using U-KTS is that not all the data is utilised. Hence M-KTS are also considered in this thesis. These have the form:

$$\{\langle \mathcal{T}_1, \{\mathcal{F}_1^t, \mathcal{KH}^t - 1\}\rangle, \langle \mathcal{T}_2, \{\mathcal{F}_2^t, \mathcal{KH}^t - 2\}\rangle, \ldots, \langle \mathcal{T}_n, \{\mathcal{F}_n^t, \mathcal{KH}^t - n\}\rangle\}$$

Again because $\mathcal{T}_i$ can be inferred from the point ordering the above can be simplified to:

$$\{\{\mathcal{F}_1^t, \mathcal{KH}^t - 1\}, \{\mathcal{F}_2^t, \mathcal{KH}^t - 2\}, \ldots, \{\mathcal{F}_n^t, \mathcal{KH}^t - n\}\}$$

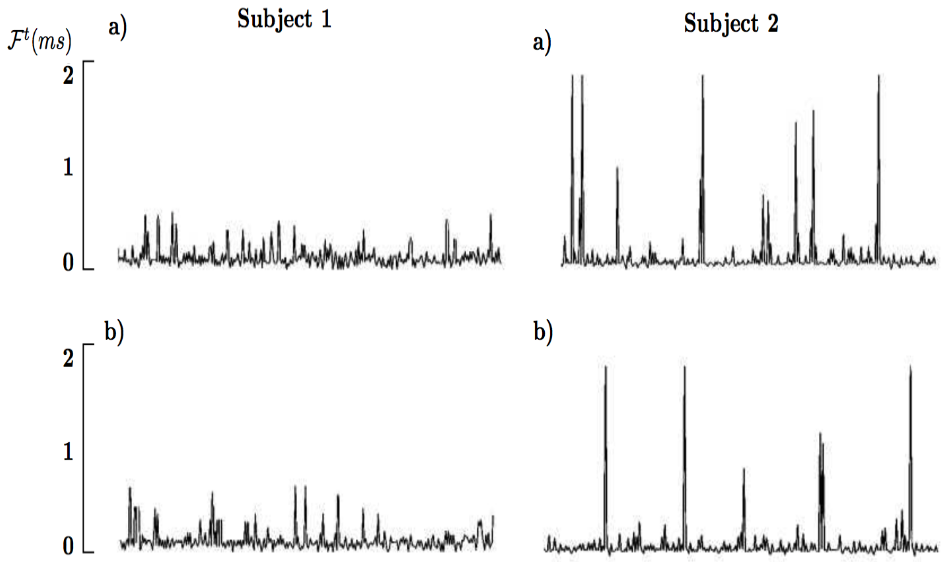


FIGURE 4.3: Four examples of U-KTS subsequences ($n = 300$), where the vertical axis records $\mathcal{F}^t$ values, for two subjects, two examples per subject (**a** and **b**), writing unspecified free text.

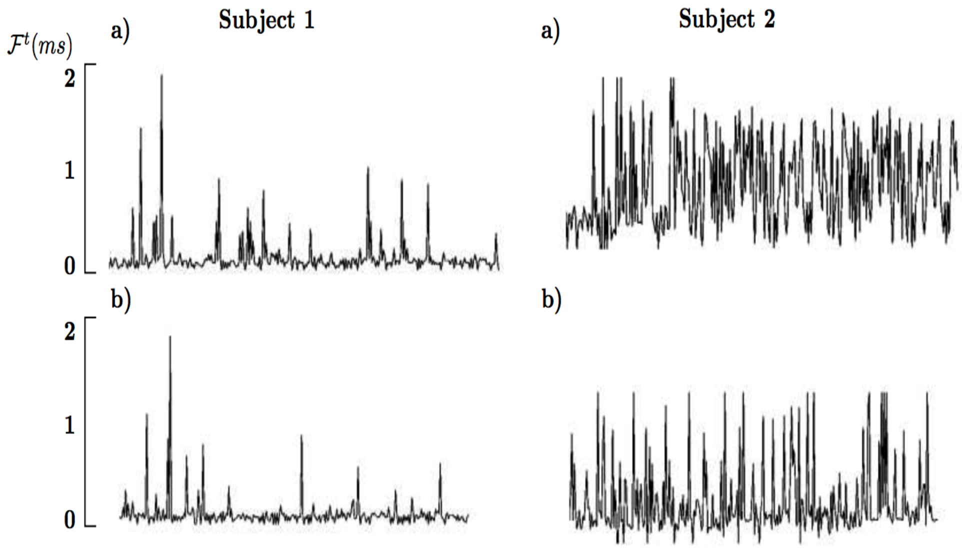


FIGURE 4.4: Four examples of U-KTS subsequences ($n = 300$), where the vertical axis records $\mathcal{KH}^t$, values, for two subjects, two examples per subject (**a** and **b**), writing unspecified free text.

4.5 Keystroke Time Series Similarity using Dynamic Time Warping

The key aspect of any form of time series analysis is the comparison of individual time series. For many applications, we wish to know the similarity between two given time series (or time series subsequences). This similarity is typically expressed as a distance measure of some kind. How this distance is calculated depends very much on the end goal of the analysis [59, 164]. The simplest mechanism is the Euclidean distance measure. This is derived as follows. Given two keystroke time series sequences $\mathcal{K}_{ts_1}$ and $\mathcal{K}_{ts_2}$ (of the same length), the Euclidean distance between each point in $\mathcal{K}_{ts_1}$ and each corresponding point in $\mathcal{K}_{ts_2}$ is calculated. The average distance is then calculated to give an overall similarity measure, sim. If $sim = 0$ the two time series are identical. However, as indicated in Chapter 2, for many applications, the Euclidean distance approach is considered to be over simplistic as it assumes a one-to-one correspondence of points. Any "offsets" (phase and amplitude differences) that might exist in the time series being compared is not considered. In the case of keystroke analytics, this can be illustrated by considering the time series subsequences given in Figures 4.3 and 4.4. Inspection of these subsequences indicates that "shapelets" (patterns) within the subsequences related to the same subject are similar, but that the "peaks" and "troughs" are offset from one another. The Euclidean distance similarity mechanisms will not capture this noticeable similarity.

To overcome this limitation, the work presented in this thesis used the Dynamic Time Warping (DTW) mechanism to measure the similarity between keystroke time series sequences. This is a well-established method [20, 37], which has been used effectively to find the similarity between pairs of point (time) series (readers may wish to refer back to Section 2.4.3 where the background to the DTW approach was given). In summary, DTW serves to warp the linearity of sequences (even of different lengths) so that any phase shifting can be taken into consideration. This is done by calculating what is referred to as a *warping path*. The length of this warping path, the *minimum warping distance*, is then treated as a similarity measure; if the length is zero the two time series under consideration are identical. Although the idea of DTW was presented in Section 2.4.3, further consideration of the DTW approach concerning keystroke time series is presented, for completeness, in the following two sub-sections. Sub-section 4.5.1 considers DTW in the context of U-KTS and Sub-section 4.5.2 considers DTW in the context of M-KTS.

4.5.1 DTW in Univariate-Keystroke Time Series (U-KTS)

Given two keystroke time series subsequences $s_1 = \{p_1, p_2, \dots, p_i, \dots, p_x\}$ and $s_2 = \{q_1, q_2, \dots, q_j, \dots, q_y\}$, where x and y are the lengths of the two series respectively, and the values represented by each point $p_i \in s_1$ and each point $q_j \in s_2$ are either $\mathcal{F}^t$ or $\mathcal{KH}^t$ values, the minimum warping distance between the two subsequences is computed as follows. A matrix $\boldsymbol{M}$ of size $(x-1) \times (y-1)$, is first constructed whereby the value held at each cell $m_{i,j} \in \boldsymbol{M}$ is the distance from point $p_i \in s_1$ to point $q_j \in s_2$:

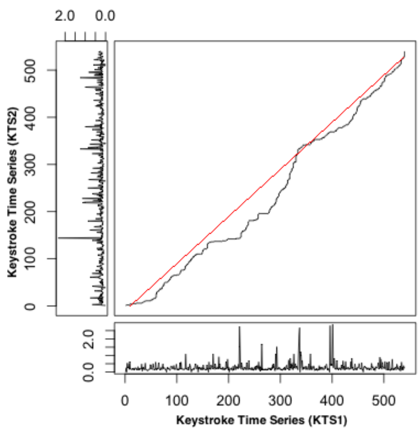
$$m_{i,j} = \sqrt{(p_i - q_j)^2} \tag{4.3}$$

The matrix $\boldsymbol{M}$ is used to find a minimum warping distance (Θ) which is then used as a similarity measure. A warping distance is the accumulated sum of the values associated with a *warping path* $(\mathbb{P})$ from cell $m_{0,0}$ to cell $m_{x-1,y-1}$. A warping path in this context is a sequence of cell locations, $\mathbb{P} = \{k_1, k_2, \dots\}$, such that given $k_n = m_{i,j}$ the follow on location k_{n+1} is either $m_{i+1,j}$, $m_{i,j+1}$ or $m_{i+1,j+1}$. The value for Θ associated with a particular $\mathbb{P}$ is then the sum of the values held at the locations in $\mathbb{P}$:

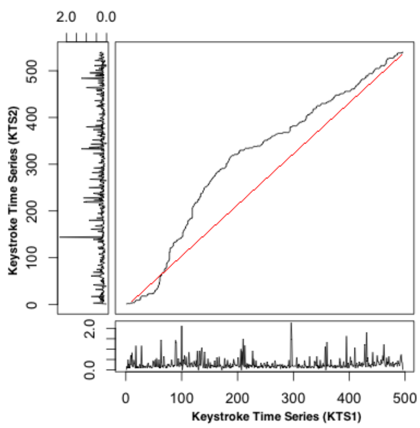
$$\Theta = \sum_{n=1}^{|\mathbb{P}|} k_n \in \mathbb{P} \tag{4.4}$$

To arrive at a minimum $\mathbb{P}$, for each location the following location is chosen so as to minimise the accumulated Θ. The "best" warping path is thus that which serves to minimise the distance from $m_{0,0}$ to $m_{x-1,y-1}$. The value of Θ, for a pair of keystroke time series, can, therefore, be interpreted as an indicator of the similarity between the two keystroke time series under consideration. Note that if $\Theta = 0$ the two keystroke time series in question will be identical. Note that the computation cost of DTW between the two time series is $\mathcal{O}(xy)$, where x and y are the lengths of the two keystroke time series.

Figure 4.5 shows examples of the application of DTW using four time series subsequences taken from the ACB evaluation data set. Figure 4.5 **(a)** shows the warping path $\mathbb{P}$ (black line) that results when the DTW process is used to compare two keystroke time series subsequences produced by the same subject, whilst Figure 4.5 **(b)** shows the warping path $\mathbb{P}$ that results when the DTW process is used to compare two keystroke time series subsequences produced by different subjects. In both cases, subjects were writing unstructured and unknown texts. The red diagonal line, included in both figures, indicates the minimum $\mathbb{P}$ that would have been obtained given two identical subsequences. Although it is not entirely clear from inspection of the figures, the minimum warping path for time series from the same subject is closer to the diagonal than that produced using different subjects.



(a)



(b)

FIGURE 4.5: Application of DTW: **(a)** the minimum warping path ($\mathbb{P}$) obtained when comparing two keystroke time series subsequences from the same subject, **(b)** the minimum warping path ($\mathbb{P}$) obtained when comparing two keystroke time series subsequences from different subjects. In both cases the typed text was unstructured.

4.5.2 DTW in Multivariate-Keystroke Time Series (M-KTS)

The method for determining Θ, using DTW, adopted with respect to the work presented in this thesis, directed at M-KTS, is to calculate two warping paths. This could be achieved using two DTW matrices. However, it is more efficient to use a single matrix with two values stored in each cell. An alternative approach would have been to store 3-D distances at each cell. Although much less storage would be required to store such 3-D distances the calculation of 3-D distances would be equivalent to calculating two 2-D distances. The main advantage offered by the proposed two 2-D distances approach is simplicity.

Thus given two M-KTS sequences, such that $s_1 = \{p_1, p_2, \ldots, p_i, \ldots, p_x\}$ and $s_2 = \{q_1, q_2, \ldots, q_j, \ldots, q_y\}$, where x and y are the lengths of the two series respectively, and the values represented by each point $p_i \in s_1$ and each point $q_j \in s_2$ comprise a tuple of the form $\langle \mathcal{F}^t, \mathcal{KH}^t \rangle$, Θ is calculated as follows. First a DTW matrix $\boldsymbol{M}$ of size $(x-1) \times (y-1)$ is constructed. Each cell $m_{i,j} \in \boldsymbol{M}$ then holds two distance values, the difference between the $\mathcal{F}^t$ value for point $p_i \in s_1$ and that for point $q_j \in s_2$; and the difference between the $\mathcal{KH}^t$ value for point $p_i \in s_1$ and that for point $q_j \in s_2$.

$$m_{ij} = \sqrt{\left((p_i^{(\mathcal{F}^t)} - q_j^{(\mathcal{F}^t)}) + (p_i^{(\mathcal{KH}^t)} - q_j^{(\mathcal{KH}^t)})\right)^2} \tag{4.5}$$

Two warping paths are then calculated, $\mathbb{P}_{\mathcal{F}^t}$ and $\mathbb{P}_{\mathcal{KH}^t}$. Each warping path is determined in the same manner as for the univariate (U-KTS) case. Two minimum warping distances are then determined, $\Theta_{\mathcal{F}^t}$ and $\Theta_{\mathcal{KH}^t}$. The final value for Θ is then the average of these two values:

$$\Theta = \frac{1}{2}(\Theta_{\mathcal{F}^t} + \Theta_{\mathcal{KH}^t}) \tag{4.6}$$

As in the univariate case, if the the value for Θ is zero both time series, s_1 and s_2 are the same. Note that the computation cost in the context of multivariate time series, as proposed here, is $\mathcal{O}2(xy)$.

4.6 Summary

This chapter has presented both the formalism and some preliminaries concerning the keystroke time series user authentication approaches presented later in this thesis. In summary keystroke time series, as conceived of in this thesis, are sequences of discrete data points such that each data point describes either: (i) the flight time $\mathcal{F}^t$ or (ii) the key-hold time $\mathcal{KH}^t$ or (iii) both $\mathcal{F}^t$ and $\mathcal{KH}^t$. Thus either univariate or multivariate time series. A key requirement for the proposed keystroke analysis approaches is to be able to find the similarity between pairs of sequences. The adopted method was the Dynamic Time Warping mechanism. Thus the chapter included detail concerning the application of DTW in both the univariate and the multivariate cases. In the following chapter, the first proposed approach, founded on the fundamentals presented in this chapter, is described and evaluated, namely the OKCA approach. This was essentially a proof of concept approach and considered user authentication in the static context; the objective being to investigate the validity of the proposed keyboard time series idea.

Chapter 5

Once-only Keystroke Continuous Authentication

5.1 Introduction

In the previous chapter, a number of preliminaries and formalisms relevant to continuous keystroke authentication were presented. In this chapter, the Once-only Keystroke Continuous Authentication (OKCA) system is presented, the first of a number of keystroke time series authentication approaches presented in this thesis. As the name suggests, the approach is directed at *once only*, or static, authentication, where the authentication takes place once typing has been completed. Although the main focus of the work presented in this thesis is continuous iterative authentication, whilst typing is taking place, the aim of the OKCA system was to establish the general principle of the proposed Keystroke Time Series (KTS) based approach in a simple (static) context before considering the more challenging iterative continuous authentication problem. In other words, OKCA was designed to provide a proof of concept of the utility of the KTS approach in which typing patterns are extracted from free text, which can be then employed for iterative continuous authentication.

The main idea of the OKCA system is that, given a new time series that it is claimed belongs to a particular subject, to compare this to a set of k profiles held in a typing template repository, that are known to belong to the claimed subject. The comparison is conducted using the DTW approach described earlier in the thesis. In this manner, a Warping Distance (WD) is determined, for each comparison, and an average WD derived accordingly. If this is below some user specified similarity threshold σ the claimed identity of the subject is confirmed; otherwise the claimed identity is rejected. Note that σ in this context is calculated by computing the average WD of the k profiles of the subject.

The remainder of this chapter is organised as follows. Section 5.2 provides a detailed description of the proposed OKCA system. This is followed by Section 5.3 where an extensive evaluation of the OKCA system is presented. In Section 5.4 some further discussion is given, and finally in Section 5.5, some conclusions concerning the material presented in the chapter are provided.

5.2 Once-only Keystroke Continuous Authentication (OKCA) System

From the preceding, the OKCA system operates simply by comparing a typing sample, that it is claimed to belong to a particular subject, to a bank of typing templates (references) that are known to belong to the subjects in the data set so that authentication can take place. Central to this approach is time series comparison using DTW as described earlier. The process is as shown in Figure 5.1.

Three variations of the OKCA were generated, two directed at Univariate Keystroke Time Series (U-KTS) representing either $\mathcal{F}^t$ or $\mathcal{KH}^t$, and one directed at Multivariate Keystroke Time Series (M-KTS). The distinction between the two univariate time series was that when using $\mathcal{F}^t$ values the φ threshold (noise reduction limit), as described in the previous chapter, was applied.

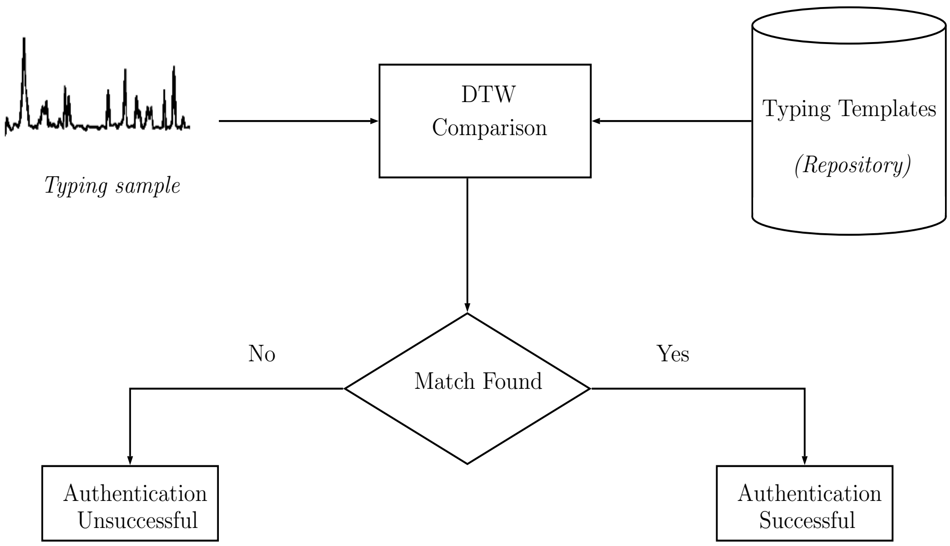


FIGURE 5.1: Schematic illustrating the operation of the OKCA system.

5.2.1 OKCA Algorithm

The basic operation of OKCA is given by the pseudo code presented in Algorithm 2. The inputs are: (i) time series (keystroke stream) $s = \{\rho_1, \rho_2, \dots\}$ where ρ_i is a single value in the case of U-KTS, and a two item tuple in the case of M-KTS, (ii) a similarity threshold σ and (iii) a noise limit φ, where $\mathcal{F}^t$ values are used. The output is a true-false binary value, true if the subject is deemed to be who he/she claims to be, and false otherwise. Note that the default result is "false" (line 1). If appropriate noise reduction is applied to s (line 3). The set of profiles S for the claimed subject are then processed in turn (loop from lines 5 to 9). On each iteration noise reduction is applied to each template $t_i \in S$ (line 6); this is not done when the templates are generated because a different value for φ may be required on each run. The similarity between S and each t_i (the minimum warping distance), sim_i, is then calculated (line 7), and a $simIndex$ value maintained (line 8). The warping distances are then averaged (line 10) and the average distance compared with the σ threshold (lines 11 to 13). Note that further detail of how to obtain the value of σ is given in Sub-section 5.3.1 below.

Algorithm 2 The General OKCA System Process.

Input: $s = \{\rho_1, \rho_2, \dots\}$, φ, σ
Output: $true$ or $false$

1: $result = false$
2: $simIndex = 0$
3: $s = NoiseReduction(s, \varphi)$
4: $S = \{t_1, \dots, t_i\}$ set of templates from repository.
5: **for** $\forall t_i \in S$ **do**
6: $\quad t_i = NoiseReduction(t_i, \varphi)$
7: $\quad sim_i = dtw(s, t_i)$
8: $\quad simIndex = simIndex + sim_i$
9: **end for**
10: $simIndex = \frac{\sum_{i=1}^{i=|S|} sim_i}{|S|}$
11: **if** $simIndex \leq \sigma$ **then**
12: $\quad result = true$
13: **end if**
14: return $result$

5.3 OKCA Validation

The proposed OKCA approach was evaluated with respect to the following three objectives:

1. To determine the effectiveness of the proposed approach in the context of Univariate keystroke Time Series (U-KTS).
2. To determine the effectiveness of the proposed approach in the context of Multivariate keystroke Time Series (M-KTS).
3. To compare the operation of the KTS approach with the feature vector based approach commonly used in the literature.

For the evaluation the data sets introduced in Chapter 3 were used. The evaluation was conducted by considering a subject identification scenario as opposed to a subject authentication scenario, the main focus of the work presented in this thesis. This was done so as to provide a more challenging evaluation of the proposed approach. Recall that authentication only requires comparison with profiles that belong to the claimed user, while identification requires comparison with all profiles in the repository.

To conduct the evaluation a "test harness" was first constructed; the nature of this test harness is discussed in Sub-section 5.3.1. Then, in Sub-section 5.3.2 the experimental setup is presented. The results, with respect to the proposed approach in the context of U-KTS, are given in Sub-section 5.3.3; whereas the reported results in the context of M-KTS are discussed in Sub-section 5.3.4. The outcomes with respect to the comparison with feature vector based approach are presented in Sub-section 5.3.5.

5.3.1 Test Harness

The fundamental idea underpinning the test harness was to divide vertically, each data set from Chapter 3, so that each record comprised three equal time series subsequences. Thus, given a time series s_i this would be divided into three subsequences s_{i_1}, s_{i_2} and s_{i_3}. One of these subsequences could then be used as the time series to be identified and the other two parts assigned to the required typing template repository (thus $k = 2$).

The partitions (groups) can be defined as follows: $a = \{s_{1_1}, \ldots, s_{1_n}\}$, $b = \{s_{2_1}, \ldots, s_{2_n}\}$ and $c = \{s_{3_1}, \ldots, s_{3_n}\}$, where n is the number of records (subjects) in the input data set. Consequently, 3-fold cross-validation could be conducted. Note that the reason for using 3-fold cross validation was to allow for a sufficient amount of data to be available to construct typing profiles.

The pseudo code for the test harness is given in Algorithm 3. The input is a data set $D = \{T_1, \ldots, T_n\}$ and a φ value (the flight time noise reduction limit). The output is a $|D| \times |D|$ similarity matrix Sim holding ranked similarity values (DTW warping distances). The process was commenced by dimensioning the matrix Sim in which to hold similarity measures (line 1) and the empty groups (a, b and c) (line 2). The groups are then populated in lines 3 to 13. In the pseudo code, it is assumed that the test group is a; this, of course, does not have to be the case. Each time series in the test group a is compared to each pair of time series in groups b and c using DTW, and an average warping distance, Θ (similarity value) obtained (line 20). If the corresponding samples to be compared are from the same subject, such that $dtw(s_{1i}, s_{2j})$ and $dtw(s_{1i}, s_{3j})$ where $i = j$, the obtained average similarity, Θ, serves as the value of threshold σ for the current subject i (lines 21-24). Recall that in the test harness groups b and c represent the typing template repository. Once all comparisons for a time series in group a have been complete the column values in the matrix representing this time series are ranked in ascending order. In this manner, we can observe whether the ranking value corresponds to the same subject, if it is ranked first then the identification has been successful.

5.3.2 Experimental Setting

A number of experiments were conducted to investigate the three OKCA evaluation objectives listed above. Experiments were conducted in terms of U-KTS and M-KTS

Algorithm 3 A Test Harness for OKCA System.

Input: $\mathcal{D} = \{T_1, \ldots, T_n\}$, φ (noise reduction limit).
Output: Return "highlighted" ranking similarity.

$Sim = |D| \times |D|$ similarity matrix
$a = \emptyset$, $b = \emptyset$ and $c = \emptyset$.
for $\forall T_i \in D$ **do**
    if $\mathcal{F}^t \in T_i$ **then**
        $T_i = NoiseReduction(T_i, \varphi)$
    end if
    $l = \left\lfloor \frac{T_i}{3} \right\rfloor$
    $gp1 = \{\rho_1, \ldots, \rho_l\} \in T_i$
    $gp2 = \{\rho_{l+1}, \ldots, \rho_{l+l}\} \in T_i$
    $gp3 = \{\rho_{l+l+1}, \ldots, \rho_{|T_i|}\} \in T_i$
    $a = a \cup gp1$
    $b = b \cup gp2$
    $c = c \cup gp3$
end for
for $i = 1$ to $|D|$ **do**
    $s_{1_i} = s_{1_i} \in a$
    for $j \in |D|$ **do**
        $s_{2_j} = s_{2_j} \in b$
        $s_{3_j} = s_{3_j} \in c$
        $\Theta = \frac{dtw(s_{1_i}, s_{2_j}) + dtw(s_{1_i}, s_{3_j})}{2}$
        if $i = j$ **then**
            $\sigma = \Theta$
            $highlight(\sigma)$
        end if
    end for
    $Sim = \Theta \cup Sim$
    Rank(Sim)
end for
return Sim

Table 5.1: Evaluation data set summary.

Data set	# Sub.	Txt.	Env.	Lang.	# Sam.	Dur.	Avg.	SD
ACB	30	Unstr.	Uncont.	English	3	1 Day	3589	529
GP	31	Unstr.	Uncont.	Italian	15	15 Days -6 mths	14315	2189
VHHS	39	Unstr.	Cont.	English	12	2 Days -11 mths	8490	712

Key: Unstr. = *Unstructured*, Uncont. = *Uncontrolled*, Cont. = *Controlled*, mths = *months*.

using the three data sets presented earlier in Chapter 3, namely: ACB, GP and VHHS. For convenience, Table 5.1, previously presented in Section 3.6 (Chapter 3), provides a statistical summary of the data sets. Note that the GP data set cannot be used to implement OKCA with respect to M-KTS as $\mathcal{KH}^t$ information was not provided in this data set.

As noted in the previous sub-section, although the pseudo code presented in Algorithm 3 implies that the test set is always group 1 (a), this does not have to be the case. Therefore, for each evaluation objective, three sets of experiments were conducted using a different group as the test set in each case. Thus: (i) a v. $\{b, c\}$, (ii) b v. $\{a, c\}$ and (iii) c v. $\{a, b\}$ (as illustrated in Figure 5.2).

The value for φ was set to 2 seconds for all the experiments. Evaluation regarding the effect of using different φ settings is presented later in this thesis in Chapter 6. The evaluation metrics used were: (i) Accuracy (Acc.), (ii) Mean Reciprocal Rank (MRR) and (iii) False Match Rate (FMR) and (iv) False Non-Match Rate (FNMR). Note also that all experiments were conducted using a 3.2 GHz Intel Core i5 processor with 24 GB RAM.

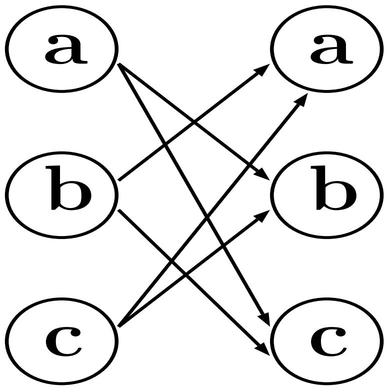


FIGURE 5.2: Schematic illustrating evaluation strategy using different groups.

5.3.3 Evaluation in Context of U-KTS

In the case of the univariate time series evaluation, the U-KTS evaluation, a total of fifteen individual experiments were conducted:

$$\begin{aligned} 3\ group\ configuration \times ((2\ datasets \times 2\ time\ series\ formats\ (\mathcal{F}^t\ and\ \mathcal{KH}^t)) \\ +(1\ dataset \times 1\ time\ series\ format\ (\mathcal{F}^t))) \end{aligned} \tag{5.1}$$

Recall that the GP data set did not include $\mathcal{KH}^t$ values, hence, in the case of GP data set, only U-KTS with $\mathcal{F}^t$ could be considered.

The complete set of results is presented in Appendix A; however, Table 5.2 gives a sample of the ranked *Sim* matrix, generated using the test harness described above, for the ACB data set when using U-KTS with $\mathcal{F}^t$ and assuming that the test group is group a. The highlighted values (in bold font) refer to the value where a subject is compared to itself. Thus, the last row in the table gives the ranking, r' (or simply the threshold σ), of the correct match; in other words, these are the values where a subject is compared to itself. Ideally, we would wish this comparison to be ranked first.

TABLE 5.2: Sample of Sim matrix for a v. $\{b, c\}$ using the ACB data set and U-KTS with $\mathcal{F}^t$ (*correct identifications* highlighted in bold font).

	S_1	S_2	S_3	S_4	...	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group b and c	**0.0528**	0.0329	**0.0635**	**0.0490**	...	0.0580	**0.0597**	**0.0385**	0.0677
	0.0578	**0.0374**	0.0655	0.0520	...	0.0667	0.0616	0.0392	0.0697
	0.0590	0.0401	0.0665	0.0530	...	0.0668	0.0646	0.0435	0.0720
	0.0598	0.0484	0.0683	0.0557	...	**0.0678**	0.0699	0.0477	0.0724
	0.0621	0.0487	0.0688	0.0579	...	0.0679	0.0699	0.0490	0.0740
	0.0631	0.0527	0.0719	0.0592	...	0.0706	0.0705	0.0680	**0.0751**
	0.0635	0.0549	0.0740	0.0619	...	0.0706	0.0712	0.0714	0.0770
	0.0646	0.0550	0.0741	0.0668	...	0.0706	0.0732	0.0761	0.0777
	0.0652	0.0551	0.0772	0.0685	...	0.0707	0.0741	0.0765	0.0782
	0.0658	0.0575	0.0773	0.0691	...	0.0739	0.0742	0.0797	0.0786
	0.0660	0.0581	0.0775	0.0710	...	0.0741	0.0748	0.0814	0.0786
	0.0682	0.0583	0.0778	0.0712	...	0.0744	0.0751	0.0819	0.0787
	0.0698	0.0640	0.0780	0.0713	...	0.0763	0.0777	0.0896	0.0788
	0.0725	0.0721	0.0781	0.0730	...	0.0767	0.0786	0.0898	0.0811
	0.0739	0.0732	0.0790	0.0756	...	0.0796	0.0803	0.0920	0.0814
	0.0748	0.0750	0.0834	0.0767	...	0.0798	0.0832	0.0952	0.0817
	0.0749	0.0799	0.0836	0.0786	...	0.0798	0.0836	0.0990	0.0826
	0.0753	0.0839	0.0846	0.0789	...	0.0801	0.0845	0.1098	0.0832
	0.0756	0.0906	0.0846	0.0837	...	0.0810	0.0850	0.1134	0.0877
	0.0805	0.0910	0.0847	0.0850	...	0.0824	0.0855	0.1135	0.0901
	0.0838	0.0961	0.0851	0.0864	...	0.0826	0.0875	0.1135	0.0928
	0.0883	0.1055	0.0864	0.0882	...	0.0826	0.0883	0.1170	0.0928
	0.0883	0.1063	0.0864	0.0882	...	0.0838	0.0892	0.1200	0.0952
	0.0892	0.1063	0.0865	0.0894	...	0.0879	0.0898	0.1236	0.0955
	0.0902	0.1099	0.0865	0.0914	...	0.0896	0.1009	0.1243	0.0972
	0.0907	0.1123	0.0870	0.0924	...	0.0925	0.1033	0.1288	0.1001
	0.0939	0.1195	0.0924	0.0940	...	0.0946	0.1079	0.1345	0.1038
	0.0984	0.1232	0.1118	0.0947	...	0.0957	0.1083	0.1649	0.1041
	0.1137	0.1307	0.1302	0.0963	...	0.1079	0.1111	0.1677	0.1085
	0.1177	0.1415	0.1310	0.1002	...	0.1134	0.1177	0.1747	0.1205
r'	**1**	**2**	**1**	**1**	...	**4**	**1**	**1**	**6**

A summary of the results obtained, in terms of r' values, is given in Tables 5.3 to 5.5 for data sets ACB, GP and VHHS respectively. Note that Table 5.4, ranking results for GP data set, only presents the results for U-KTS with $\mathcal{F}^t$. In each table, the rows

TABLE 5.3: The r' values obtained from the U-KTS evaluation using the ACB data set.

	U-KTS with $\mathcal{F}^t$			U-KTS with $\mathcal{KH}^t$		
S	a v. $\{b,c\}$	b v. $\{a,c\}$	c v. $\{a,b\}$	a v. $\{b,c\}$	b v. $\{a,c\}$	c v. $\{a,b\}$
1	1	1	2	1	1	2
2	2	1	2	2	3	2
3	1	1	1	4	3	1
4	1	1	2	2	4	3
5	2	2	2	3	2	3
6	1	2	1	5	4	6
7	5	4	2	10	8	8
8	1	1	1	5	7	6
9	2	1	1	6	6	4
10	2	2	2	1	2	1
11	3	3	2	5	5	6
12	3	3	3	4	6	7
13	2	2	1	13	10	12
14	1	1	1	2	5	1
15	2	2	1	10	9	7
16	3	4	3	3	4	4
17	2	2	4	5	4	6
18	6	6	5	1	2	3
19	2	2	1	6	5	7
20	2	3	3	7	6	7
21	5	5	4	5	2	4
22	1	1	1	6	5	4
23	3	3	4	2	1	1
24	1	1	1	2	1	5
25	7	5	8	2	1	3
26	1	1	2	3	1	1
27	4	3	3	1	2	1
28	1	2	1	1	1	1
29	1	2	2	2	3	2
30	6	4	5	1	1	2

represent the subject to be identified (the identifier of the subject is listed in column one). Results are presented with respect to each test group combination.

5.3.3.1 Subject Identification Accuracy

To obtain a single overall subject identification accuracy value (Acc) the ratio between the number of *incorrect identifications* (ℓ) ranked prior to a *correct match* (r') and the total number of comparisons (τ) was computed:

$$Acc = \frac{\tau - \ell}{\tau} \times 100 \tag{5.2}$$

TABLE 5.4: The r' values obtained from the U-KTS evaluation using the GP data set.

	U-KTS with $\mathcal{F}^t$		
S	a v. $\{b, c\}$	b v. $\{a, c\}$	c v. $\{a, b\}$
1	1	2	2
2	1	2	1
3	2	2	1
4	3	2	2
5	2	3	2
6	3	3	3
7	2	1	1
8	5	4	4
9	4	1	2
10	3	3	3
11	2	2	2
12	8	4	6
13	5	3	5
14	2	4	4
15	7	5	6
16	1	1	1
17	4	5	5
18	1	4	3
19	4	3	3
20	1	2	1
21	5	4	3
22	2	1	2
23	1	4	1
24	8	8	7
25	6	3	5
26	2	2	2
27	4	4	4
28	1	1	2
29	2	4	1
30	2	1	3
31	1	1	1

In this context, ℓ and τ are calculated as in Equations 5.3 and 5.4 as follows:

$$\ell = \sum_{i=1}^{N} (r'_i - 1) \tag{5.3}$$

$$\tau = 2(N^2) \tag{5.4}$$

where N is the number of subjects in the data sets. Recall that r'_i is the ranking of the desired match for subject S_i. Thus, with respect to Table 5.2, we have $\ell = 56$ and $\tau = 1800$, and a consequent *Acc* value of 96.89% ($\frac{1800-56}{1800} \times 100 = 96.89\%$).

Table 5.5: The r' values obtained from the U-KTS evaluation using the VHHS data set.

	U-KTS with $\mathcal{F}^t$			U-KTS with $\mathcal{KH}^t$		
S	a v. $\{b,c\}$	b v. $\{a,c\}$	c v. $\{a,b\}$	a v. $\{b,c\}$	b v. $\{a,c\}$	c v. $\{a,b\}$
1	1	2	1	2	3	1
2	1	1	1	2	1	1
3	2	2	2	3	3	5
4	3	2	3	1	2	3
5	2	3	2	1	2	2
6	9	8	8	11	9	9
7	2	3	2	3	5	4
8	4	4	6	12	10	9
9	1	1	2	5	2	2
10	2	1	2	8	9	7
11	2	3	3	3	5	5
12	3	3	2	8	6	6
13	1	2	1	3	2	1
14	6	5	5	6	4	8
15	2	1	1	4	2	3
16	1	1	1	1	2	1
17	9	7	9	12	10	11
18	3	4	2	1	2	2
19	2	1	4	15	13	13
20	3	2	2	1	3	2
21	1	1	1	5	5	2
22	7	6	4	1	2	3
23	2	2	2	3	2	5
24	3	3	4	1	1	3
25	8	6	5	2	1	1
26	7	8	6	3	1	3
27	2	2	1	10	7	8
28	1	4	3	4	4	6
29	3	3	2	5	4	5
30	1	1	1	3	3	3
31	6	4	6	2	1	3
32	1	2	2	6	7	4
33	2	2	2	8	9	9
34	1	1	1	10	12	11
35	4	3	4	5	6	3
36	1	1	2	1	1	2
37	3	2	1	2	4	1
38	1	1	4	1	1	4
39	1	1	2	4	2	1

The complete set of subject identification accuracy results obtained are given in Table 5.6. The table includes the overall average accuracy and associated Standard Deviation

(SD) with respect to this average accuracy. Inspection of the table demonstrates that the proposed OKCA system produced promising results, indicating the potential benefits that can be gained were the KTS representation to be adopted for user authentication; the best overall recorded accuracy was 97.21% (with an associated SD = 0.17) for the VHHS data set. The table also indicates that better results were obtained using U-KTS with $\mathcal{F}^t$ than U-KTS with $\mathcal{KH}^t$; thus an argument can be made that $\mathcal{F}^t$ is a more reliable feature than $\mathcal{KH}^t$ in the context of the evaluation of keystroke time series.

TABLE 5.6: The subject identification accuracy (**Acc.**) results from U-KTS evaluation.

	U-KTS with $\mathcal{F}^t$			U-KTS with $\mathcal{KH}^t$		
	ACB	GP	VHHS	ACB	GP	VHHS
a v. $\{b, c\}$	96.89	96.67	97.01	95.67	x	95.99
b v. $\{a, c\}$	97.06	96.98	97.34	96.00	x	96.12
c v. $\{a, b\}$	97.22	97.03	97.27	95.50	x	95.96
Avg.	97.06	96.90	97.21	95.72	x	96.02
SD	0.17	0.20	0.17	0.25	x	0.09

5.3.3.2 Mean Reciprocal Rank Evaluation

Table 5.7 presents Mean Reciprocal Rank (MRR) results obtained. MRR is a standard evaluation measure used in Information Retrieval where it is typically used as a measure of how close the position of a desired result is to the top of a ranked list [33]. MRR is calculated as follows:

$$MRR = \frac{1}{|Q|} \cdot \sum_{i=1}^{|Q|} \frac{1}{r_i} \tag{5.5}$$

where: (i) Q is a set of queries (in our case queries as to whether we have the correct subject or not), and (ii) r is the generated rank of the desired response to Q_i. An MRR of 1.00 would indicate that all the results considered are correct; thus we wish to maximise MRR.

From Table 5.7, it can be seen that best MRR values were obtained with respect to U-KTS with $\mathcal{F}^t$ confirming the observation made earlier that $\mathcal{F}^t$ is a better feature than $\mathcal{KH}^t$ with respect to the analysis of keystroke time series data. The MRR results obtained also indicate the potential of the proposed KTS representation for recognising typing patterns from free text for the purpose of the continuous authentication of subjects taking (say) online assessments.

5.3.3.3 FMR and FNMR Evaluation

FMR and FNMR were calculated with respect to each combination of groups for each data set and each U-KTS format supported by the data sets. FMR is the ratio of the number of false accepts (*false positives*) against the number of comparisons. In the context of OKCA, if the equivalent sample's rank was not equal to 1, this meant that the

TABLE 5.7: The **MRR** results obtained for U-KTS evaluation.

	U-KTS with $\mathcal{F}^t$			U-KTS with $\mathcal{KH}^t$		
	ACB	GP	VHHS	ACB	GP	VHHS
a v. $\{b, c\}$	0.570	0.530	0.540	0.460	x	0.440
b v. $\{a, c\}$	0.540	0.570	0.530	0.480	x	0.430
c v. $\{a, b\}$	0.550	0.530	0.500	0.490	x	0.420
Avg.	0.560	0.550	0.530	0.470	x	0.430
SD	0.02	0.02	0.02	0.01	x	0.01

sample has been falsely rejected. In contrast, FNMR was calculated by computing the ratio of the number of subjects that were ranked higher than the desired sample against the total number of comparisons. The subjects ranked before the desired subject were considered to represent false negatives, subjects that had been falsely accepted as real subjects. Thus FMR and FNMR were computed as follows:

$$FMR = \frac{false\ positives}{n} \tag{5.6}$$

$$FNMR = \frac{false\ negatives}{n} \tag{5.7}$$

where n is the total number of comparisons, for each group, such that:

$$n = true\ positives + true\ negatives + false\ positives + false\ negatives$$

Table 5.8 presents the reported FMR and FNMR results obtained. The table gives the overall average FMR and FNMR results from all three group combinations across the ACB, GP and VHHS data sets. From the table, the obtained results are shown for both U-KTS with $\mathcal{F}^t$ and U-KTS with $\mathcal{KH}^t$. Note again that GP data set has no results for the $\mathcal{KH}^t$ feature. From the table, it can be seen that, regardless of which data set was considered, best results were obtained using U-KTS with $\mathcal{F}^t$; the best recorded values were 2.78% and 0.91% for FMR and FNMR using the VHHS data set. The results thus indicate that good performance can be realised where the proposed KTS approach is used for keystroke continuous authentication where subjects are typing free text.

TABLE 5.8: The **FMR** and **FNMR** results obtained in U-KTS representation while applying $\mathcal{F}^t$ and $\mathcal{KH}^t$ features.

Representation	*U-KTS with $\mathcal{F}^t$*		*U-KTS with $\mathcal{KH}^t$*	
Metrics / Data set	FMR	FNMR	FMR	FNMR
ACB	2.94	1.13	4.28	1.24
GP	2.80	1.11	x	x
VHHS	2.78	0.91	3.98	1.02

5.3.4 Evaluation in Context of M-KTS

This sub-section presents the results obtained in the context of M-KTS; the second evaluation objective. Recall that for the M-KTS evaluation only the ACB and VHHS data sets were used. In this context, the constructed ACB and VHHS multivariate (bivariate) time series comprised $\mathcal{F}^t$ and $\mathcal{KH}^t$ values. The evaluation was conducted in a similar manner to the U-KTS evaluation reported on earlier. All three group combinations were considered in turn and ranked similarity (minimum warping distance) matrices constructed. The value for φ, as in the case of the U-KTS evaluation, was again set to 2 seconds. The complete set of results concerning the similarity matrices are presented in the Appendix A.

5.3.4.1 Subject Identification Accuracy

Table 5.9 presents the subject identification accuracy results obtained. From the table, and with reference to Table 5.6, it can be seen that M-KTS accuracy results obtained were better than the U-KTS results obtained. The overall average accuracies were $Acc = 98.04\%$ and $Acc = 98.21\%$ for the ACB and VHHS data sets receptively; obviously better than $Acc = 97.06\%$ and $Acc = 97.21\%$ as was recorded in the case of U-KTS with $\mathcal{F}^t$ for the same data sets. The obtained results again provided further evidence of the potential of the proposed KTS approach concerning continuous subject authentication.

TABLE 5.9: The subject identification accuracy (**Acc.**) results from M-KTS evaluation.

	M-KTS	
	ACB	VHHS
a v. $\{b, c\}$	97.94	98.09
b v. $\{a, c\}$	98.00	98.22
c v. $\{a, b\}$	98.17	98.32
Avg.	98.04	98.21
SD	0.12	0.12

5.3.4.2 Mean Reciprocal Rank Evaluation

Table 5.10 presents the MRR results obtained. From the table, and comparing with Table 5.7, it can be seen that when using M-KTS better MRR results were also obtained than in the case of U-KTS. The average obtained MRR average values were 0.640 and 0.570 using the ACB and VHHS data sets respectively, compared with 0.560 and 0.530 in the context of U-KTS.

5.3.4.3 FMR and FNMR Evaluation

Table 5.11 presents the FMR and FNMR results obtained. Inspection of the table indicates that superior performance, in terms of FMR and FNMR, is achieved in the context of M-KTS as opposed to U-KTS in all cases. The best recorded FMR value was

TABLE 5.10: The **MRR** results obtained from the M-KTS evaluation.

[redacted]	M-KTS	
	ACB	VHHS
a v. $\{b, c\}$	0.630	0.570
b v. $\{a, c\}$	0.640	0.590
c v. $\{a, b\}$	0.640	0.560
Avg.	0.640	0.570
SD	0.01	0.02

1.79%, whilst the best FNMR value was 0.89% (both in the context of the VHHS data set).

TABLE 5.11: The **FMR** and **FNMR** results obtained from the M-KTS evaluation.

Representation	M-KTS	
Metrics / Data set	FMR	FNMR
ACB	1.96	0.98
VHHS	1.79	0.89

5.3.5 OKCA and Feature Vector Representation (FVR)

The third objective of the evaluation presented here was to compare the operation of the proposed KTS approach with that of the statistical feature vector based representation (FVR) frequently encountered in the literature. In this context, feature vectors were constructed using the average flight time $\mu(f^t)$ for the shared occurring di-graphs found in two typing samples to be compared. Thus:

$$\mu(f^t) = \frac{1}{d}\sum_{i=1}^{i=d} \mathcal{F}_i^t \tag{5.8}$$

where d is the number of identified shared di-graphs and $\mathcal{F}^t$ is the flight time value between the identified di-graphs. Each identified di-graph was thus a feature (dimension) in a feature space with the range of average $\mu(f^t)$ values as the dimension.

The similarity between vectors was measured using the Cosine Similarity (CS) metric (DTW would clearly be inappropriate); a well established method for finding similarity between two non-zero vectors [143] extensively used in the context of information retrieval, data mining and pattern recognition. Recall that CS was used in this context as a baseline to compare the proposed methods. Thus other similarity mechanisms, concerning FVR, were not considered because the main focus of the thesis was the use of time series methods.

Whatever the case, CS can be calculated using the following equation:

$$CS(x, y) = \frac{x \cdot y}{||x|| \times ||y||} \tag{5.9}$$

Table 5.12: Sample of CS similarity matrix for a v. $\{b, c\}$ using the ACB data set and the feature vector representation (*correct matches* highlighted in bold font).

	S_1	S_2	S_3	S_4	...	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group b and c	0.6821	0.6409	0.7964	0.7776	...	0.6875	0.8155	0.7323	0.7654
	0.6690	0.6403	0.7307	0.7667	...	0.6229	0.7202	**0.6495**	**0.7525**
	0.6680	0.6207	0.7241	0.7634	...	0.6051	0.6852	0.5867	0.7506
	0.6567	0.6171	0.7063	0.7093	...	0.5985	0.6731	0.5537	0.7018
	0.6137	0.6146	0.6725	0.6727	...	0.5713	0.6669	0.5249	0.6986
	0.5881	0.5967	0.6443	0.6683	...	0.5696	**0.6556**	0.5234	0.6936
	0.5689	0.5957	0.6327	0.6651	...	0.5610	0.6491	0.5233	0.6890
	0.5530	0.5934	0.6134	0.6595	...	0.5553	0.6471	0.5146	0.6778
	0.5487	0.5926	0.5880	0.6587	...	0.5481	0.6439	0.5044	0.6732
	0.5038	**0.5895**	0.5826	0.6480	...	0.5374	0.6243	0.4677	0.6690
	0.5032	0.5860	0.5809	0.6463	...	0.5244	0.6221	0.4672	0.6518
	0.4953	0.5741	**0.5805**	0.6463	...	0.5199	0.5876	0.4666	0.6432
	0.4906	0.5726	0.5804	0.6391	...	0.5134	0.5840	0.4662	0.6316
	0.4830	0.5655	0.5597	**0.6379**	...	**0.5110**	0.5807	0.4598	0.6264
	0.4824	0.5568	0.5516	0.6229	...	0.5047	0.5797	0.4576	0.6161
	0.4770	0.5316	0.5505	0.6043	...	0.4931	0.5760	0.4498	0.6143
	0.4763	0.5310	0.5434	0.5964	...	0.4676	0.5614	0.4491	0.5897
	0.4580	0.5291	0.5388	0.5944	...	0.4635	0.5582	0.4416	0.5842
	0.4575	0.5239	0.5351	0.5895	...	0.4577	0.5582	0.4386	0.5766
	0.4530	0.5171	0.5212	0.5882	...	0.4530	0.5560	0.3948	0.5751
	0.4509	0.5063	0.5137	0.5816	...	0.4444	0.5462	0.3937	0.5737
	0.4384	0.4980	0.5094	0.5782	...	0.4374	0.5451	0.3908	0.5546
	0.4371	0.4964	0.5013	0.5698	...	0.4333	0.5368	0.3908	0.5369
	0.4236	0.4804	0.4897	0.5695	...	0.4273	0.5282	0.3898	0.5306
	0.4236	0.4804	0.4750	0.5325	...	0.4013	0.5249	0.3893	0.5284
	0.4180	0.4691	0.4333	0.5269	...	0.4013	0.4999	0.3845	0.5127
	0.3945	0.4655	0.4247	0.5114	...	0.3967	0.4992	0.3770	0.4799
	0.3897	0.4443	0.4064	0.5103	...	0.3780	0.4888	0.3700	0.4675
	0.3821	0.4355	0.4064	0.5058	...	0.3521	0.4830	0.3667	0.4385
	0.3449	0.3492	0.3898	0.4493	...	0.3463	0.4554	0.2727	0.4385
r'	**7**	**10**	**12**	**14**	...	**14**	**6**	**2**	**2**

where $x \cdot y$ is the dot product between two feature vectors x and y, and $|x|$ ($|y|$) is the magnitude of the vector x (y). Note that for the CS calculation to operate correctly the feature vectors need to be of the same length (unlike in the case of DTW).

Table 5.13: The average subject identification Acc., MRR, FMR and FNMR using FVR approach across the ACB, GP and VHHS data sets.

Data set \ Metrics	Acc.	MRR	FMR	FNMR
ACB	81.54	0.148	18.46	1.61
GP	83.33	0.176	16.67	1.54
VHHS	87.72	0.249	12.28	1.12

For the comparison reported on in this section, a similarity matrix, of the form constructed with respect to the experiments reported on earlier in this chapter, was constructed using the feature vector based approach and CS values. Table 5.12 presents a sample of the resulting CS similarity matrix. The entire matrix is presented in Appendix A. As in earlier cases, the last row in the table gives the ranking, r', of the correct match (highlighted in bold font); in other words, the ranking where a subject is compared to itself.

Using the rankings presented in the similarity matrix, subject identification accuracy (Acc.), MRR, FMR and FNMR values were calculated in the same manner as before. The average results, average from three group combinations, are presented in Table 5.13. Comparing these results with the results presented earlier in the context of U-KTS and M-KTS, it can be seen that worse results were attained with respect to all four metrics considered. It can thus be argued that the proposed KTS approach is superior to the established feature vector approach in the context of subject identification using keystroke dynamics, indicating that this is also likely to be the case concerning subject authentication.

5.4 Evaluation Summary and Discussion

For completeness, this section summarises the main findings concerning the experiments reported on in the previous section. The summary is firstly presented in terms of the obtained r' values, and secondly is given in the context of the average subject identification accuracy (Acc.), MRR, FMR and FNMR.

The r' analysis is presented using the sequence of "whisker plots" illustrated in Figures 5.3, 5.4 and 5.5, one for each data set. Each figure shows the whisker plot for the ranking values with respect to the four representations used in the context of the OKCA approach as follows:

1. U-KTS with $\mathcal{F}^t$ (referred to in the figures as U-KTS+$\mathcal{F}^t$).
2. U-KTS with $\mathcal{KH}^t$ (referred to in the figures as U-KTS+$\mathcal{KH}^t$).
3. M-KTS.
4. FVR.

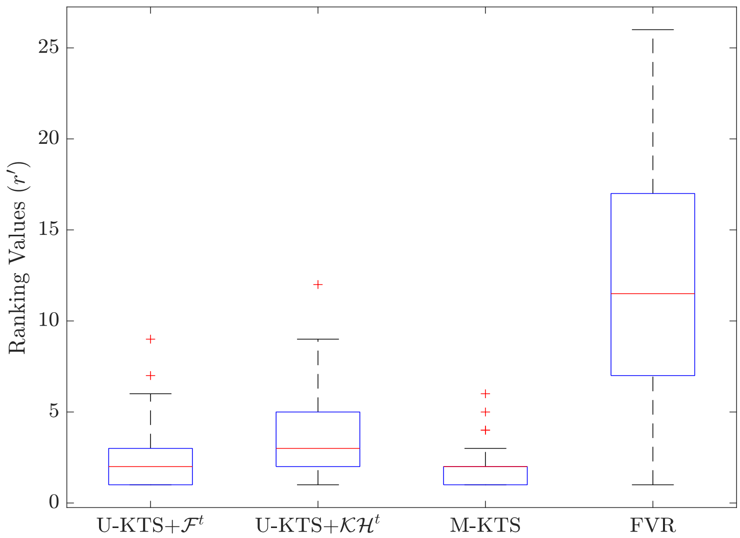


FIGURE 5.3: The statistical distribution for ranking values (r') for U-KTS with $\mathcal{F}^t$ (U-KTS+$\mathcal{F}^t$), U-KTS with $\mathcal{KH}^t$ (U-KTS+$\mathcal{KH}^t$), M-KTS and FVR applied to the ACB data set.

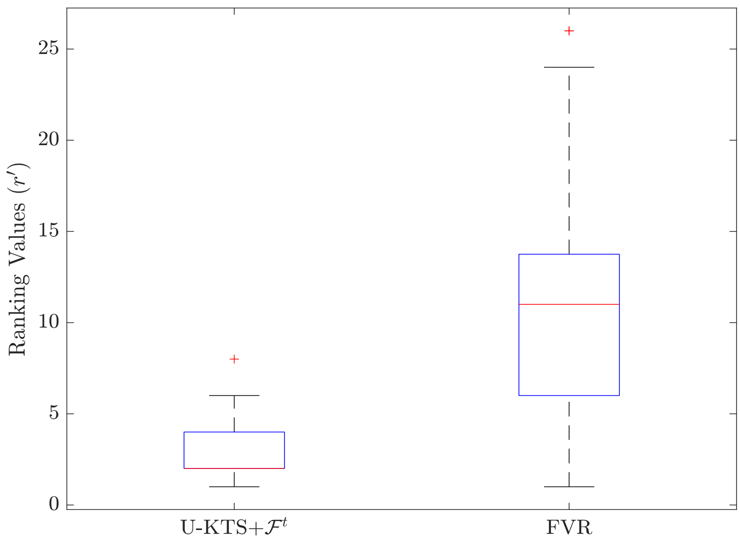


FIGURE 5.4: The statistical distribution for ranking values (r') for U-KTS with $\mathcal{F}^t$ (U-KTS+$\mathcal{F}^t$) and FVR applied to the GP data set.

The four representations are thus listed along the horizontal axis in each figure, except in the case of GP data set figure (Figure 5.4) whereas only the results for U-KTS+$\mathcal{F}^t$ and FVR were used because the GP data set only featured $\mathcal{F}^t$ values. The plots provide a statistical interpretation of the distribution of the r' values obtained in each case. More specifically, the "red" line in each box shows the median of the ranking values (r'), while the top and bottom of the box show the 75% and 25% quartile of the obtained rankings. Recall that rankings of value 1 indicate a correct match. Thus the closer the median

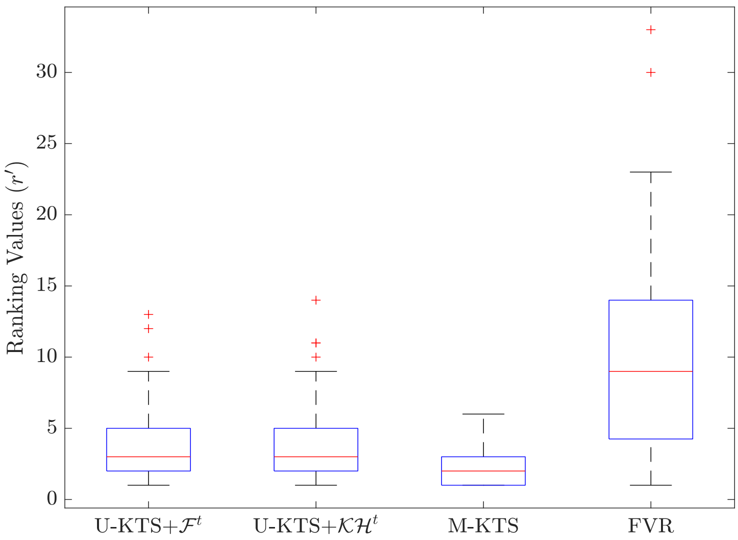


FIGURE 5.5: The statistical distribution for ranking values (r') for U-KTS with $\mathcal{F}^t$ (U-KTS+$\mathcal{F}^t$), U-KTS with $\mathcal{KH}^t$ (U-KTS+$\mathcal{KH}^t$), M-KTS and FVR applied to the VHHS data set.

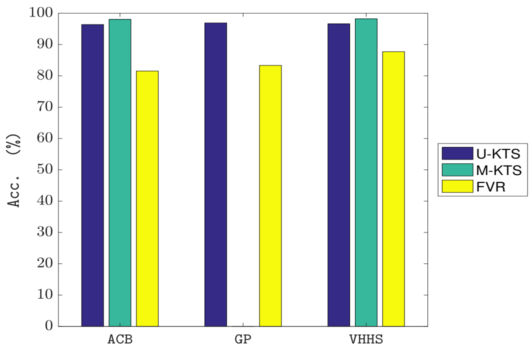


FIGURE 5.6: Subject identification Acc. (%) summary.

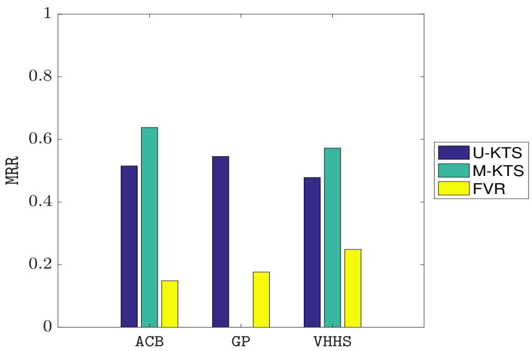


FIGURE 5.7: MRR summary.

is to one, the more accurate the performance of the approach. It can be seen from the figures that M-KTS has yielded the best performance.

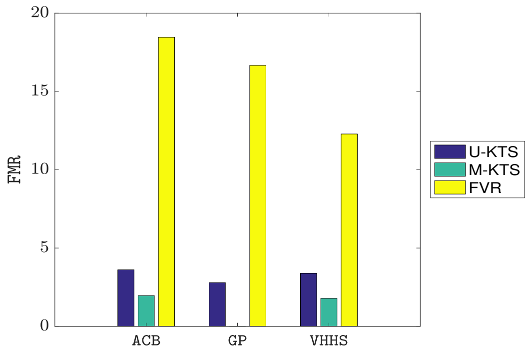


Figure 5.8: FMR summary.

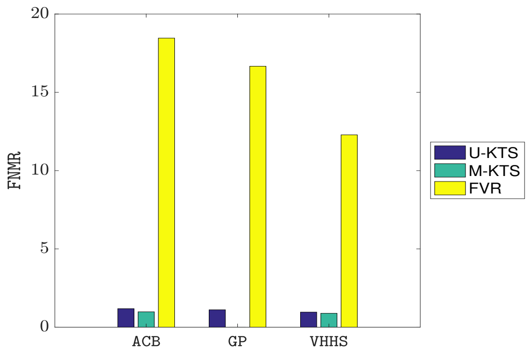


Figure 5.9: FNMR summary.

Figures 5.6, 5.7, 5.8 and 5.9 present the average accuracy (Acc.), MRR, FMR and FNMR results in the form of a set of histograms. Overall, the results, as presented in the figures, confirm the earlier observation that the proposed KTS approach outperformed the established FVR approach.

5.5 Summary

This chapter has presented the Once-only Keystroke Continuous Authentication (OKCA) approach, a KTS "proof of concept" system. The operation of the OKCA system was fully described and evaluated in the context of subject identification. Subject identification was used, instead of subject authentication, the primary focus of the thesis, because this was considered to be much more challenging. The evaluation was conducted using the three evaluation data sets presented previously in Chapter 3. The results presented demonstrated that the proposed KTS representation, coupled with DTW, produced good results. More specifically M-KTS outperformed U-KTS, whilst U-KTS with $\mathcal{F}^t$ outperformed U-KTS with $\mathcal{KH}^t$. In the following chapter, the first iterative keystroke continuous authentication approach proposed in this thesis is presented.

Chapter 6

Iterative Keystroke Continuous Authentication in Time Series

6.1 Introduction

In the previous chapter, the OKCA system was described, a static subject authentication system. This was a proof of concept system design to evidence the utility of the proposed time series approach to analysing keystroke dynamics. This was done by conducting a series of experiments directed at subject identification. The experimental results demonstrated that the keystroke time series representation can be successfully applied to recognising typing patterns from free text; without the disadvantages associated with the feature vector approach. It was thus conjectured that the time series approach could also be successfully applied for iterative continuous authentication using keystroke dynamics, the target application that the research presented in this thesis seeks to address. This chapter thus presents the first proposed approach to keystroke continuous authentication presented in this thesis, namely the Iterative Continuous Keystroke Authentication (IKCA) approach (the second approach is presented in the following Chapter). Using the IKCA approach, typing activity is conceptualised in terms of a continuous data stream (a time series), comprised of a sequence of press-and-release temporal events. This data stream is then sampled so as to extract non-overlapping shapelets using moving window mechanism. These shapelets can be generated from both univariate (U-KTS) and multivariate (M-KTS) keystroke data streams.

The rest of this chapter is structured as follows. Section 6.2 presents the primary framework for the IKCA approach. Then, in Section 6.3, the technical operation of the IKCA is presented. Section 6.4 presents an extensive evaluation concerning the implementation of the IKCA using the real keystroke data sets given in Chapter 3. Finally, Section 6.5 provides a comprehensive summary of the material presented in this chapter.

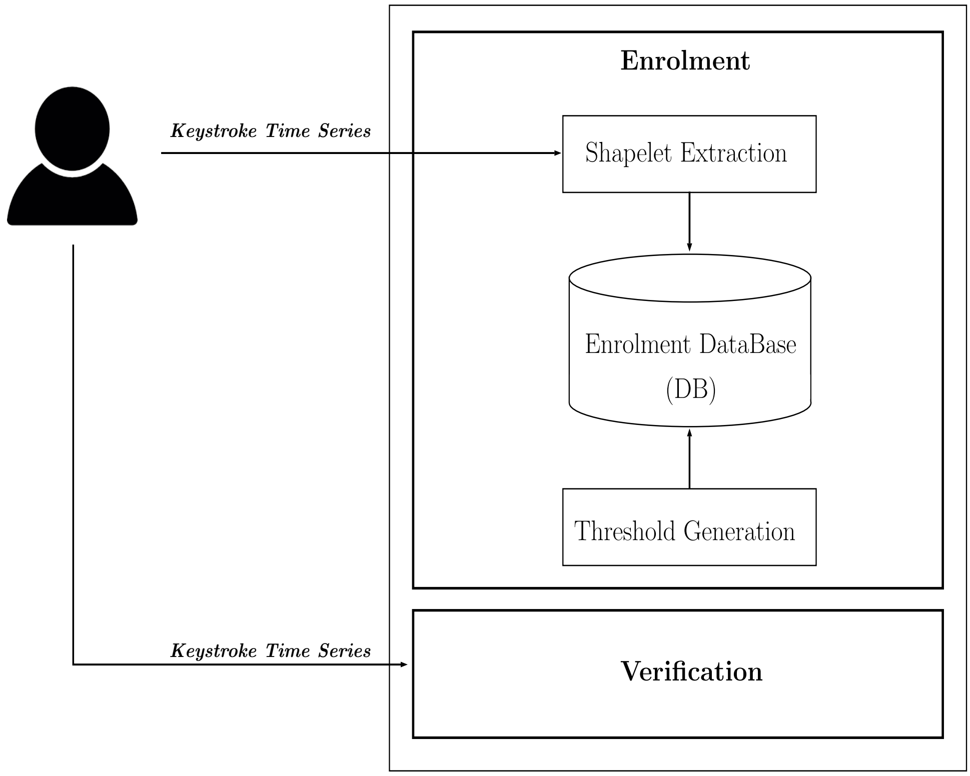


FIGURE 6.1: The IKCA operational framework.

6.2 IKCA Framework

This section presents the operational framework for the IKCA system. An overview of the framework is presented in Figure 6.1. At a high-level, the framework operates in a similar manner as other biometric pattern recognition mechanisms (see Sub-section 2.2.3, Chapter 2) in that it features two central components: (i) enrolment and (ii) verification. Enrolment is the process whereby an enrolment database, a database of profiles for legitimate subjects, is built up. Verification is the process whereby subjects are authenticated. The first precedes the second.

With respect to the proposed IKCA system, the building up of the profile is conducted by extracting time series subsequences, namely "shapelets", from a sample keyboard time series provided by each legitimate subject. The concept of shapelets is discussed further in Sub-section *shapelets* below. Enrolment also involves the generation of individual thresholds (σ values).

Further detail concerning the construction of the enrolment DataBase (DB), and σ value generation, is given in Sub-section 6.2.2. The verification process is given in more detail in Figure 6.2. The process iterates over the keyboard input stream periodically collecting shapelets. The first shapelet collected is compared to the templates for the claimed subject in the DB profile, this is referred to as "start-up" authentication and is described further in Sub-section 6.2.3. If a match is found the process continues, otherwise the process exits. Each subsequent shapelet is compared to the immediately

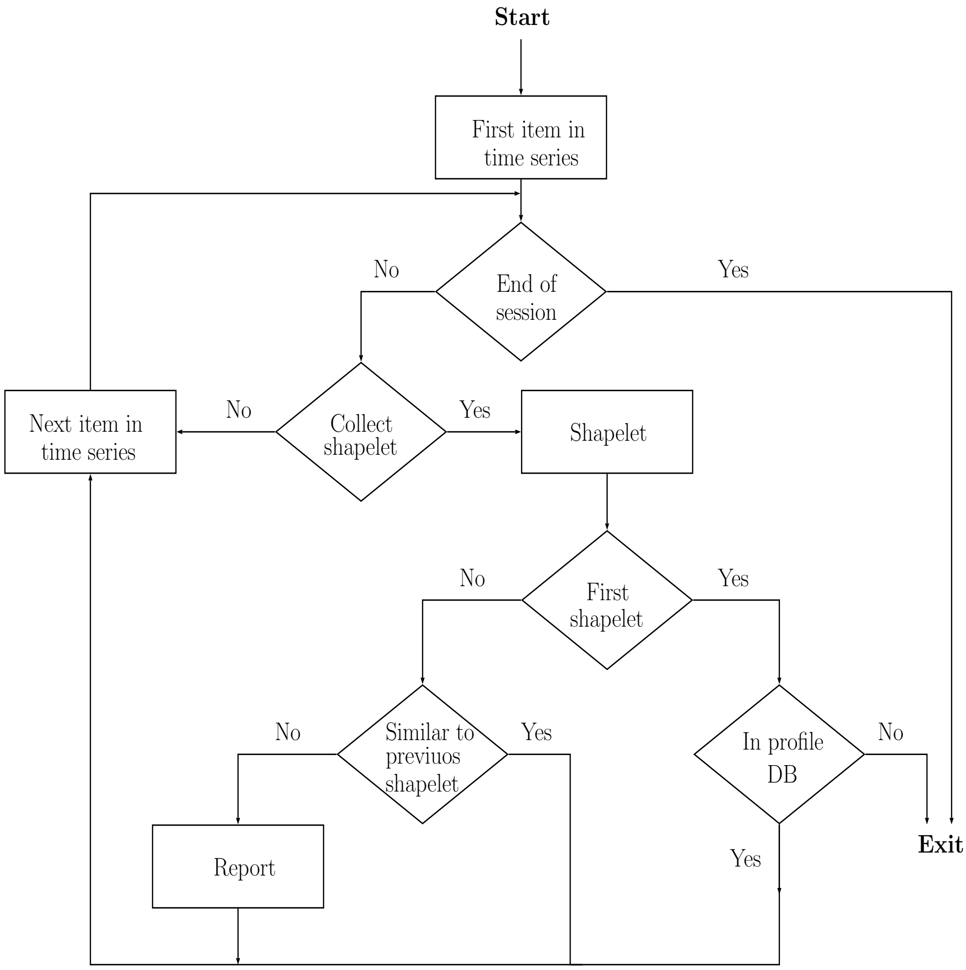


Figure 6.2: The IKCA subject verification process.

preceding shapelet, if the match is unsuccessful a report is generated. The process continues until the time series ends.

6.2.1 Shapelets

A key aspect of the IKCA methodology, with respect to both enrolment and verification, is the usage of time series subsequences called shapelets. The concept of shapelets is well established within the time series analysis community [95, 169]; in the context of the proposed IKCA system, shapelets are extracted using a sliding window of length ω, where ω is user defined. More formally, given a keystroke time series $\mathcal{K}_{ts_i} = \{\rho_1, \rho_2, \dots, \rho_n\}$, where ρ_i is some keystroke feature or set of features, a shapelet w of length ω is the subseries $\{\rho_i, \dots, \rho_{i+\omega}]$. The extracted shapelet can be then used as a typing template, or to authenticate a subject, as appropriate. For the experiments reported later in this chapter, a range of ω values was considered from 25 to 150 increasing in steps of 25 ($\{25, 50, 75, 100, 125, 150\}$) so as to examine the effect of ω on performance regarding accuracy. It was anticipated that a small window size would provide efficiency gains, desirable in the context of the IKCA approach; whilst a larger window would provide accuracy gains. A trade-off between efficiency and accuracy was therefore expected.

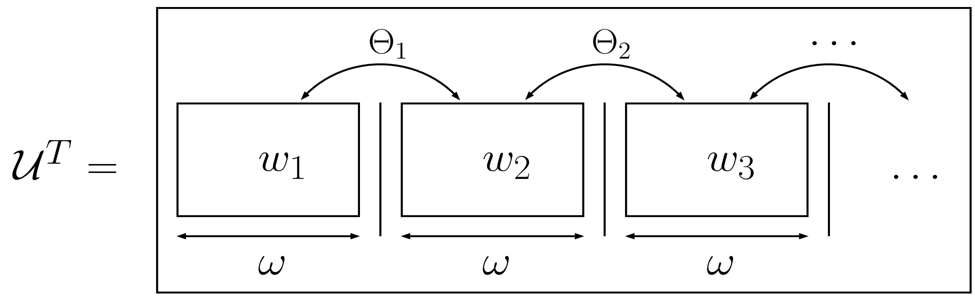


FIGURE 6.3: A schematic illustrating the process of constructing a user typing profile $\mathcal{U}^T$ for a single subject.

6.2.2 Enrolment Database Construction and σ Value Generation

From the foregoing, the IKCA system operates using an enrolment database, a "bank" of user (subject) typing templates (profiles), one per user, defined as follows:

Definition 6.1. *A user typing template (profile) $\mathcal{U}^T$ is a set of m shapelets such that $\mathcal{U}^T = \{w_1, w_2, \dots, w_m\}$.*

Note that the total length of the time series from which templates are generated must be substantially greater than the window size ω. Figure 6.3 simplifies the process whereby a profile $\mathcal{U}^T$ is generated. In this example, the windows are non-overlapping and abutting, this does not have to be the case, but this was the mechanism adopted with respect to the evaluation presented later in this chapter.

The profiles stored in the enrolment databases, as noted above, are also used to derive a bespoke similarity threshold, σ, for each user. This is calculated by comparing all shapelets within a profile $\mathcal{U}^T$ using the DTW method, and obtaining an average warping distance $\bar{\Theta}$ which is then used as the value for σ:

$$\sigma = \bar{\Theta} = \frac{1}{|\mathcal{U}^T|} \sum_{i=2}^{|\mathcal{U}^T|} dtw(w_{i-1}, w_i) \tag{6.1}$$

6.2.3 Authentication

The IKCA verification process is conducted by comparing the most recent shapelet window with previous shapelets extracted during the typing session, as shown in Figure 6.2. On "start-up", as also shown in the figure, the subject's identity is first confirmed; in other words to confirm that the subject is who (s)he says (s)he is. As noted above this initial process is called start-up authentication. In this context, the start-up authentication is done by comparing, using DTW, the first shapelet window collected, w_1, with the relevant user template profiles in $\mathcal{U}^T$ (stored in the enrolment database) and obtaining an average similarity value (minimum warping distance). If the average similarity value is less than or equal to σ, the validation process proceeds accordingly. It has been shown that averaging the warping distances associated with a set of time series can lead to an effective and more accurate classification of streaming data than if only one warping

Algorithm 4 IKCA Process.

Input: f = sampling frequency, ω = shapelet window length, σ = similarity threshold
Output: Similarity "highlight" reporting

```
s1 = ∅
s2 = ∅
t = 0
isCollectingTS = false
loop
        ▷ Get next point
    p_t = Next point in data stream
        ▷ End of input stream
    if p_t == End of data stream marker then
        Exit loop
        ▷ Start collecting a subsequence
    else if remainder t/f ≡ 0 then
        s2 = {p_t}
        isCollectingTS = true
        ▷ Conduct comparison
    else if remainder t/ω ≡ 0 then
        s2 = noiseReduction(s2)
        if s1 = ∅ then
            Compare s2 with U
        else
            wd = dtw(s1,s2)
            if wd > σ then
                highlight
            end if
        end if
        s1 = s2
        isCollectingTS = false
        ▷ Continue collecting subsequence
    else if isCollectingTS then
        s2 = s2 ∪ p_t
    end if
        ▷ Increment counter
    t = t + 1
end loop
```

distance is considered [116]. Each subsequent shapelet window w_k (where $k > 1$) is then compared with the preceding, previously collected, shapelet window w_{k-1} again utilising DTW; in this way changes in typing behaviour can be detected.

6.3 The IKCA Process

The IKCA process is presented in more detail in this section. From the preceding, the fundamental idea is that, as typing proceeds, we repeatedly collate shapelets of length ω and compare, on start-up, with $\mathcal{U}^T$; and then, once the process is underway, with the previously collected shapelet to the current shapelet. The process is given by the pseudo code in Algorithm 4. The algorithm takes as input: (i) the frequency f with which keystroke time series shapelets are collected, (ii) the length ω of a collected shapelet and (iii) a similarity threshold σ value. Note that f can be set so that either: shapelets overlap ($f < \omega$), shapelets abut up against each other ($f = \omega$) or shapelets are separated by additional keystrokes ($f > \omega$). For the evaluation presented in Section 6.4, $f = \omega$ was used (as illustrated in Figure 6.3).

On start-up the sets s_1 and s_2 are initialised to $\emptyset$ (lines 1 and 2), a keystroke counter t is initialised with the value 0 (line 3) and a Boolean flag $isCollectingTS$, indicating whether a shapelet is in the process of being collected or not, initialise to the value $false$. The procedure then operates on a continuous loop. On each iteration, a single key press is processed; there are five options: (i) stop processing as the end of the data stream has been reached (lines 7 and 8), (ii) start collating a new shapelet (lines 9 to 11), (iii) shapelet collation is complete therefore test the shapelet (lines 12 to 23), (iv) add the current keystroke to the current shapelet (lines 24 and 25) or (v) do nothing (the process is between shapelet collection phases). At the end of each iteration, the keystroke counter is incremented by one (line 27).

Whenever a shapelet window of length ω has been obtained, the algorithm tests whether the shapelet contains noise using the function *noiseReduction*() (line 13). Recall that the noise reduction is only applicable with respect to the flight time, $\mathcal{F}^t$, feature (Section 4.3, in Chapter 4). If there is no previous shapelet window, start-up authentication takes place (lines 14 and 15). Otherwise, the current shapelet is compared to the previously collected shapelet (line 17 to 19).

6.4 IKCA Evaluation

The outcomes of the evaluation of the proposed IKCA approach are presented in this section. A series of experiments were conducted to evaluate the IKCA approach to determine how well it performed in terms of subject authentication and, conversely, the detection of impersonators. The objectives of the evaluation were:

1. **Enrolment Database Generation Time**: To evaluate the processing time required to generate a user typing template, $\mathcal{U}^T$ and an associated threshold value σ using a range of ω and φ values.

2. **Authentication Performance**: To determine how well the IKCA approach performed concerning subject authentication; again using a range of ω and φ values.

3. **Comparison with Feature Vector Approach**: To compare the operation of IKCA approach with the established feature vector based approach, where feature vectors are generated using n-graphs (as described in [54]), for keystroke continuous authentication.

The IKCA evaluation was conducted with respect to both U-KTS and M-KTS so as to analyse which representation yielded the better results. Although U-KTS can be presented in either $\mathcal{F}^t$ or $\mathcal{KH}^t$ feature, in the evaluation presented here $\mathcal{F}^t$ was used. The reason for this was that with respect to the OKCA system (given in Chapter 5), it was demonstrated that the use of $\mathcal{F}^t$ outperformed $\mathcal{KH}^t$. This was also found to be the case with respect to IKCA operation. However, for completeness, the U-KTS with $\mathcal{KH}^t$ results are presented in Appendix B, for further inspection and comparison with the results of U-KTS with $\mathcal{F}^t$.

The evaluation experiments were conducted using the three data sets presented in Chapter 3 (ACB, GP and VHHS). However, as in the case of the evaluation of the OKCA system, the GP data set was not used for U-KTS with $\mathcal{KH}^t$, nor for M-KTS, as the $\mathcal{KH}^t$ feature was not present in this data set. Nevertheless, the metrics used for the evaluation were: (i) Authentication accuracy (Acc.), (ii) False Match Rate (FMR) and (iii) False Non-Match Rate (FNMR). Recall that FMR and FNMR are the standard metrics used to measure the performance of Biometric systems [154], although some researchers, in the literature, have used the terms FMR (False Acceptance Rate) and FRR (False Rejection Rate) instead.

The remainder of this section is organised as follows. The experimental setting is presented in Sub-section 6.4.1. The results with respect to each of the above objectives are then discussed Sub-sections 6.4.2, 6.4.3 and 6.4.4 respectively.

6.4.1 Experimental Setting

For the evaluation, for each subject in each data set, the records were split into two so that one-half could be used to create the enrolment database, or simply to construct $\mathcal{U}^T$, and the other to simulate a typing stream. Thus the size of $\mathcal{U}^T$ (number of m shapelets) depends on the defined ω and the length (n) of one-half record, for each subject, such that $m = \lfloor \frac{n}{\omega} \rfloor$.

The simulation was achieved by constructing a "test harness" whereby the keystroke data was released in a manner that simulated real-time keyboard activity. Real-life experiments were not conducted as the author wished to repeatedly use the same keyboard input data so that comparisons could be made. Wherever applicable two-fold cross-validation was conducted (using different halves of the keystroke time series to generate the enrolment database). Hence the results presented in the following subsections are average results from the two cross-validations. For comparison with the established feature vector based approach, the vector approach was re-implemented especially. All experiments were conducted using a 3.2 GHz Intel i5 processor with 24 GB RAM.

6.4.2 Enrolment Database Generation Time

The first evaluation objective was to analyse the processing time required to generate an enrolment database, including the associated σ threshold value. Tables 6.1 and 6.2

give the run-time values (seconds) obtained using a range of ω values (shapelet sizes) and $\mathcal{F}^t$ U-KTS and M-KTS respectively. The "per subject" values, in each table, were obtained by dividing the total run time with the number of subjects (records) in the relevant data set (see Table 3.4 given in Chapter 3). Recall that the reason the GP data set is not included in Table 6.2 was the $\mathcal{KH}^t$ feature was not available in this case.

From the tables, it can be seen that, regardless of the KTS representation used, processing time increased with corresponding increases in the size of ω. This was to be expected because the computation time required by the DTW would increase as the size of the subsequences (ω) considered increased (even though there might be less of them). The reason the processing time in the context of M-KTS was worse than in the case of U-KTS was also because the computational cost of DTW with respect to M-KTS was more expensive. Figure 6.4 provides two comparisons, one for data set ACB and one for data set VHHS, of run time against ω. Inspection of the figure indicates that, as was anticipated, the use of the U-KTS was more efficient than the use of M-KTS.

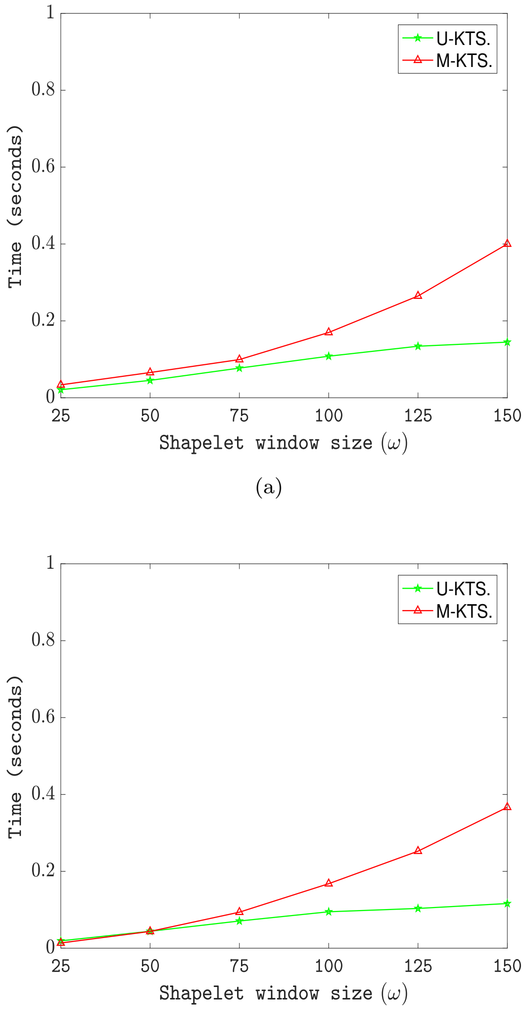
Regardless of whether U-KTS or M-KTS was used, the results show that the run time requirements for enrolment database generation, using the time series based approach, are within acceptable limits in which the authentication can be conducted within seconds (an acceptable limit for authentication whether in the context of online learning or other kind of application). From the results, it can also be seen that the time to compare two shapelets will be less than the time required to compare m shapelets (where $m > 2$). Therefore, it can be argued here that the proposed IKCA approach can be effectively used for iterative authentication.

TABLE 6.1: Enrolment database generation run time (seconds) using $\mathcal{F}^t$ U-KTS.

ω	Entire data set			Average per Subject		
	ACB	GP	VHHS	ACB	GP	VHHS
25	0.629	0.918	0.895	0.021	0.030	0.023
50	1.356	1.761	1.716	0.045	0.057	0.044
75	2.316	2.760	2.476	0.077	0.089	0.064
100	3.243	3.696	3.333	0.108	0.119	0.086
125	4.022	4.034	3.998	0.134	0.130	0.103
150	4.324	4.727	4.534	0.144	0.153	0.116

TABLE 6.2: Enrolment database generation run time (seconds) using M-KTS.

ω	Entire data set		Per Subject	
	ACB	VHHS	ACB	VHHS
25	1.007	1.180	0.034	0.030
50	1.969	1.570	0.066	0.040
75	2.980	3.523	0.099	0.090
100	5.095	6.244	0.170	0.160
125	7.940	9.742	0.265	0.250
150	12.000	14.130	0.400	0.362



(a)

(b)

FIGURE 6.4: Enrolment database generation comparison using $\mathcal{F}^t$ U-KTS and M-KTS: (a) ACB data set, (b) VHHS data set.

6.4.3 Authentication Performance

To evaluate the effectiveness of iterative subject authentication (in terms of impersonation detection) using the proposed IKCA system, for each data set and each subject, the continuous typing process was simulated by presenting the keystroke dynamics for each subject in the form of a data stream. In each case, the data stream was appended to a randomly selected second data stream from another subject. The idea was to simulate one subject being impersonated by another halfway through a typing session. For every comparison of every shapelet window w_i with a previously stored shapelet window w_{i-1} (line 17 in Algorithm 4, and both of length ω) it was determined whether the result was a True Positive (TP), a False Positive (FP), a False Negative (FN) or a True Negative (TN). In this manner, a *confusion matrix* was built up from which accuracy (Acc.), False Match Rate (FMR) and False Non-Match Rate (FNMR) could be calculated (using Equations 6.2, 6.3 and 6.4 below).

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{6.2}$$

$$FMR = \frac{FP}{FP + TN} \tag{6.3}$$

$$FNMR = \frac{FN}{FN + TP} \tag{6.4}$$

The experiments were run using the same range of ω values as used previously and a range of φ values from 0.75 to 2.00 incrementing in steps of 0.25. Each experiment was also run twice, each time using a different half of the data for authentication (the other half being used for user typing template generation); the results presented are therefore averages.

In the context of the U-KTS with $\mathcal{F}^t$ representation, Figure 6.5 gives the accuracy results obtained, in the form of 3-D histograms, one per data set. In the figure, the vertical axis indicates accuracy while the horizontal axises represent the shapelet window size (ω) and the limit value (φ). From the figure, it can be seen that $\omega = 100$ tended to produce best accuracy results for all three data sets; the red bars in the figure shows the best results with $\omega = 100$ across a range of φ values. From the figures it can also be observed that the selection of φ does not seem to have as much impact as the selection of ω; best accuracy results were recorded with $\varphi = 1.25$. It can also be observed from the figure that accuracy tends to decrease if the selected ω value was too small; in contrast, the performance tended to stabilise when $\omega > 100$. This meant that a best performance could be obtained using relatively short keystroke shapelets. Overall, the best accuracy results obtained (at: $\omega = 100$ and $\varphi = 1.25$) were 96.20%, 95.17% and 94.83% for ACB, GP and VHHS data sets respectively.

With respect to FMR and FNMR, the obtained results are presented in Tables 6.3 and 6.4 with respect to each data set; the best-recorded values are displayed in bold font. From the tables, it can be seen that the best FMR and FNMR values were obtained using ω settings of 100, 125 and 150, and $\varphi = 1.25$. The best recorded FMR results were 0.580, 0.731 and 1.045 whereas the best FNMR results were 1.970, 1.895 and 2.020 for ACB, GP and VHHS respectively. Note that these results were obtained at $\omega = 100$ and $\varphi = 1.25$; the same values as obtained in the case of Acc.

In the context of the M-KTS, Figure 6.6 presents the accuracy results obtained, again in the form of 3-D histograms. Recall that in this context the GP data set was not included. It should also be noted that, in the same way as the experiments directed at U-KTS, the experiments were run twice, each time with a different half of the data; the results presented are therefore averages. From the figure, it can be observed that a ω value of 125 tended to produce best results for all the data sets (highlighted in red); the best accuracy result recorded for ACB was 98.39%, whereas the best recorded result for VHHS was 97.32%, both with $\varphi = 1.25$. Similar to what has been observed in the case

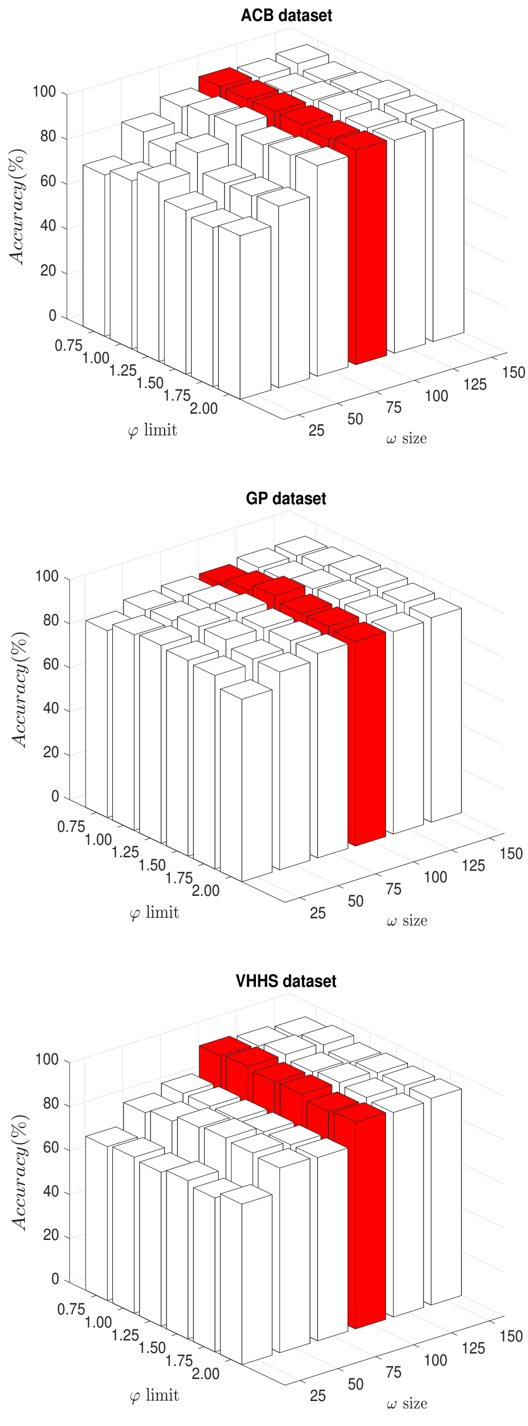


FIGURE 6.5: 3-D histograms showing recorded accuracy against a range of values for ω (shapelet window size) and φ ($\mathcal{F}^t$ limit), for ACB, GP and VHHS, when using U-KTS with $\mathcal{F}^t$.

of U-KTS, the selection of φ did not seem to have as much impact on accuracy as the selection of ω, and that the accuracy tended to decrease if the selected ω value was too small. With regards to FMR and FNMR, Tables 6.5 and 6.6 show the obtained results; the best results are highlighted in bold font. From the tables, it can be observed that

Table 6.3: The obtained **FMR** values (%) using different settings of ω and φ, in the context U-KTS, for the ACB, GP and VHHS data sets.

	ω \ φ	0.75	1.00	1.25	1.50	1.75	2.00
ACB	25	0.668	0.658	0.598	0.658	0.737	1.011
	50	0.665	0.655	0.595	0.655	0.734	0.948
	75	0.661	0.651	0.591	0.651	0.73	0.944
	100	0.656	0.646	**0.580**	0.59	0.725	0.939
	125	0.655	0.645	0.599	0.645	0.724	0.938
	150	0.648	0.638	0.586	0.638	0.717	0.931
GP	25	0.742	0.764	0.74	0.74	0.741	0.742
	50	0.743	0.743	0.739	0.739	0.74	0.74
	75	0.74	0.739	0.739	0.738	0.739	0.74
	100	0.736	0.735	**0.731**	0.733	0.738	0.735
	125	0.738	0.739	0.735	0.734	0.734	0.736
	150	0.739	0.738	0.736	0.735	0.735	0.736
VHHS	25	1.267	1.228	1.048	1.078	1.071	1.092
	50	1.178	1.125	1.045	1.075	1.068	1.089
	75	1.106	1.107	1.047	1.077	1.07	1.091
	100	1.077	1.076	**1.045**	1.075	1.078	1.078
	125	1.075	1.068	1.06	1.068	1.061	1.062
	150	1.075	1.077	1.069	1.077	1.07	1.071

Table 6.4: The obtained **FNMR** values (%) using different settings of ω and φ, in the context U-KTS, for the ACB, GP and VHHS data sets.

	ω \ φ	0.75	1.00	1.25	1.50	1.75	2.00
ACB	25	2.000	2.070	2.025	2.030	2.025	2.029
	50	2.020	2.090	1.990	1.980	1.980	1.990
	75	1.992	2.022	1.990	1.980	1.984	1.980
	100	1.989	2.059	**1.970**	1.972	1.980	1.980
	125	1.995	2.065	1.979	1.979	1.979	1.980
	150	1.980	2.050	1.978	1.976	1.979	1.981
GP	25	1.956	1.971	1.901	1.960	1.961	1.936
	50	1.951	1.966	1.955	1.955	1.956	1.961
	75	1.950	1.965	1.950	1.954	1.955	1.954
	100	1.901	1.897	**1.895**	1.897	1.896	1.896
	125	1.957	1.942	1.940	1.942	1.942	1.945
	150	1.949	1.945	1.943	1.948	1.948	1.943
VHHS	25	2.133	2.131	2.052	2.077	2.068	2.06
	50	2.035	2.033	2.033	2.034	2.034	2.035
	75	2.027	2.025	2.025	2.027	2.026	2.026
	100	2.021	2.022	**2.020**	2.023	2.023	2.024
	125	2.027	2.025	2.024	2.024	2.025	2.025
	150	2.031	2.029	2.029	2.03	2.03	2.029

best FMR and FNMR were recorded using a combination of $\omega = 125$ and $\varphi = 1.50$.

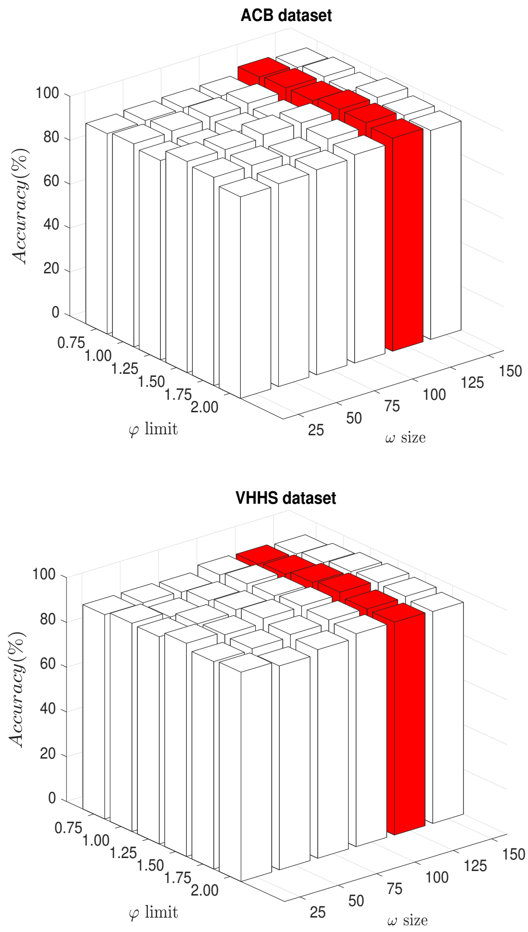


FIGURE 6.6: 3-D histograms showing recorded accuracy against a range of values for ω (shapelet window size) and φ (the $\mathcal{F}^t$ limit), for ACB and VHHS, using M-KTS.

Overall, the performance of M-KTS tended to be more accurate than U-KTS with $\mathcal{F}^t$. For comparison purposes, the best obtained results for all metrics, using the IKCA system with respect to $\mathcal{F}^t$ U-KTS and M-KTS are summarised in Table 6.7. The table also gives the average (Avg.) and Standard Deviation (SD) in each case.

6.4.4 Comparison with Feature Vector Approach

This sub-section reports on the results obtained when the operation of the proposed IKCA approach was compared with a Feature Vector Representation (FVR) based approach of the form frequently referenced in the literature. As noted in Chapter 2 (Subsection 2.3.5), there are a number of reports where FVRs, made up of keystroke dynamics related to n-graphs, have been applied to the keystroke continuous authentication problem, although not in an iterative manner (authentication was conducted on typing completion, thus static authentication as in the case of the OKCA systems). Of these, the approach described in [54] was selected (as a baseline) for the comparison reported on here for the following reasons:

Table 6.5: **FMR** values (%) using different settings of ω and φ, in the context of M-KTS, for the ACB and VHHS data sets.

	ω \ φ	0.75	1.00	1.25	1.50	1.75	2.00
ACB	25	0.051	0.052	0.049	0.048	0.048	0.048
	50	0.049	0.049	0.047	0.046	0.046	0.047
	75	0.048	0.0.47	0.048	0.048	0.049	0.049
	100	0.047	0.046	0.046	0.046	0.046	0.046
	125	0.047	0.046	0.046	**0.045**	0.046	0.046
	150	0.046	0.046	0.046	**0.045**	0.046	0.047
VHHS	25	0.031	0.030	0.031	0.030	0.030	0.031
	50	0.031	0.032	0.032	0.032	0.032	0.032
	75	0.030	0.030	0.030	0.030	0.030	0.030
	100	0.029	0.029	0.028	0.028	0.028	0.028
	125	0.028	0.028	0.028	**0.027**	0.028	0.028
	150	0.029	0.029	0.029	0.029	0.029	0.030

Table 6.6: **FNMR** values (%) using different settings of ω and φ, in the context of M-KTS, for ACB and VHHD data sets.

	ω \ φ	0.75	1.00	1.25	1.50	1.75	2.00
ACB	25	1.167	1.167	1.101	1.101	1.101	1.101
	50	1.099	1.099	1.099	1.099	1.099	1.099
	75	1.097	1.097	1.096	1.096	1.973	1.096
	100	1.096	1.096	1.096	1.096	1.096	1.096
	125	1.095	1.094	1.094	**1.093**	1.094	1.094
	150	1.096	1.095	1.095	1.095	1.095	1.095
VHHS	25	1.100	1.100	1.100	1.099	1.099	1.100
	50	1.098	1.098	1.097	1.097	1.097	1.098
	75	1.097	1.097	1.096	1.096	1.097	1.097
	100	1.097	1.096	1.096	1.096	1.096	1.096
	125	1.096	1.096	1.096	**1.095**	1.096	1.096
	150	1..097	1.097	1.097	1.096	1.096	1.097

1. The study obtained, to the best knowledge of the author, the best reported FMR and FNMR (although referred to as FAR and FRR) results to date.

2. The data set used for the study was publicly available (although with some records missing).

3. The approach was well explained in the literature; therefore it was easy to reproduce.

Because, as noted above, the study reported in [54] was conducted in 2005 it was considered inappropriate to compare directly with the reported results; instead, the approach was re-implemented and re-run in an iterative manner, as described for the IKCA proposed system, so that meaningful comparisons could be made. The features used were

Table 6.7: Summarisation of IKCA results in the context of U-KTS with $\mathcal{F}^t$, and M-KTS using the ACB, GP and VHHS data sets.

Data set	U-KTS			M-KTS		
	Acc.	FMR	FNMR	Acc.	FMR	FNMR
ACB	96.20	0.580	1.970	98.39	0.045	1.093
GP	96.31	0.731	1.895	x	x	x
VHHS	94.83	1.045	2.020	97.32	0.027	1.095
Avg.	95.78	0.785	1.962	97.85	0.036	1.094
SD	0.825	0.237	0.063	0.757	0.013	0.001

Table 6.8: Summarisation of IKCA results using: (i) U-KTS, (ii) M-KTS and (iii) FVR across the ACB, GP and VHHS data sets.

Data set	U-KTS			M-KTS			FVR		
	Acc.	FMR	FNMR	Acc.	FMR	FNMR	Acc.	FMR	FNMR
ACB	96.20	0.580	1.970	98.39	0.045	1.093	80.82	8.60	11.10
GP	96.31	0.031	1.094	x	x	x	90.15	6.05	2.06
VHHS	94.83	1.040	2.049	97.32	0.027	1.095	89.15	9.09	7.13
Avg.	95.78	0.550	1.704	97.85	0.036	1.094	86.70	7.918	7.422
SD	0.825	0.505	0.530	0.757	0.013	0.001	5.122	1.634	4.748

the $\mathcal{F}^t$ values for all shared digraph, trigraphs and quadgraphs in each data set. Authentication was otherwise conducted for each subject (user) as described in [54], see also Sub-section 2.3.5. The metrics used were, again, accuracy (Acc.), FMR and FNMR.

The results are presented in Table 6.8. For comparison purposes, the table includes the Acc., FMR and FNMR results presented earlier with respect to the proposed U-KTS and M-KTS representations in Table 6.7. From Table 6.8, it can be observed that the proposed IKCA system produced much better performances than the feature vector-based approach, with respect to all three data sets. This, therefore, confirmed the hypothesis posed by the author, in the introduction to the thesis, that a time series representation would serve to more effectively encapsulate keystroke dynamics than the feature vector based approach that has typically been adopted to date. Furthermore, from the table, it can be observed that a better performance was obtained using M-KTS than U-KTS with respect to all three data sets. In other words, M-KTS serves to better encapsulate keystroke dynamics than U-KTS with $\mathcal{F}^t$ and the feature vector based approach that has typically been used to date. For completeness, Figure 6.7 presents the best accuracy results in a graphical form.

It should also be noted here that the FMR and FNMR result obtained for the GP data set, of 6.05% and 2.06% respectively, are worse than the 5.00% and 0.005% reported in the original study [54], despite using the same mechanism. It was conjectured that this was because of the change in the data set size that had occurred since the original study and that reported on here, where the original data set consisted of 40 subjects, but the publicly available version only had 31 subjects. Also, in [54], an additional 165 samples that served as imposter samples were used; however, they were not used in the

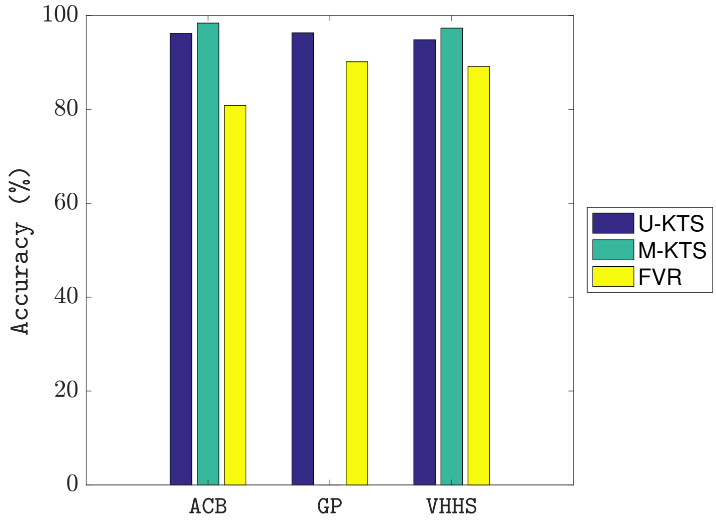


FIGURE 6.7: The accuracy results obtained using U-KTS, M-KTS and FVR with respect to the IKCA approach.

evaluation given in this chapter (the reasons for not including the imposter samples in the GP data set were given in Sub-section 3.4.2 of Chapter 3).

Moreover, with respect to the efficiency, where it is desirable to construct $\mathcal{U}^T$ and verification, a better efficiency was obtained in the case of IKCA than FVRs. According to [54] the run-time for verifying a single typing sample was given as 140 seconds, although in the case of the re-implementation, the run time was approximately 120 seconds due to the processing power used not being the same as that used with respect to that used to obtain the results reported on in [54]. Using the same data set (GP), in the IKCA approach a single verification takes less than 0.362 seconds to compute (in the worst case) thus, we have a worst case runtime ratio 1 : 23; a significant "speed-up". The reason for the speed up when using the proposed time series analysis approach was that there was no need to search through the time series to identify the n-graphs of interest or any requirement for the subsequent statistical feature calculation.

6.5 Summary

This chapter has presented a novel approach to keystroke continuous authentication, namely the Iterative Keystroke Continuous Authentication (IKCA) system. Intuitively, IKCA serves as a biometric recognition system whereby impersonation can be effectively detected. The essential advantage of the IKCA, in comparison with the established feature vector representations, was that it was more accurate and more efficient, which in turn meant that it would be suitable for authentication in the context of the online assessments frequently used in eLearning and MOOC environments.

The IKCA system operated in an iterative manner as typing proceeds by extracting keystroke time series subsequences called shapelets. Each extracted shapelet, of length

ω, is compared to the previously extracted shapelet except on start-up, when the first shapelet is compared with the content of a previously constructed enrolment database for the subject. For the system, DTW similarity comparison was adopted, although other techniques might also be appropriate.

The evaluation demonstrated that M-KTS produced a better performance than U-KTS with $\mathcal{F}^t$, although both approaches improved upon the Feature Vector based Representation (FVR) approach. However, the efficiency of M-KTS was not as good as that of U-KTS. The evaluation was also conducted using a range of values for the shapelet size, ω, and the noise reduction limit, φ, to determine the effect of these parameters on IKCA performance. The reported results showed that $\omega = 100$ and $\varphi = 1.25$ were the best values that can be used for IKCA.

In the next chapter, the KCASA approach is presented, an extension of the IKCA approach presented in this chapter. The motivation for the KCASA approach was to improve both the authentication performance and the efficiency of the authentication, by transforming the keystroke time series data from the time domain to the spectral domain.

Chapter 7

Keystroke Continuous Authentication based Spectral Analysis

7.1 Introduction

In the preceding chapter, a novel approach for keystroke continuous authentication was introduced, the Iterative Keystroke Continuous Authentication IKCA approach. The IKCA approach used time series analysis techniques to recognise typing patterns from free text in a manner suited to iterative continuous authentication. The keystroke data was processed in its raw form as a continuous data stream; periodically this data stream was sampled and shapelets extracted which were then analysed for authentication purposes by comparing to either a previously extracted shapelet or, on start-up, to a bank of shapelets that are known to belong to the claimed user. The data streams (shapelets) were considered either in terms of flight time ($\mathcal{F}^t$) or hold time ($\mathcal{KH}^t$) or both; thus either: (i) Univariate Keystroke Time Series (U-KTS) or (ii) Multivariate Keystroke Time Series (M-KTS). However, a limitation of the approach was that it became less efficient as the authentication accuracy increased. This was clearly observed when comparing the effectiveness and efficiency of using M-KTS with using U-KTS. Moreover, although IKCA was shown to be a good biometric approach, it was thought that there was potential to enhance the authentication accuracy using time series analytics with keystroke dynamics.

Consequently, it was conjectured that better results might be obtained if the time series data (shapelets) were transformed from the temporal domain to the spectral domain. The intuition behind this idea was the observation that with respect to other time series applications, such as time series indexing [45] and time series pattern extraction [146], usage of the spectral domain could significantly improve the analysis in terms of both speed and accuracy (see also [2, 27, 28, 77, 159, 164]). The idea was thus built into an alternative keystroke continuous authentication approach, the Keystroke Continuous Authentication based Spectral Analysis (KCASA) approach.

The proposed KCASA approach was also motivated by the desire to provide an answer to the following subsidiary research question given in Chapter 1:

- What is the most effective process whereby continuous, real-time, authentication can be conducted in the context of online assessment and more generally?

Providing an answer to this subsidiary research question would contribute to addressing the main research question also postulated in the introductory chapter of this thesis. Note that to the best knowledge of the author, the idea of spectral analysis in the context of keystroke continuous authentication (free text) has not been considered previously.

The proposed KCASA approach is fully described in this chapter. Two types of spectral transform were considered within the context of the proposed KCASA approach: (i) Discrete Fourier Transformation (DFT) and (ii) Discrete Wavelet Transform (DWT). Also, as in the case of the IKCA approach, variations directed at both U-KTS and M-KTS were considered. The remainder of this chapter is organised as follows. Section 7.2 details the basic framework and fundamentals of the KCASA approach. Algorithmic detail concerning the KCASA approach is then given in Section 7.3. This is followed by Section 7.4 where the evaluation of the proposed KCASA approach is presented. In Section 7.5 some discussion is provided concerning the evaluation, whilst Section 7.6 summarises and concludes the chapter.

7.2 KCASA Framework

The generic framework for the KCASA approach is presented in this section. Figure 7.1 illustrates the fundamental framework in terms of its steps:

1. Shapelet Extraction.
2. Noise Reduction.
3. Transformation.
4. Similarity Comparison.
5. Template Construction.

Each of the above steps is considered in further details in the following five sub-sections, Sub-section 7.2.1 to 7.2.5.

7.2.1 Shapelet Extraction

As in the case of the IKCA approach, the proposed KCASA approach operated by periodically extracting shapelets from the input data stream using a windowing mechanism.

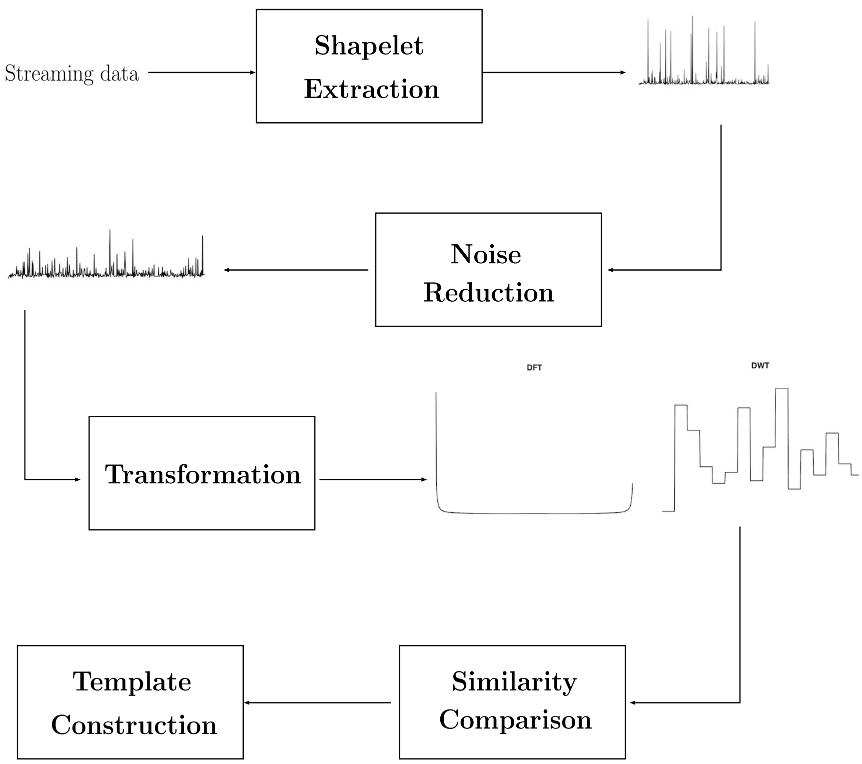


FIGURE 7.1: The main components within the KCASA framework.

In Chapter 4 it was noted that the process of typing produces a keystroke time series $\mathcal{K}_{ts}$, such that $\mathcal{K}_{ts} = \{\rho_1, \rho_2, \dots, \rho_i, \dots, \rho_n\}$ where ρ_i is a parametrised point that represents a typing event, $\mathcal{F}^t$ and/or $\mathcal{KH}^t$, and n is the length of $\mathcal{K}_{ts}$. Using the widowing mechanism, an order collection of keystroke time series subsequences (shapelets) was produced $\{s_1, s_2, \dots\}$ such that for each shapelet $s_j = \{\rho_i, \rho_{i+1}, \dots, \rho_{i+\omega-n}\}$, i is some "start" point in $\mathcal{K}_{ts}$ and ω is the window length (and $\forall s_k \subset \mathcal{K}_{ts}$). The frequency with which shapelets were collected was again defined by a sampling rate parameter f (where $f = \omega$).

7.2.2 Noise Reduction

Prior to any further processing being conducted, as in the case of the IKCA approach, each newly collated shapelet was cleaned. As noted in Chapter 4 the issue here was that $\mathcal{F}^t$ values can be large, for example when the subject has paused typing or as a consequence of an "away from keyboard" event. As in the case of the IKCA approach a limit was placed on the $\mathcal{F}^t$ values using a maximum flight time threshold value φ. Given an $\mathcal{F}^t$ value in excess of φ, the value was replaced with the φ value. For the evaluation presented later in this chapter, a range of values for φ were considered, from 0.750 to 2.00 seconds, increasing in steps of 0.25 seconds (thus $\{0.75, 1.00, 1.25, 1.50, 1.75, 2.00\}$).

7.2.3 Transformation

The next component in the KCASA approach, and the novel element of the approach, was the transformation of the extracted shapelets to the spectral domain. As noted in the introduction to this chapter two spectral transforms were considered: (i) Discrete Fourier Transformation (DFT) and (ii) Discrete Wavelet Transform (DWT). Both are considered in further details in the following two sub-sections, Sub-sections 7.2.3.1 and 7.2.3.2.

7.2.3.1 DFT for Keystroke Time Series

The fundamental idea of the DFT is to transform a given time series (shapelet) from the temporal domain into the frequency domain. The resulting frequency-domain representation shows how much of a given signal lies within each given frequency band over a range of frequencies, we refer to the result as the Fourier spectrum. The fundamental benefit is that the DFT serves to compact the data without loosing any salient information [69]. The compression is conducted by first representing the time series as a linear combination of sinusoidal coefficients, and then computing the similarity between the transformed coefficients for any pair of corresponding signals.

Given a KCASA sequence of keystroke dynamics $s = \{\rho_1, \rho_2, \ldots, \rho_i, \ldots, \rho_\omega\}$, where ρ_i is some keystroke timing feature, and ω is the length of the subsequence, the DFT transform (S) for the subsequence s is given by:

$$S_m = \frac{1}{\sqrt{\omega}} \sum_{i=1}^{\omega} \rho_i \; e^{\frac{-j2\pi im}{\omega}} \;, \quad m = 0, 1, \ldots, \omega - 1 \tag{7.1}$$

where $j = \sqrt{-1}$ is the imaginary part. In more detail, the DFT transform typically compresses the subsequence s into a linear set of sinusoidal functions X with amplitudes p, q and phase w, thus the DFT for a given s can be alternatively written as in the following equation:

$$X = \sum_{i=1}^{\omega} (\rho_i Cos(2\pi w_i \rho_i) + q_i Sin(2\pi w_i \rho_i)) \tag{7.2}$$

Note that the time complexity to transform (each) s is $\mathcal{O}(\omega \; \texttt{log} \; \omega)$ using the Radix 2 DFT algorithm [31, 69].

Using the DFT transform, the obtained keystroke subsequence s is composed of a new magnitude (the amplitude of the discrete coefficients) and phase spectral shape in which can be compared with other transformed keystroke time series subsequences. Figures 7.2 and 7.3 illustrate the intuition behind the DFT as applied to keystroke time series within the context of the proposed KCASA approach. The figures show keystroke subsequences and their DFT transforms for two subjects taken from the ACB data set used throughout this thesis; Figure 7.2 for subject **A** and Figure 7.3 for subject **B**, where for each subject, two typing samples from unstructured text were provided. In

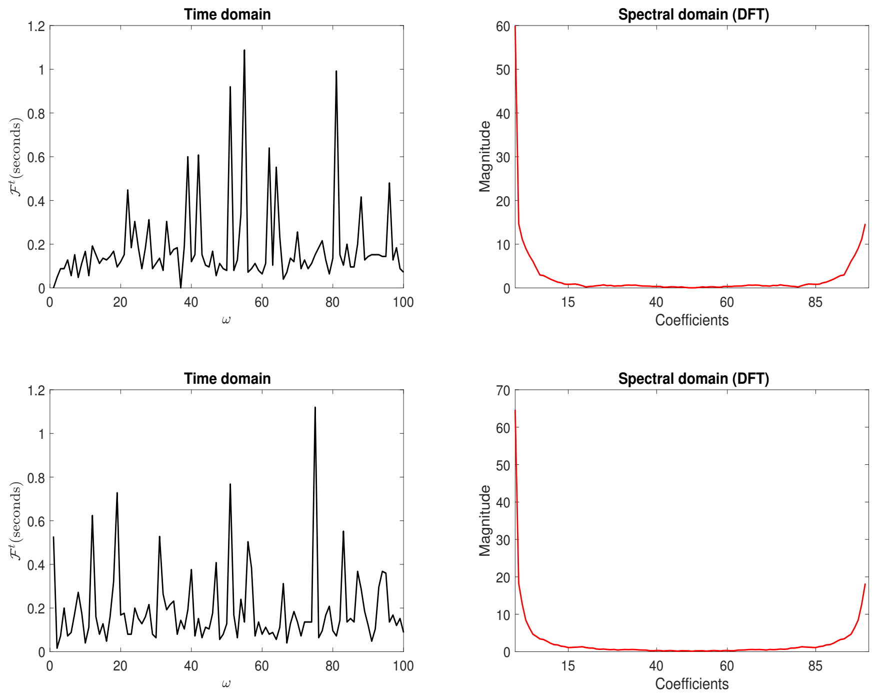


FIGURE 7.2: Example application of DFT for subject **A** from ACB data set.

each figure, on the left-hand side the raw, time domain, keystroke subsequences are shown, whilst on the right-hand side the DFT equivalents are shown. Form the figures, it can be noted that the DFT signals describe distinctive patterns of typing behaviour for the same subject.

7.2.3.2 DWT for Keystroke Time Series

The Discrete Wavelet Transform (DWT) is an alternative form of time series representations that considers the time series according to the frequencies that are present. DWT is sometimes claimed to provide a better transformation than DFT in that it retains more information [28]. DWT can be applied to time series according to different scales, orthogonal [57] and non-orthogonal [49]. For the work presented in this chapter the orthogonal scale was used, more specifically the well known Haar transform [57] as described in [28]. Fundamentally a Haar Wavelet is simply a sequence of functions which together form a wavelet comprised of a series of square shapes. The Haar transform is considered to be the simplest form of DWT; however, it has been shown to offer advantages with respect to time series analysis where the time series feature sudden changes. The transformation is usually described as in Equation 7.3 where, in the context of this thesis, x is a keystroke timing feature.

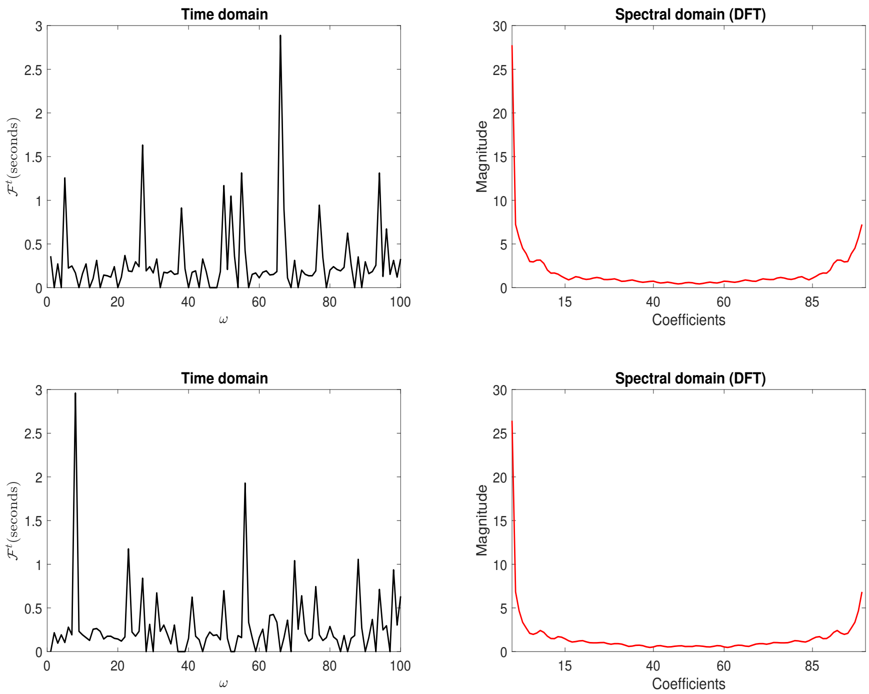


FIGURE 7.3: Example application of DFT for subject **B** from ACB data set.

$$\phi(x) = \begin{cases} 1, & \text{if} \quad 0 < t < \frac{1}{2} \\ -1, & \text{if} \quad \frac{1}{2} < t < 1 \\ 0, & \quad\;\; \text{otherwise} \end{cases} \tag{7.3}$$

The time complexity for the Haar transform is $\mathcal{O}(\omega)$ for each $\mathcal{K}_{ts}$. Note that in the context of the Haar transform, the length of a given time series should be an integral power of 2 [77], thus 2, 4, 8, 16 and so on. For further detail concerning the DWT interested readers are to referred to [24] and [42].

The principle of DWT, as adopted with respect to the KCASA approach, is illustrated in Figures 7.4 and 7.5. The figures show the DWT coefficients for keystroke subsequences obtained from two subjects, **A** and **B**; the same keystroke subsequences as given in Figures 7.2 and 7.3. The figures clearly show that DWT coefficients are distinctive in the context of keystroke data from the same subjects.

7.2.4 Similarity Comparison

As in the case of the IKCA approach, DTW as described in Chapter 4 was adopted for the purpose of measuring the similarity between DFT/DWT keystroke time series. Recall that using DTW a minimum warping path $\mathbb{P}$ is found from which a minimum warping distance Θ is extracted which is used as a similarity measure. Figures 7.6 and 7.7 showing two warping paths ($\mathbb{P}$) resulting from the the application of the DTW process to

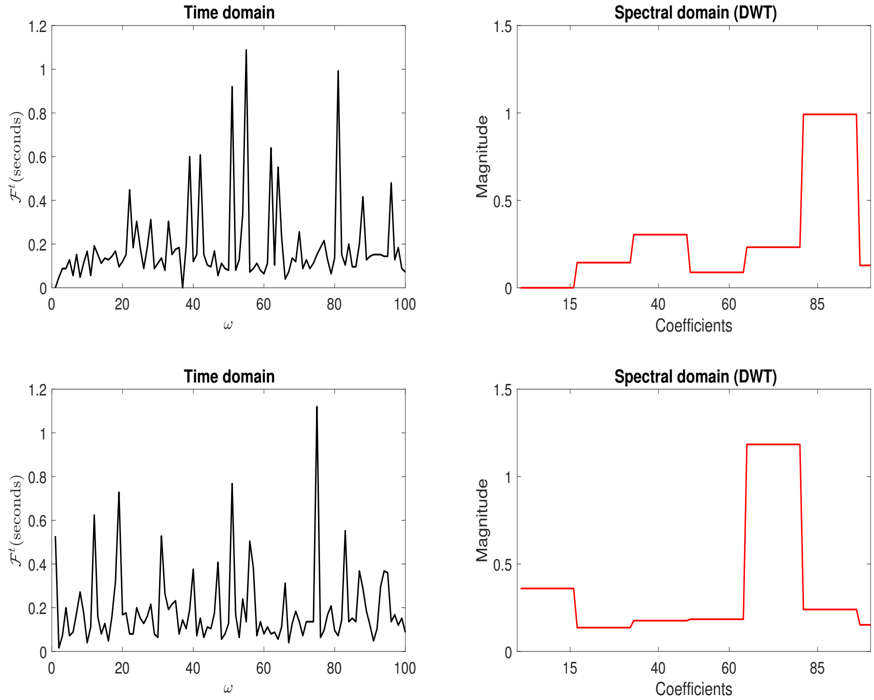


FIGURE 7.4: Example application of DWT for subject **A** from ACB data set.

DFT and DWT represented time series. Figures 7.6(a) and 7.7(a) show warping paths obtained when DTW was applied to time series for the same subject writing free text; Figures 7.6(b) and 7.7(b) when DTW was applied to time series from different subjects writing free text. From the figures, it can be seen that the minimum warping paths $\mathbb{P}$ associated with the same subject is closer to the diagonally than in the case of the minimum warping paths associated different subjects.

7.2.5 Template Construction

As in the case of the previous IKCA approach, the proposed KCASA approach uses a set of user profiles for initial authentication on start-up of a typing session. The bank of user profiles was constructed in much the same way as for the IKCA approach, except that they were stored in the form of DFT or DWT keystroke time series. Thus each subject has a user typing template $\mathcal{U}^T$ (as defined in Sub-section 6.2.2 of Chapter 6). As before a bespoke σ threshold was computed by summing the warping distances between the collected samples, for each subject, and calculating an average value.

7.3 KCASA Operation

The fundamental operation of the KCASA approach was similar to the IKCA approach, see Figure 7.8. Keystroke time series sub-sequences, of length ω, were sampled at regular

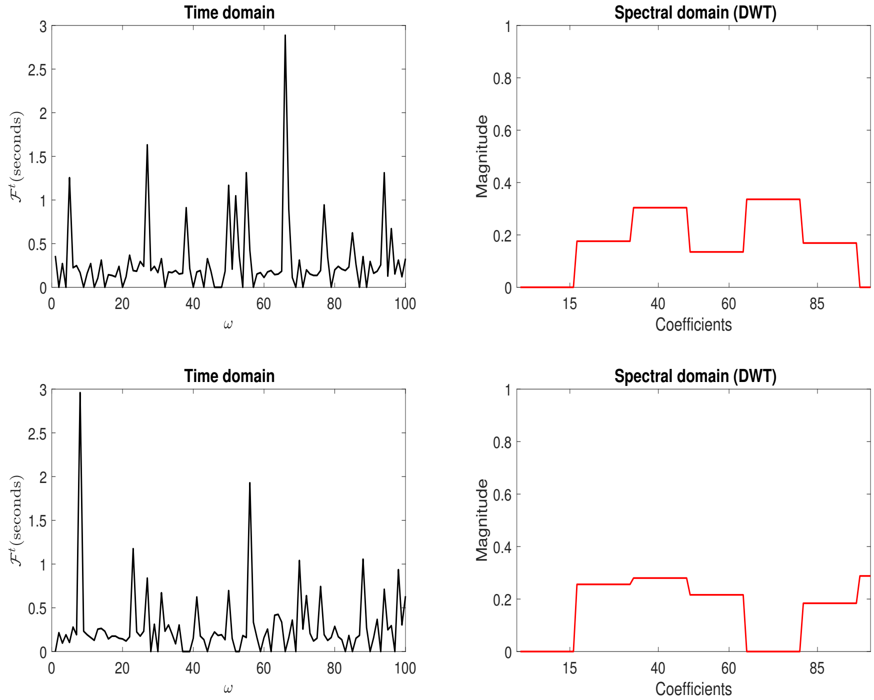
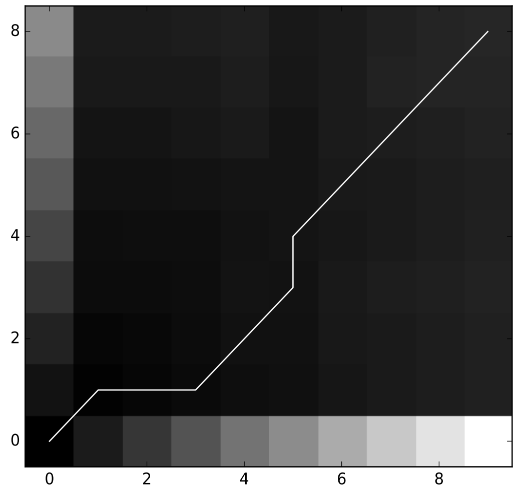


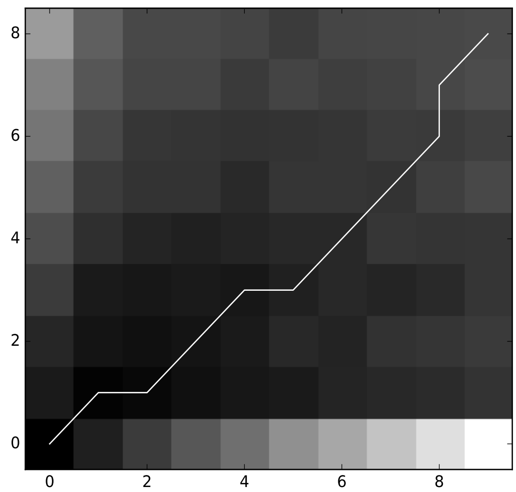
FIGURE 7.5: Example application of DWT for subject **B** from ACB data set.

intervals and transformed into the spectral domain. As before, the first collected sub-sequence, w_i, was compared with the subjects profile (see above), while each subsequent sub-sequence w_i was compared with the previous sub-sequence w_{i-1}.

The operation of the proposed KCASA approach is presented in the form of pseudo code in Algorithm 5. The algorithm takes as input: (i) the window size ω and (ii) the similarity threshold σ (derived as described above in Sub-Section 7.2.5) and (iii) a φ threshold for limiting the $\mathcal{F}^t$ feature. The process operates continuously, following a loop, until the typing session is terminated (the user completes the assessment, times out or logs-out) (lines 4-6). Values for ρ are recorded as soon as the typing session starts (line 7). Note that in the case of flight time the value will be checked, and if necessary reduced, according to φ (lines 8 to 10). The ρ value is then appended to the time series $\mathcal{K}_{ts}$. The *counter* is monitored and subsequences are extracted whenever ω keystrokes have been obtained. Each extracted subsequence w is then transformed into DFT or DWT as required. The first transformed time series sub-sequence ($w_1 \in \mathcal{K}_{ts}$), the start-up time series, is compared with the stored profile for the subject in question; while each subsequent sub-sequence w_i is compared, using DTW, with the previous w_{i-1} sub-sequence.



(a)



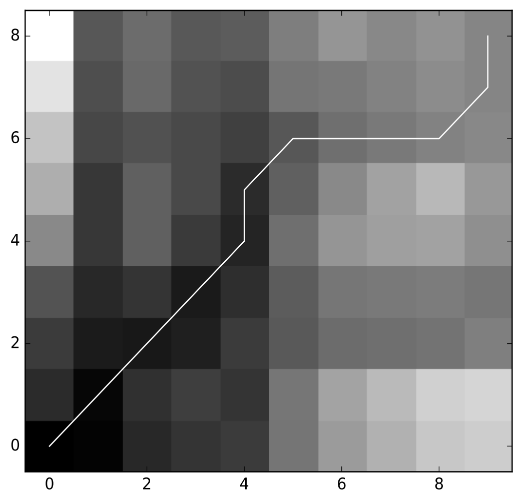
(b)

FIGURE 7.6: Examples of DTW applied to DFT data: **(a)** warping path obtained from the comparison of two DFT keystroke time series from the same subject typing free text, **(b)** warping path obtained from the comparison of two DFT keystroke time series from two different subjects typing free text.

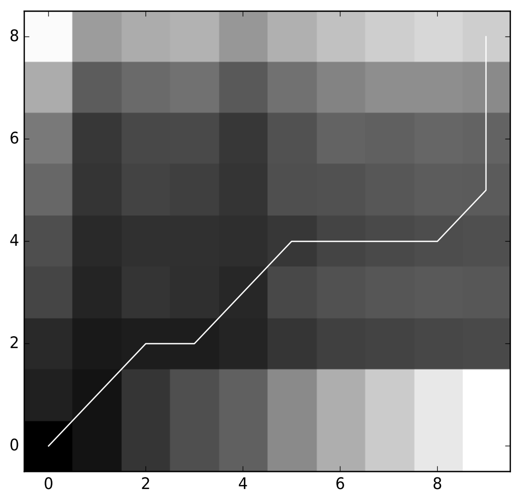
7.4 Evaluation

This section presents a review of the evaluation conducted with respect to the KCASA approach. The objectives of the evaluation were:

1. **Typing Template Construction Efficiency:** To determine the efficiency of constructing the typing templates using the proposed KCASA approach.



(a)



(b)

FIGURE 7.7: Example of DTW applied to DWT data: **(a)** warping path obtained from the comparison of two DWT keystroke time series from the same subject typing free text, **(b)** warping path obtained from the comparison of two DWT keystroke time series from two different subjects typing free text.

2. **Authentication Performance**: To evaluate the effectiveness of the iterative authentication, in terms of impersonation detection, using different values for ω (the sampling window size) and φ (the noise reduction threshold value).

The metrics used for the evaluation were: (i) Authentication accuracy (Acc.), (ii) False Match Rate (FMR) and (iii) False Non-Match Rate (FNMR).

The proposed approach was applied to both U-KTS and M-KTS using either DFT or DWT. In other words, four variations of the proposed KCASA approach were considered:

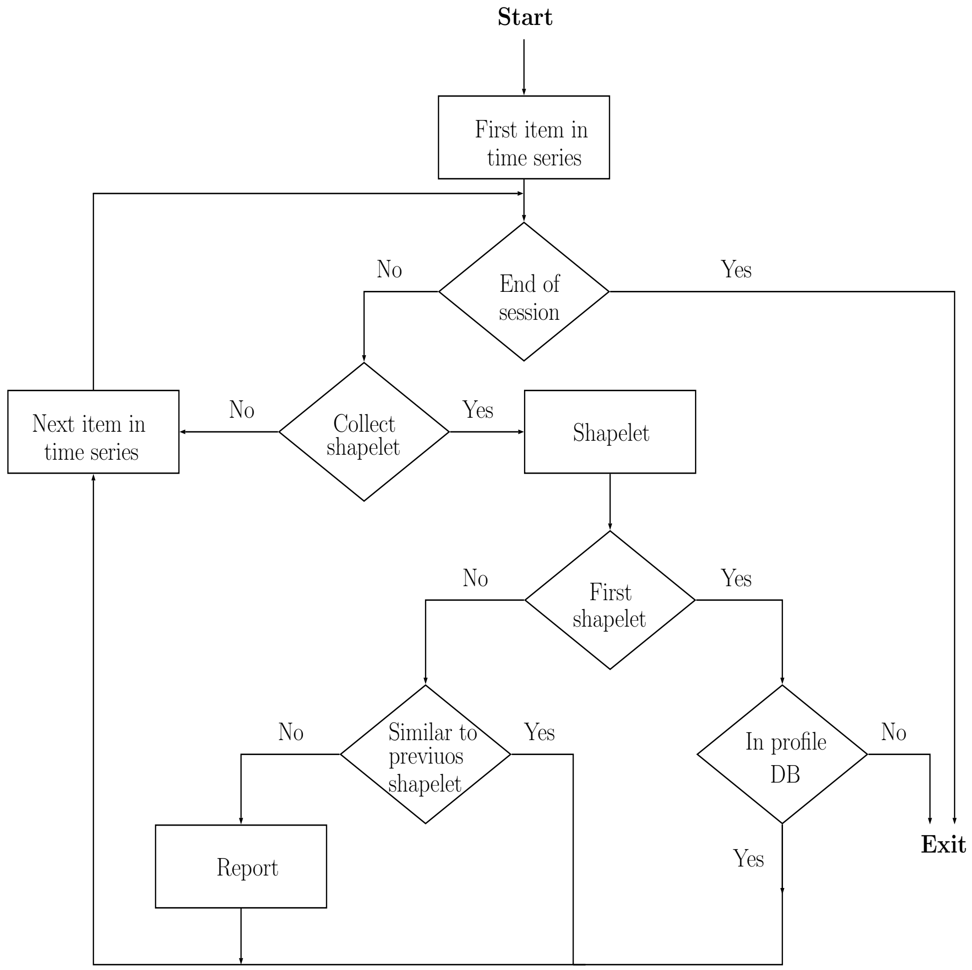


Figure 7.8: The KCASA subject verification process. Note that it is similar to the IKCA verification process proposed in Chapter 6.

1 **KCASA_DFT+U-KTS**: KCASA using DFT applied to U-KTS.

2 **KCASA_DWT+U-KTS**: KCASA using DWT applied to U-KTS.

3 **KCASA_DFT+M-KTS**: KCASA using DFT applied to M-KTS.

4 **KCASA_DWT+M-KTS**: KCASA using DWT applied to M-KTS.

Note that, in the context of the U-KTS representation, results using $\mathcal{F}^t$ are presented here because the evaluation presented in earlier chapters demonstrated that this produced better results than when $\mathcal{KH}^t$ was used. However, for completeness, the results obtained using U-KTS $\mathcal{KH}^t$ are presented in Appendix C.

The rest of this section is organised as follows. The experimental setup is detailed in Sub-section 7.4.1. The results with respect to the efficiency evaluation are then presented in Sub-section 7.4.2, whilst the results with respect to the authentication performance evaluation are presented in Sub-section 7.4.3.

Algorithm 5 KCASA Algorithm

Input: ω, σ, φ.
Output: Continuous authentication commentary.

$counter = 0$
$\mathcal{K}_{ts} = \emptyset$
loop
  if terminated signal received **then**
    break
  end if
  $\rho =$ keystroke feature (e.g. $\mathcal{F}^t$ or $\mathcal{KH}^t$)
  if $(\mathcal{F}^t \in \rho) > \varphi$ **then**
    $\rho = \varphi$ ▷ Noise reduction.
  end if
  $\mathcal{K}_{ts} = \mathcal{K}_{ts} \cup \langle counter, k \rangle$
  $counter + +$
  if $REM(counter/\omega) == 0$ **then**
    $w_i =$ subsequence $\{\mathcal{K}_{ts_{counter-\omega}} \ldots \mathcal{K}_{ts_{counter}}\}$
    if $counter = \omega$ **then** ▷ Start-up situation
      $Transform(w)$ ▷ Transform w to (DFT)/(DWT)
      Start-up: authenticate w_i w.r.t $\mathcal{U}^T$ and σ, and report
    else
      Authenticate w_i w.r.t. w_{i-1} and σ, and report
    end if
  end if
end loop

7.4.1 Experimental Setting

The evaluation was conducted using the three data sets given in Chapter 3: ACB, GP and VHHS. Recall that for the GP data set only the $\mathcal{F}^t$ feature was available; therefore the performance of KCASA using M-KTS could not be conducted using the GP data set.

As indicated in Sub-section 7.2.3.2 that DWT transform can only support time series data whose length is defined as an integral power of 2, thus for the evaluation the range of ω values considered was $\{16, 32, 64, 128, 256, 512\}$. The range of φ values considered was $\{0.750, 1.00, 1.25, 1.50, 2.00\}$ seconds.

As for the previously reported experiments conducted to evaluate the IKCA approach, each record in each data set was divided into two, the first half used to construct the typing template (profile), and the second half used for the continuous authentication evaluation. Two-fold cross-validation was therefore conducted; the results reported in the following sub-sections are thus averages. All experiments were conducted using a 3.2 GHz Intel i5 processor with 24 GB RAM.

7.4.2 Typing Template Construction Efficiency

Tables 7.1 and 7.2 present the average run-time complexity (seconds) for the construction of the typing template for each subject; in other words, the tables can be interpreted as the average time required to create the typing template for a single subject. More specifically, Table 7.1 reports the time complexity for KCASA applied to U-KTS, whilst Table 7.2 reports the time complexity for KCASA applied to M-KTS. From the tables, it can be seen that the time complexity increases as ω increases. This was to be expected, as noted in Chapter 6 (in the context of the IKCA approach), because the time complexity to compute the DTW increases as the value for ω increases. Nonetheless, the process of constructing the typing template gained considerable efficiency; the worst run-time is less than one second. It can also be observed that DFT was more efficient than DWT, and that KCASA U-KTS required less resource than KCASA M-KTS.

Table 7.1: Typing template generation complexity (seconds) for KCASA applied to U-KTS (KCASA_U-KTS+DFT and KCASA_U-KTS+DWT).

ω	KCASA_U-KTS+DFT			KCASA_U-KTS+DWT		
	ACB	GP	VHHS	ACB	GP	VHHS
16	0.001	0.015	0.001	0.005	0.029	0.005
32	0.002	0.025	0.001	0.010	0.056	0.010
64	0.003	0.063	0.002	0.015	0.078	0.016
128	0.006	0.081	0.004	0.019	0.114	0.024
256	0.009	0.110	0.006	0.023	0.125	0.029
512	0.013	0.120	0.009	0.027	0.140	0.037

Table 7.2: Typing template generation complexity (seconds) for KCASA applied to M-KTS (KCASA_M-KTS+DFT and KCASA_M-KTS+DWT).

ω	KCASA_M-KTS+DFT		KCASA_M-KTS+DWT	
	ACB	VHHS	ACB	VHHS
16	0.013	0.012	0.021	0.022
32	0.022	0.023	0.042	0.043
64	0.051	0.035	0.071	0.052
128	0.076	0.065	0.098	0.071
256	0.095	0.089	0.122	0.094
512	0.102	0.099	0.132	0.105

7.4.3 Authentication Performance

For evaluating the authentication performance, the typing process was simulated by streaming the data for each subject in each data set. The idea was therefore to pretend one subject being alleged as another subject through typing session. For each comparison of a collected (and transformed) time series subsequence with the previously collected subsequence, the outcome would be either: (i) True Positive-TP, (ii) False Positive-FP, (iii) False Negative-FN or (iv) True Negative-TN. Consequently, the accuracy, FMR and FNMR could be calculated using the equations given previously in Sub-section 6.4.3 (Chapter 6). For convenience, these equations are restated here:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{7.4}$$

$$FMR = \frac{FP}{FP + TN} \tag{7.5}$$

$$FNMR = \frac{FN}{FN + TP} \tag{7.6}$$

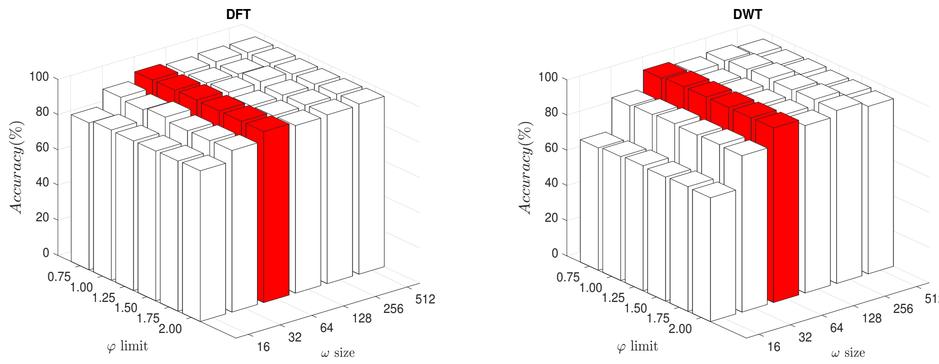
Recall that the authentication performance was evaluated using ranges of values for the ω and φ parameters. Recall also that the evaluation of the proposed KCASA approach was conducted by applying it to both U-KTS and M-KTS. Thus the obtained results are considered in terms of U-KTS and M-KTS in the following two sub-sections starting with U-KTS. Some further comparison is presented in Sub-section 7.4.3.3.

7.4.3.1 Results in Context of U-KTS

Figures 7.9, 7.10 and 7.11 present the performance results obtained when KCASA was applied to U-KTS, in the form of 3D bar charts, with respect to the three data sets considered. Each figure includes two such charts, with DFT on the left and DWT on the right. The x- and y-axis are ω and φ, whilst the vertical axis (the z-axis) shows overall accuracy. Inspection of the figures shows that $\omega = 64$ (highlighted in red) produced the best accuracy results for all three data sets with respect to both DFT and DWT. It can also be observed that when the value for ω increases beyond 64 the effect on accuracy is marginal.

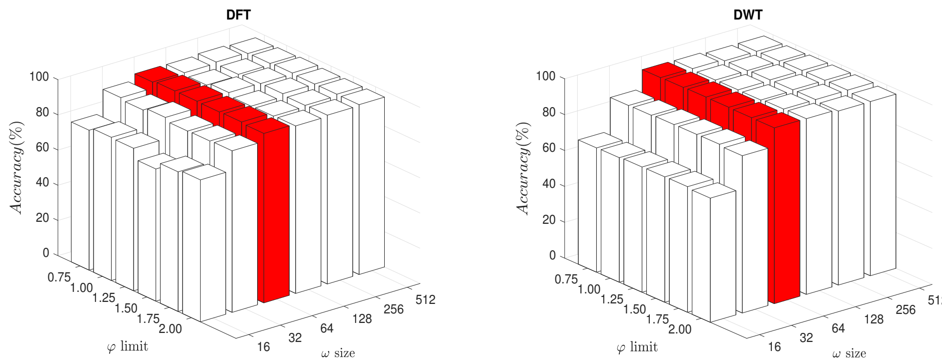
With respect to the φ parameter, it can be seen that this has less effect on the overall authentication performance, although $\varphi = 1.25$ (seconds) tended to produce better results. In other words, the selection of φ does not seem to have as much impact as the selection of the ω parameter.

The accuracy (Acc.), FMR and FNMR results obtained, in the context of U-KTS when using $\omega = 64$ and $\varphi = 1.25$, are presented in tabular form in Table 7.3. The table includes overall average values (Avg.), and the associated Standard Deviation (SD), across all data sets. From the table, it can be observed that the DWT representation produced the best overall accuracy (average accuracy of 98.24% with an associated SD of 1.08). Moreover, it can also be noted that the DWT representation gave the best FMR and FNMR results with an average of 0.032 and 1.500 respectively; with associated SD values of 0.008 and 0.141.



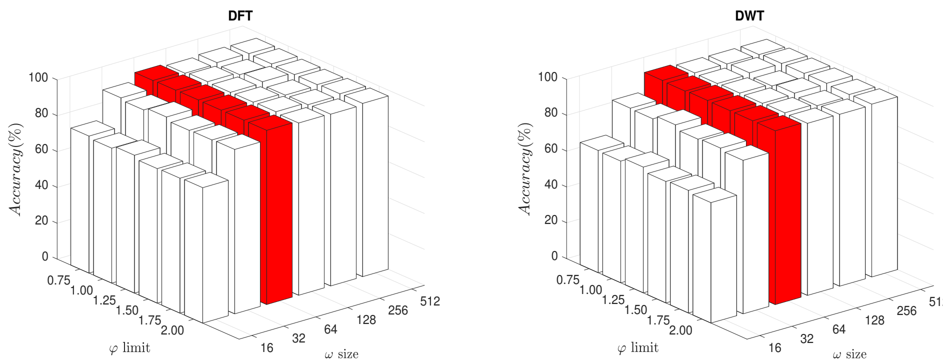
(a) (b)

FIGURE 7.9: The effect of ω and φ parameter settings on accuracy using **U-KTS** and the **ACB** data set.



(a) (b)

FIGURE 7.10: The effect of ω and φ parameter settings on accuracy using **U-KTS** and the **GP** data set.



(a) (b)

FIGURE 7.11: The effect of ω and φ parameter settings on accuracy using **U-KTS** and the **VHHS** data set.

TABLE 7.3: Reported performance results of KCASA, in the context of U-KTS, using $\omega = 64$ and $\varphi = 1.25$.

Data set	DFT			DWT		
	Acc.	FMR	FNMR	Acc.	FMR	FNMR
ACB	97.43	0.030	1.500	99.22	0.029	1.370
GP	96.94	0.034	1.720	98.41	0.032	1.480
VHHS	97.42	0.045	1.850	97.09	0.036	1.650
Avg.	97.26	0.036	1.690	98.24	0.032	1.500
SD	0.28	0.008	0.177	1.08	0.004	0.141

7.4.3.2 Results in Context of M-KTS

The reported results of the KCASA approach, when applied to M-KTS, are presented in this sub-section. Recall that the GP data set was not used here because it was not possible to construct multivariate data for this data set. Moreover, it should be mentioned that the evaluation was conducted using the same selections of ω and φ values as in the case of the U-KTS evaluation given above.

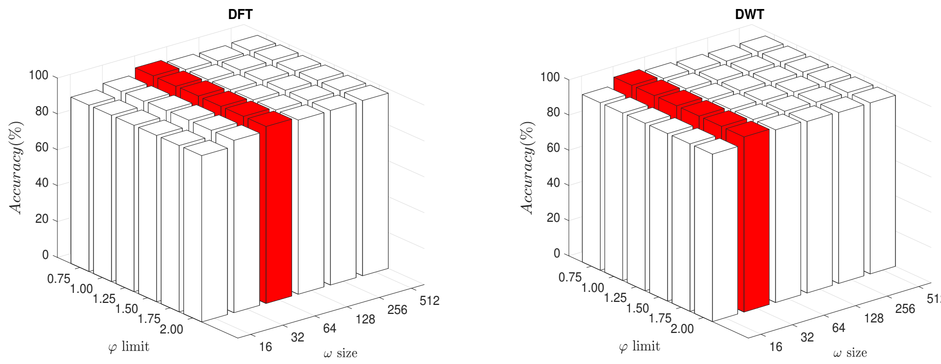
As before, the obtained accuracy results are given in the form of 3D bar charts in Figures 7.12 and 7.13 for the ACB and VHHS data sets respectively. Again, best results are highlighted in red. From the figures, it can be observed that in the context of DFT a best accuracy was obtained when using $\omega = 64$ whilst in the context of DWT a best accuracy was obtained when using $\omega = 32$ (with respect to both data sets). This means that good accuracy authentication can be gained using short time series subsequences; in other words, an accurate authentication can be obtained using only a small portion of the keystroke data stream; an important advantage for "real-time" authentication as might be desirable in the context of the online assessments frequently used concerning eLearning and MOOCs.

With respect to the φ parameter, the best recorded performance was obtained using $\varphi = 1.25$ seconds, although, as in the case of the U-KTS experiments, it can be noted that the φ setting had less effect on authentication performance than the ω setting.

The accuracy (Acc.), FMR and FNMR results obtained, in the context of M-KTS when using $\omega = 32$ and $\varphi = 1.25$, are presented in tabular form in Table 7.3. The table also gives the overall average values and the associated SD in each case. The table clearly shows that the DWT produced the best performance, with an average accuracy of 99.12% (and an associated SD of 0.77). For FMR and FNMR, the best obtained results were 0.010 and 0.816, again using DWT.

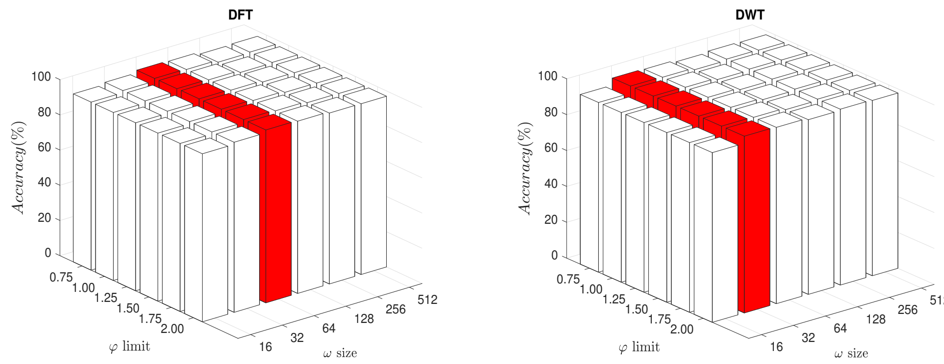
7.4.3.3 Comparison of KCASA U-KTS versus KCASA M-KTS

Tables 7.5, 7.6 and 7.7 present a summary of the results presented in the previous two sub-sections so that some comparisons can be made between the operation of KCASA with U-KTS and KCASA with M-KTS. The results presented in the tables are those obtained using the most appropriate parameter settings for ω and φ, namely $\omega = 64$ and



(a) (b)

FIGURE 7.12: The effect of ω and φ parameter settings on accuracy using **M-KTS** and the **ACB** data set.



(a) (b)

FIGURE 7.13: The effect of ω and φ parameter settings on accuracy using **M-KTS** and the **VHHS** data set.

TABLE 7.4: Reported performance results of KCASA, in the context of M-KTS, using $\omega = 32$ and $\varphi = 1.25$.

Data set	**DFT**			**DWT**		
	Acc.	FMR	FNMR	Acc.	FMR	FNMR
ACB	98.78	0.016	0.868	99.67	0.009	0.700
VHHS	98.30	0.018	0.941	98.58	0.011	0.932
Avg.	98.54	0.017	0.904	99.12	0.010	0.816
SD	0.34	0.002	0.051	0.77	0.001	0.165

$\varphi = 1.25$ for KCASA U-KTS, and $\omega = 32$ and $\varphi = 1.25$ for KCASA M-KTS. Overall, the tables show that DWT, in the context of M-KTS, outperformed the other considered variations of the KCASA approach.

Table 7.5: Summary of the obtained accuracy (**Acc.**) results for the KCASA approach.

Data set	KCASA_DFT +U-KTS	KCASA_DWT +U-KTS	KCASA_DFT +M-KTS	KCASA_DWT +M-KTS
ACB	97.43	99.22	98.78	**99.67**
GP	96.94	98.41	–	–
VHHS	97.42	97.09	98.30	98.58

Table 7.6: Summary of the obtained **FMR** results for the KCASA approach.

Data set	KCASA_DFT +U-KTS	KCASA_DWT +U-KTS	KCASA_DFT +M-KTS	KCASA_DWT +M-KTS
ACB	0.030	0.029	0.016	**0.009**
GP	0.034	0.032	–	–
VHHS	0.045	0.036	0.018	0.011

Table 7.7: Summary of the obtained **FNMR** results for the KCASA approach.

Data set	KCASA_DFT +U-KTS	KCASA_DWT +U-KTS	KCASA_DFT +M-KTS	KCASA_DWT +M-KTS
ACB	1.500	1.370	0.868	**0.700**
GP	1.720	1.480	–	–
VHHS	1.850	1.650	0.941	0.932

7.5 KCASA versus IKCA

As indicated in the introduction to this chapter, the proposed KCASA approach is an alternative to the IKCA approach presented in Chapter 6. This section presents a comparison of the approaches. Both approaches were directed at keystroke iterative authentication using keystroke time series and both operated, at least at a high level, in a similar manner. However, the distinction between the two is that the KCASA approach transforms the collected time series into the spectral domain; both approaches were applied to both U-KTS and M-KTS keystroke time series. The two approaches, and their variations, can thus be summarised as follows:

1 **IKCA+U-KTS**: Iterative authentication where IKCA was applied to univariate keystroke time series.

2 **IKCA+M-KTS**: Iterative authentication where IKCA was applied to multivariate keystroke time series.

3 **KCASA_DFT+U-KTS**: Iterative authentication where KCASA was coupled with DFT and applied to univariate keystroke time series.

4 **KCASA_DWT+U-KTS**: Iterative authentication where KCASA was coupled with DWT and applied to univariate keystroke time series.

5 **KCASA_DFT+M-KTS**: Iterative authentication where KCASA was coupled with DFT and applied to multivariate keystroke time series.

6 **KCASA_DWT+M-KTS**: Iterative authentication where KCASA was coupled with DWT and applied to multivariate keystroke time series.

Each of the above was evaluated, in terms of its authentication performance, using the three test data sets, ACB, GP and VHHS, introduced earlier in this thesis. Moreover, the Binomial proportion confidence intervals were calculated[1] at $p < 0.01$ so as to determine the statistical significance for each method.

The best results obtained are summarised in Tables 7.8, 7.9 and 7.10 in terms of accuracy, FMR and FNMR. Tables 7.8, considers the ACB data set, Table 7.9 the GP data set and Table 7.10 the VHHS data set. In each case the best performing ω and φ parameters are also listed. From the tables, it can be observed that KCASA_M-KTS+DWT variation produced the best performance statistically out of all the variations considered with respect to all the evaluation metrics. The best accuracy was 99.67% with FMR and FNMR of 0.009 and 0.700 respectively for ACB data set. Best results, in the context of KCASA_M-KTS+DWT , were obtained using $\omega = 32$ and $\varphi = 1.25$.

TABLE 7.8: Results of KCASA and IKCA variations for **ACB** data set.

Method	Acc.	FMR	FNMR	Best Parameters	
				ω	φ
IKCA+U-KTS	96.20	0.580	1.970	100	1.25
IKCA+M-KTS	98.39	0.045	1.093	125	1.50
KCASA_DFT+U-KTS	97.43	0.130	1.500	64	1.25
KCASA_DWT+U-KTS	99.22	0.029	1.070	64	1.25
KCASA_DFT+M-KTS	98.78	0.036	1.091	64	1.25
KCASA_DWT+M-KTS	**99.67**	**0.009**	**0.700**	32	1.25

TABLE 7.9: Results of KCASA and IKCA variations for **GP** data set.

Method	Acc.	FMR	FNMR	Best Parameters	
				ω	φ
IKCA+U-KTS	96.31	0.730	1.895	100	1.25
KCASA_DFT+U-KTS	96.94	0.034	1.720	64	1.25
KCASA_DWT+U-KTS	**98.41**	**0.032**	**1.480**	64	1.25

TABLE 7.10: Results of KCASA and IKCA variations for **VHHS** data set.

Method	Acc.	FMR	FNMR	Best Parameters	
				ω	φ
IKCA+U-KTS	94.83	1.045	2.020	100	1.25
IKCA+M-KTS	97.32	0.057	1.095	125	1.50
KCASA_DFT+U-KTS	97.42	0.045	1.085	64	1.25
KCASA_DWT+U-KTS	97.09	0.059	1.098	64	1.25
KCASA_DFT+M-KTS	98.30	0.018	0.941	64	1.25
KCASA_DWT+M-KTS	**98.58**	**0.011**	**0.932**	32	1.25

[1]A well-known statistical significant test, see [160, 163] for further details.

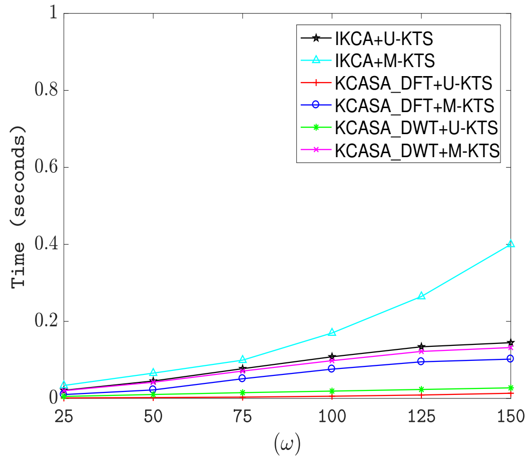


FIGURE 7.14: Template construction runtime (seconds) comparison using variations of IKCA and KCASA with respect to **ACB** data set.

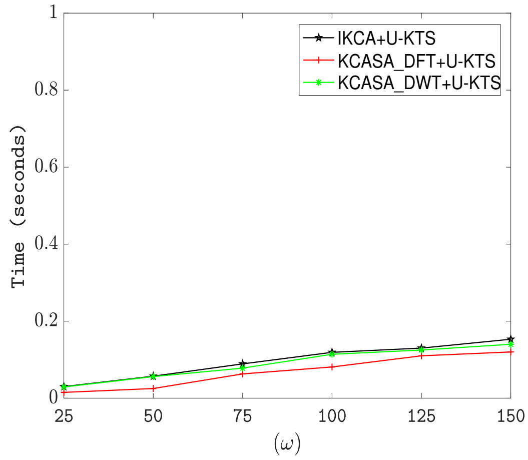


FIGURE 7.15: Template construction runtime (seconds) comparison using variations of IKCA and KCASA with respect to **GP** data set.

Evaluation was also conducted to measure the efficiency of constructing the required typing templates (the biometric database) and the bespoke threshold generation for each subject. Figures 7.14, 7.15 and 7.16 summaries the run-time (seconds) results obtained with respect to the ACB, GP and VHHS data sets receptively. From the figures, it can be seen that, regardless of which approach was used and which variation, in all cases, the runtime increased as ω increased. As noted earlier, this was to be anticipated because the computation time required for the DTW would increase as the ω value increased. Overall the template construction efficiency results indicate that when using the proposed KCASA approach, efficiency gains were made over the IKCA approach, with DFT producing better runtime results than DWT.

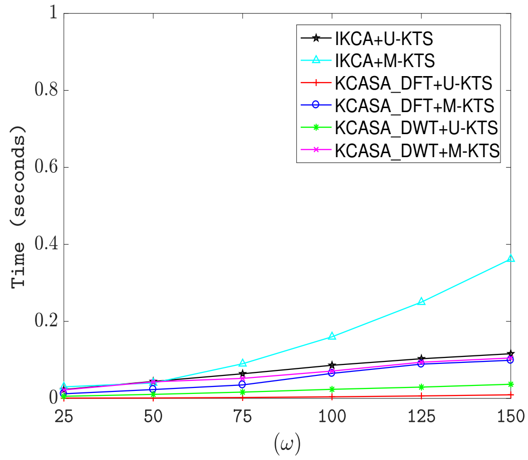


FIGURE 7.16: Template construction runtime (seconds) comparison using variations of IKCA and KCASA with respect to **VHHS** data set.

7.6 Summary

This chapter has introduced a novel approach for iterative continuous keystroke authentication, namely the Keystroke Continuous Authentication based Spectral Analysis (KCASA) approach. The fundamental idea of KCASA was to sample keystroke time series subsequences, of length ω, and then transform each collected time series subsequence from the temporal domain to the spectral domain. The intuition was that such time series transformations would provide for efficiency gains and improved performance. Two spectral transformations were considered: (i) Discrete Fourier Transform (DFT) and (ii) Discrete Wavelet Transform (DWT). The comparison between transformed keystroke signals, as in the case of the IKCA approach, was conducted using Dynamic Time Warping (DTW) due to the advantages that DTW offered with respect to capturing time shifting (offsets) between corresponding sub-sequences.

The proposed KCASA approach was evaluated with respect to both univariate and multivariate keystroke time series (U-KTS and M-KTS). The evaluation also considered the effect of different parameter settings for the window size (ω) and the noise reduction limit (φ). The experimental results indicated that KCASA coupled with DWT significantly outperformed KCASA coupled with DFT in terms of authentication performance. The best result was obtained using $\omega = 32$ keystrokes, and $\varphi = 1.5$ seconds, when KCASA was applied to M-KTS. However, KCASA coupled with DFT was found to be the most efficient. It was observed that the KCASA approach produced significantly superior performance, in terms of authentication and efficiency, than the IKCA analogue approach proposed in the previous chapter (Chapter 6) and by extension the feature vector based approach form the literature.

In the following chapter, the thesis is concluded with a summary and an overview of the main findings in terms of the original research question and subsidiary research questions postulated in Chapter 1. The chapter also considers some potential directions for future work whereby the work presented in this thesis can be extended.

Chapter 8

Conclusion and Future Work

8.1 Introduction

This chapter concludes the work presented in this thesis. The chapter commences, Section 8.2, with a summary of the material presented. The chapter then goes on in Section 8.3 to present the main findings and contributions of the work in the context of the postulated research question, and subsidiary research questions, presented in Chapter 1. The chapter is concluded, in Section 8.4, with a review of potential areas for future research that build upon the work presented in the thesis.

8.2 Summary of Thesis

The work presented in this thesis commenced with a scene-setting chapter, Chapter 1, where a "roadmap" for the thesis was presented, including the motivations, the research question and subsidiary research questions, the main contributions of the research and the adopted research methodology. The central theme proposed in the thesis is iterative user keyboard authentication using techniques taken from the domain of time series analysis.

The main motivation for keyboard user authentication, or simply keystroke authentication, was the substantial increase of internet-facilitated remote learning domains, such as eLearning platforms and Massive Open Online Courses (MOOCs). This, in turn, has led to the requirement for techniques whereby the claimed identity of distance learners can be reliably authenticated. A commonly utilised authentication method, in this context, is through the use of user credentials (usernames and passwords). This type of methods is typically known as "once only" authentication. However, in the case of online assessments and exams there is a requirement to monitor the identity of users throughout the course of an entire assessment; in other words, "iterative" continuous authentication is required. The aim was thus to detect impersonations in the context of remote learning assessments.

From the literature, the "standard" solution to provide iterative continuous authentication was to use some form of biometrics such as continuous iris recognition or fingerprint recognition, but this requires specialist equipment and technology not readily available to the typical distance learner. Consequently, the selected focus of the thesis was keystroke dynamics (typing patterns), a form of behavioural biometric, because this was seen as a promising cost-effective solution to the iterative continuous authentication problem for online assessment. A promising solution because, again from the literature, it had been shown that individuals have distinctive keyboard usage styles.

The aim of iterative keystroke continuous authentication is to recognise typing patterns as typing progresses and regardless of what text is actually being typed. In other words, typing patterns need to be recognised from free text so that keystroke continuous (iterative) authentication can be realised. This is a non-trivial task. From the literature, the prominent methods used to recognise typing patterns in the context of continuous authentication was by constructing feature vectors based on keystroke timing information. Feature vectors, in this context, usually comprise statistical information concerning keystroke dynamics, such as the mean and standard deviation of the keystroke dynamics associated with sets of n-graphs. However, the main criticism of the feature vector approach was that the feature vector values are either typing pattern abstractions (for example average hold times) or only consider a subset of the data (for example only frequently occurring digraphs). There are also more specific limitations of the feature vector based approach depending on how the feature vectors are constructed. For example, where the feature vectors are constructed using the average flight time of n-graphs contained in a training data set, it can be argued that these might not be representative of the n-graphs in the samples to be authenticated. Consequently, this thesis presented a novel approach to keystroke continuous authentication using time series analysis whereby all keystroke features are taken into consideration, rather than those associated with specific n-graphs. More specifically three approaches were presented:

1. A benchmark, proof-of-concept, system, the Once-only Keystroke Continuous Authentication (OKCA) system, directed at static, as opposed to continuous, user authentication.

2. The Iterative Keystroke Continuous Authentication (IKCA) system, an improved variation of the benchmark system directed at continuous authentication.

3. The Keystroke Continuous Authentication-based Spectral Analysis (KCASA) system, an extension of the previous system where time series is transformed into the spectral domain as opposed to being considered in the temporal domain.

In Chapter 2, the background to the work presented in the thesis was presented. The chapter commenced with a review of keyboard authentication fundamentals; in particular the distinction between *static* authentication and *continuous* authentication as well as the distinction between *user authentication* and *user identification*. This was followed by a review of existing authentication methods. These methods were classified as being either: (i)Token-based or (ii) Biometric-based; further detail was presented with respect to each category. The chapter also considered related work concerning keystroke continuous authentication approaches; especially the state-of-the-art of keystroke dynamics in the context of free text analysis. Time series techniques, pertinent to the work presented in this thesis, were also reviewed including similarity comparison methods. The adopted similarity method used with respect to all the systems presented in the thesis was Dynamics Time Warping (DTW) which was therefore considered in detail in the chapter.

Chapter 3 then went to introduce the evaluation data sets that were used for evaluating the proposed keystroke continuous authentication approaches. The chapter commenced with a discussion of the factors to be considered when attempting to collect typing samples. It was noted that, in the context of user authentication directed at those completing online assessments, it was most appropriate to obtain unstructured text within uncontrolled environments as this best simulated the process of online assessment. The presented data sets were as follows: (i) The University of Liverpool ACB data set collated by the author, (ii) The University of Torino GP data set collected by [54] and (iii) The Clarkson University VHHS data set collected by [161]. The mechanism and process used to collect each data set were fully described in the chapter. Moreover, statistical information for each data set was also presented so as to provide the reader with a better understanding of the characteristics of each data set.

Chapter 4 presented a formalism for the keyboard continuous authentication problem in terms of time series. Intuitively, a keystroke time series is a sequence of discrete real-valued data points representing either: (i) the flight time $\mathcal{F}^t$ or (ii) the key-hold time $\mathcal{KH}^t$ or (iii) both $\mathcal{F}^t$ and $\mathcal{KH}^t$. Thus each keystroke point was conceptualised to be a dimensional discrete (indexed) temporal event (ρ_i) forming a sequence of multidimensional events $\{\rho_1, \rho_2, \dots\}$ where each ρ holds timing information concerning the $\mathcal{F}^t$ and/or $\mathcal{KH}^t$ timing features. Consequently, Univariate-Keystroke Time series (U-KTS) and Multivariate-Keystroke Time Series (M-KTS) could be constructed.

Chapter 5 introduced the first time series-based keystroke authentication system considered in the thesis, namely the Once-only Keystroke Continuous Authentication (OKCA) system. The principle idea of OKCA was to provide a proof of concept system to evaluate the effectiveness of the idea of using time series analysis techniques for the purpose of keyboard user authentication. For this purpose, the OKCA system was directed at static user identification. Three variations of the OKCA system were constructed according to the nature of the time series to be considered as follows: (i) U-KTS with $\mathcal{F}^t$, (ii) U-KTS with $\mathcal{KH}^t$ and (ii) M-KTS. The evaluation of the OKCA system was conducted using the three evaluation data sets presented earlier in the thesis in Chapter 3. A comparison was performed with the feature vector representation method as proposed in the earlier work on user authentication/identification directed at keystroke dynamics from free text. The reported evaluation indicated that the use of time series showed promise with respect to typing patterns extraction from free text which could then be fruitfully employed for iterative (real-time) keystroke continuous authentication.

In Chapter 6, the Iterative Keystroke Continuous Authentication (IKCA) approach was introduced. The IKCA approach was founded on the OKCA proof-of-concept approach from the previous chapter, but directed at continuous keyboard user authentication. More specifically, the IKCA approach was designed to operate in an iterative manner, as typing proceeded, by repeatedly extracting keystroke time series subsequences, of length ω, called shapelets. The extracted shapelets were either in the form of U-KTS or M-KTS. The idea was that the first extracted shapelet would be compared with a previously constructed biometric enrolment database to verify the identity of the claimed subject; this process was defined as a *start-up* authentication. As typing proceeded, each further extracted shapelet was compared with the previously extracted shapelet. In this manner, continuous authentication could be realised. A novel feature of the IKCA approach was that a bespoke similarity threshold σ was computed for each subject in the data set. The evaluation of the IKCA approach demonstrated that IKCA coupled with M-KTS produced a better performance compared with IKCA coupled with U-KTS and a comparator feature vector-based approach. Moreover, it was concluded that the performance of the IKCA approach was better than the established feature vector based approaches for keystroke continuous authentication in terms of both accuracy and efficiency. This confirmed the hypothesis central to the thesis that time series analytics is an effective iterative keystroke authentication mechanism.

In Chapter 7, the Keystroke Continuous Authentication based Spectral Analysis (KCASA) approach was presented. The central idea underpinning the KCASA approach was to apply a spectral transform to the collected keystroke time series prior to any authentication taking place. The conjecture was that transformation into the spectral domain would lead to more accurate and faster authentication. In the KCASA approach, two spectral transformations were considered: (i) Discrete Fourier Transform (DFT) and (ii) Discrete Wavelet Transform (DWT). The approach was evaluated by applying both U-KTS and M-KTS. The reported results showed that DWT, coupled with M-KTS, outperformed DFT, in all cases, with respect to authentication performance. However, DFT was found to be more efficient. Moreover, it was observed that the KCASA approach obtained a significantly superior performance, in terms of both authentication and efficiency, comparing to the IKCA approach from the previous chapter (Chapter 6).

8.3 Main Findings and Contributions

The main findings and contributions from the work presented in this thesis are given in this section. The principle motivation for the work was formulated in a single research question as follows:

"Is it possible to continuously authenticate individuals, according to their keyboard usage patterns; and if so what are the most appropriate mechanisms for achieving this?"

The provision of an answer to this research question entailed the resolution of a number of subsidiary questions. Thus, the main findings for the work presented in this thesis are presented here in the light of each of the subsidiary questions, and then in terms of the overriding initial research question, as follows:

1. *How can we best represent keyboard usage patterns in a way that avoids the disadvantages associated with the feature vector representation used to date?*

   Keyboard usage patterns can be extracted, as noted in Chapter 2, from either fixed text or free text. The "standard" approach is to use a feature vector representation. However, in the context of free text, there is a substantial overhead with respect to the generation of the feature vectors; there is also a question about how effective they are. The proposed solution was to consider keystroke dynamics in terms of time series data, in which typing patterns can be recognised from free text and used for continuous authentication. **Thus, in the proposed representation, each keystroke was represented as a temporal event (a press-and-release event) where two principal timing features, flight time $\mathcal{F}^t$ and key-hold time $\mathcal{KH}^t$, were considered as these encapsulate other features that might also be considered**. The intuition of using keystroke time series was that the sequence of temporal typing events produced distinctive shapelets for each user; thus these shapelets would provide for an authentication mechanism. Two categories of keystroke time series were considered univariate and multivariate (U-KTS and M-KTS). Time series in both the temporal domain and transformed to the spectral domain were experimented with. With respect to the spectral domain, two transforms were considered: Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT). The most appropriate mechanism for representing keyboard usage patterns was found to be the DWT representation when applied to M-KTS in terms of authentication accuracy, and the DFT representation in terms of efficiency.

2. *Given a collection of patterns, what is the most appropriate mechanism whereby a new pattern can be compared with an existing pattern, for the purpose of user authentication?*

   Given a typing stream represented as a keystroke time series, user authentication requires frequent comparison of time series subsequences. The established, and most straightforward, mechanism for comparing two time series is to use ED measurement. This requires the two time series to be compared to be of the same length; however, a more significant issue with respect to keyboard time series comparison is that it does not allow for offsets in the time series (where distinctive subsequences are not aligned). **The adopted method was therefore to use Dynamic Time Warping (DTW) which addresses the disadvantages associated with ED similarity measurement**. The derived similarity measure (the DTW minimum warping path distance) was then compared to a threshold σ. If the distance was less than the threshold then the subject was deemed to be who they claimed to be; otherwise, a warning was issued. An interesting "twist" to the proposed process was that a bespoke σ value was determined for each subject, as a oppose to a "blanket" threshold. This was in recognition of the observation that there is no "one size fits all" σ threshold value. The approach produced some excellent results; a best accuracy of 99.12% was recorded.

3. *Given solutions to 1 and 2 above, how do we go about evaluating whether a good solution has been discovered (or not)?*

   Biometric systems are typically evaluated using Accuracy (Acc.), False Match Rate (FMR) and False Non-Match Rate (FNMR); these metrics were therefore also selected for evaluating the proposed keyboard continuous user authentication approaches. More specifically, a number of data sets were used, one collated by the author and two more from the literature. Each record for each subject was split into two, one for building the user typing template required by the authentication mechanism and one to represent the keystroke data stream. To simulate the presence of an imposter, a test stream from another subject was appended to that of the current subject. Experiments were then conducted to determine whether the imposter was detected or not. Using this approach, the comparison between the two keystroke continuous authentication approaches proposed in the thesis (IKCA and KCASA) were conducted and compared with the established (from the literature) feature vector-based approach. This evaluation mechanism seemed to work well.

4. *What is the most effective process whereby continuous, real-time, authentication can be conducted in the context of online assessment and more generally?*

   The mechanism presented in the thesis to address continuous, real-time, authentication was to periodically sample the keystroke input stream and extract time series subsequences called shapelets. Each shapelet was then compared to the previously collected shapelet. The exception was the first shapelet in a typing session, which was compared with a previously collected set of profiles that was known to belong to the claimed subject. As noted above, an interesting "twist" to the proposed mechanism was that a bespoke σ similarity threshold was determined for each subject. Some excellent results were thus produced.

5. *In the context of 4 above, how should the process deal with "away from keyboard" events?*

The issue of “away from keyboard” events was significant as their occurrence distorted the nature of the collected time series in the context of flight time $\mathcal{F}^t$. This was established by a process of inspection of keystroke time series that included $\mathcal{F}^t$ values. **The proposed solution was to introduce a limit, φ, for acceptable values of $\mathcal{F}^t$** . More specifically, in the event of a $\mathcal{F}^t$ value that exceeded the φ value the idea was to reduce the “outlier” value to the φ value. This was conceived of as a form of noise reduction. A number of potential values for φ were evaluated ranging from 0.750 seconds to 2.00 seconds increasing in steps of 0.25 seconds. The reported evaluation demonstrated that the limit of $\varphi = 1.25$ seconds produced good authentication results with respect to the proposed IKCA and KCASA keystroke continuous authentication approaches.

6. *Can we recognise typing patterns in a way that avoids the knowledge of the text that is being typed by the user so as to avoid data privacy concerns?*

   For reasons of data confidentiality, it was seen as important that what was actually being typed was not used in the context of subject authentication. This was not the case with respect to the n-graphs approaches reported on in the literature. **Therefore, the proposed approaches, OKCA, IKCA and KCASA, made no use of the content of the text being typed using time series context**. The proposed approaches could therefore be applied by a third party without contravening any data protection requirement as the system has no cognisance of what was being typed. This demonstrated the advantage that the proposed keystroke time series authentication approaches presented in this thesis can avoid data privacy concerns.

Returning to the central research question of the thesis, it can be stated that it is possible to continuously authenticate individuals according to their keyboard usage patterns. More specifically, it is possible to successfully authenticate individuals in an iterative manner according to their keyboard usage patterns. The most appropriate mechanisms for achieving this is by considering keystroke dynamics from free text as a time series and to use then the time series represented keystroke data to find typing patterns that are indicative of individuals which can, in turn, be used for authentication purposes. The evaluation conducted, and reported on in this thesis, indicated that good results can consistently be achieved.

For completeness, the main contributions for the work presented in this thesis are restated from Chapter 1 as follows:

1. Enhancement of the accuracy of current pattern detection mechanisms for keystroke continuous authentication; reported keyboard dynamics continuous authentication accuracy to date has been generally poor [151].

2. Confirmation that keystroke dynamics can be effectively encapsulated in the form of a discrete time series representation.

3. A system for real-time/continuous keystroke authentication that is independent of the keyboard layout, and therefore generic, as required with respect to eLearning and MOOC systems; some of reported work to date has only been directed at controlled environments using prescribed keyboards,

4. A keystroke modeling technique that is a privacy-preserving in that knowledge of which keys are being pressed is not required.

5. A keyboard authentication mechanism that, although intended for use with respect to online assessment, has general applicability. For example, it may equally well be used to detect certain human conditions, such as detecting keyboard user emotions as described in [132].

6. A keystroke online data capture tool, namely the Web-Based Timestamp Keystroke Recorder (WBTKR), which facilitates the collection of keystroke data sets.

7. A keystroke dynamics data set, collected from real users in an uncontrolled environment, available for public use.

8. Eight competing techniques directed at keystroke dynamics authentication as follows:

   i Static authentication using univariate keystroke time series representation with flight time (the benchmark algorithm).

   ii Static authentication using multivariate keystroke time series representation with flight time and key-hold time (the benchmark algorithm).

   iii Continuous authentication using univariate keystroke time series representation with flight time.

   iv Continuous authentication using multivariate keystroke time series representation with flight time and hold time.

   v Continuous authentication using DFT spectral (transformed) univariate keystroke time series with flight time.

   vi Continuous authentication using DFT spectral (transformed) multivariate keystroke time series with flight time and hold time.

   vii Continuous authentication using DWT spectral (transformed) univariate keystroke time series with flight time.

   viii Continuous authentication using DWT spectral (transformed) multivariate keystroke time series with flight time and hold time.

8.4 Future Work

The work presented in this thesis has proposed an approach to keystroke continuous authentication directed at detecting impersonation with respect to applications such as online assessments. In this section, a number of potential directions for future work are introduced. Generally speaking, the future work can be divided into **immediate** and **longer** contexts. The immediate context is concerned with aspects where the current work can be extended/improved. The longer context is concerned with directions whereby the proposed approaches can be used to explore more speculative areas. Therefore, the future work discussed in this subsection is presented with respect to this categorisation as follows:

- **Future Work in The Immediate Context**

  1. **DTW Complexity**: DTW was used to compare time series subsequences (shapelets) because of the advantage that it offers with respect capturing the similarity between time series where offsets exist. This is because of the ability of DTW to "warp" the linearity of the time series. However, this advantage is gained at the expensive of computation time (Euclidean similarity calculation is much faster although it does not take account of offsets). In the literature, there are various solutions to mitigate against the complexity of DTW, for example, the work presented in [62] and [138]. In this thesis, no such mitigation was applied with respect to the experiments reported on, although this could clearly be done. It is thought that adopting such DTW mitigation techniques, to the approaches presented in the thesis, can provide for additional efficiency gains, especially where real-time authentication is desired. Thus, the time complexity of DTW, in the context of the proposed keystroke time series representations, remains an open topic for future work.
  2. **Overlapped Windows**: In the proposed approaches for continuous authentication, the continuous authentication processes operated by extracting time series subsequences (shapelets) using a window of length ω. The frequency with which shapelets were collected can be defined according to a variable f. Thus when $f > \omega$ we talk of non-overlapping time series, when $f = \omega$ we talk about abutting time series and when $f < \omega$ we talk of overlapping time series. However, in the thesis, the use of such a variable and consequently the use of overlapped time series ($f < \omega$) was not considered. It might be the case that this produces a better performance, and thus this is identified as another area suitable for further investigation.
  3. **Alternative Representation**: The work described in this thesis used a time series representation to represent keystroke dynamics; this was indeed one of the main motivations for this work because there was no other work in the literature that considered time series representation for keystroke continuous authentication. Although good experimental results were reported using the proposed time series representations, some alternative time series representations might produce further improvements. For example, with respect to the considered keystroke DWT representation (included in the KCASA

approach), there is a drawback, when using the Haar transform, that the keystroke time series must have a length which is an integral power of two. An alternative, not considered in the thesis, is Piecewise Aggregate Approximation (PAA) [78] which operates using any time series length using an approximation of the DWT representation [77]. Consequently, the use of alternative time series representations, with respect to, keystroke time series authentication is another fruitful topic for further research.

4. **Controlled Environment**: The focus for the work presented in this thesis was continuous authentication using typing samples collected within an uncontrolled environment, such as when individuals are typing free text using whatever keyboard they have at hand. This was seen as desirable concerning applications such as user authentication in the context of online assessments. However, the use of typing samples collected in controlled conditions, such as transcribed (fixed) text, is an alternative way whereby the keystroke dynamics approach can operate; this is especially the case where *static* authentication is required. Therefore, the question remains as to how the proposed keystroke representation could perform under such conditions. The importance of providing an answer to this question is that there are many applications where static authentication is a requirement. Thus an investigation of using the proposed keystroke representation for alternative (static) applications would be a useful direction for future work.

- **Future Work in The Longer Context**

1. **Alternative Domains**: It was noticed in Section 1.5, Chapter 1, that one contribution of the work presented in this thesis was to introduce a general framework in which keystroke dynamics can be applied in the context of different domains rather than the authentication domain. For example, the proposed keystroke time series representation could be utilised to detect certain human conditions such as Parkinson's disease (see for example [51]). A further alternative domain where the proposed work presented in this thesis can equally well be used, is in the context of user emotions detection (see for instance [132] where keystroke dynamics were used for user emotion detection). Investigation of these alternative domains thus provides for a further opportunity to extend the investigation presented in this thesis.

References

[1] Norman Adams, Daniela Marquez, and Gregory Wakefield. Iterative deepening for melody alignment and retrieval. 2005.

[2] Rakesh Agrawal, Christos Faloutsos, and Arun Swami. Efficient similarity search in sequence databases. In *International Conference on Foundations of Data Organization and Algorithms*, pages 69–84. Springer, 1993.

[3] Ahmed A Ahmed and Issa Traore. Biometric recognition based on free-text keystroke dynamics. *Cybernetics, IEEE Transactions on*, 44(4):458–472, 2014.

[4] Ahmed Awad E Ahmed and Issa Traore. A new biometric technology based on mouse dynamics. *Dependable and Secure Computing, IEEE Transactions on*, 4(3):165–179, 2007.

[5] M Akila, V Suresh Kumar, N Anusheela, and K Sugumar. A novel feature subset selection algorithm using artificial bee colony in keystroke dynamics. In *Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011) December 20-22, 2011*, pages 813–820. Springer, 2012.

[6] Jeffrey D Allen. *An analysis of pressure-based keystroke dynamics algorithms.* PhD thesis, Southern Methodist University, 2010.

[7] Jonathan Alon, Vassilis Athitsos, Quan Yuan, and Stan Sclaroff. A unified framework for gesture recognition and spatiotemporal gesture segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 31(9):1685–1699, 2009.

[8] Walid G Aref, Mohamed G Elfeky, and Ahmed K Elmagarmid. Incremental, online, and merge mining of partial periodic patterns in time-series databases. *IEEE Transactions on Knowledge and Data Engineering*, 16(3):332–342, 2004.

[9] S Asha and C Chellappan. Authentication of e-learners using multimodal biometric technology. In *Biometrics and Security Technologies, 2008. ISBAST 2008. International Symposium on*, pages 1–6. IEEE, 2008.

[10] Karl Johan Åström. On the choice of sampling rates in parametric identification of time series. *Information Sciences*, 1(3):273–278, 1969.

[11] Anthony J Bagnall and Gareth J Janacek. Clustering time series from arma models with clipped data. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 49–58. ACM, 2004.

[12] Claus Bahlmann and Hans Burkhardt. The writer independent online handwriting recognition system frog on hand and cluster generative statistical dynamic time warping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3):299–310, 2004.

[13] Ned Bakelman and Charles Tappert. Keystroke biometrics studies on short password and numeric passcode input, and on long spreadsheet, browser, and text application input. *ProQuest Dissertations and Theses*, page 178, 2014.

[14] Kiran S Balagani, Vir V Phoha, Asok Ray, and Shashi Phoha. On the discriminability of keystroke feature vectors used in fixed text keystroke authentication. *Pattern Recognition Letters*, 32(7):1070–1080, 2011.

[15] Salil P Banerjee and Damon L Woodard. Biometric authentication and identification using keystroke dynamics: A survey. *Journal of Pattern Recognition Research*, 7(1):116–139, 2012.

[16] Nick Bartlow and Bojan Cukic. Keystroke dynamics-based credential hardening systems. *Handbook of Remote Biometrics for Surveillance and Security*, pages 329–347, 2009.

[17] Norbert Beckmann, Hans-Peter Kriegel, Ralf Schneider, and Bernhard Seeger. The r*-tree: an efficient and robust access method for points and rectangles. In *Acm Sigmod Record*, volume 19, pages 322–331. ACM, 1990.

[18] Francesco Bergadano, Daniele Gunetti, and Claudia Picardi. User authentication through keystroke dynamics. *ACM Transactions on Information and System Security (TISSEC)*, 5(4):367–397, 2002.

[19] Francesco Bergadano, Daniele Gunetti, and Claudia Picardi. Identity verification through dynamic keystroke analysis. *Intelligent Data Analysis*, 7(5):469–496, 2003.

[20] Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, 1994.

[21] Saleh Bleha, Charles Slivinsky, and Bassam Hussien. Computer-access security systems using keystroke dynamics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(12):1217–1222, 1990.

[22] Saleh Ali Bleha and Mohammad S Obaidat. Dimensionality reduction and feature extraction applications in identifying computer users. *Systems, Man and Cybernetics, IEEE Transactions on*, 21(2):452–456, 1991.

[23] Lori Breslow, David E Pritchard, Jennifer DeBoer, Glenda S Stump, Andrew D Ho, and Daniel T Seaton. Studying learning in the worldwide classroom: Research into edx's first mooc. *Research & Practice in Assessment*, 8, 2013.

[24] C Sidney Burrus, Ramesh A Gopinath, and Haitao Guo. Introduction to wavelets and wavelet transforms: a primer. *Prentice-Hall, Inc.*, 1997.

[25] Neisha A Chadwick, David A McMeekin, and Tele Tan. Classifying eye and head movement artifacts in eeg signals. In *Digital Ecosystems and Technologies Conference (DEST), 2011 Proceedings of the 5th IEEE International Conference on*, pages 285–291. IEEE, 2011.

[26] Chi-Kwong Chan and Lee-Ming Cheng. Cryptanalysis of a remote user authentication scheme using smart cards. *IEEE Transactions on Consumer Electronics*, 46(4):992–993, 2000.

[27] FK-P Chan, AW-C Fu, and Clement Yu. Haar wavelets for efficient similarity search of time-series: with and without time warping. *IEEE Transactions on knowledge and data engineering*, 15(3):686–705, 2003.

[28] Kin-Pong Chan and Ada Wai-Chee Fu. Efficient time series matching by wavelets. In *Data Engineering, 1999. Proceedings., 15th International Conference on*, pages 126–133. IEEE, 1999.

[29] Dar-Jen Chang, Ahmed H Desoky, Ming Ouyang, and Eric C Rouchka. Compute pairwise manhattan distance and pearson correlation coefficient of data points with gpu. In *Software Engineering, Artificial Intelligences, Networking and Parallel/Distributed Computing, 2009. SNPD'09. 10th ACIS International Conference on*, pages 501–506. IEEE, 2009.

[30] Lei Chen and Raymond Ng. On the marriage of lp-norms and edit distance. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, pages 792–803. VLDB Endowment, 2004.

[31] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

[32] Coursera. Coursera. https://blog.coursera.org/about/. Accessed: 2017-08-01.

[33] Nick Craswell. Mean reciprocal rank. In *Encyclopedia of Database Systems*, pages 1703–1703. Springer, 2009.

[34] Heather Crawford. Keystroke dynamics: Characteristics and opportunities. In *Privacy Security and Trust (PST), 2010 Eighth Annual International Conference on*, pages 205–212. IEEE, 2010.

[35] John Daniel. Making sense of moocs: Musings in a maze of myth, paradox and possibility. *Journal of interactive Media in education*, 2012(3), 2012.

[36] Kresimir Delac and Mislav Grgic. A survey of biometric recognition methods. In *Electronics in Marine, 2004. Proceedings Elmar 2004. 46th International Symposium*, pages 184–193. IEEE, 2004.

[37] Hui Ding, Goce Trajcevski, Peter Scheuermann, Xiaoyue Wang, and Eamonn Keogh. Querying and mining of time series data: experimental comparison of representations and distance measures. *Proceedings of the VLDB Endowment*, 1(2):1542–1552, 2008.

[38] Paul S Dowland and Steven M Furnell. A long-term trial of keystroke profiling using digraph, trigraph and keyword latencies. In *Security and Protection in Information Processing Systems*, pages 275–289. Springer, 2004.

[39] Ted Dunstone and Neil Yager. *Biometric system and data analysis: Design, evaluation, and data mining*. Springer Science & Business Media, 2008.

[40] Simon Eberz, Kasper B Rasmussen, Vincent Lenders, and Ivan Martinovic. Evaluating behavioral biometrics for continuous authentication: Challenges and metrics. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, pages 386–399. ACM, 2017.

[41] Anna Eckerdal, Päivi Kinnunen, Neena Thota, Aletta Nylén, Judy Sheard, and Lauri Malmi. Teaching and learning with moocs: computing academics' perspectives and engagement. In *Proceedings of the 2014 conference on Innovation & technology in computer science education*, pages 9–14. ACM, 2014.

[42] Tim Edwards. Discrete wavelet transforms: Theory and implementation. *Universidad de*, 1991.

[43] edX. Schools and partners. https://www.edx.org/schools-partners. Accessed: 2017-08-01.

[44] Clayton Epp, Michael Lippold, and Regan L Mandryk. Identifying emotional states using keystroke dynamics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 715–724. ACM, 2011.

[45] Christos Faloutsos, Mudumbai Ranganathan, and Yannis Manolopoulos. *Fast subsequence matching in time-series databases*, volume 23. ACM, 1994.

[46] Luis Freire, Alexis Roche, and J-F Mangin. What is the best similarity measure for motion correction in fmri time series? *IEEE transactions on medical imaging*, 21(5):470–484, 2002.

[47] Tak-chung Fu. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24(1):164–181, 2011.

[48] SM Furnell and T Karweni. Security issues in online distance learning. *Vine*, 31(2):28–35, 2001.

[49] Dennis Gabor. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441, 1946.

[50] R Stockton Gaines, William Lisowski, S James Press, and Norman Shapiro. Authentication by keystroke timing: Some preliminary results. Technical report, DTIC Document, 1980.

[51] L Giancardo, T Arroyo-Gallego, I Butterworth, CS Mendoza, P Montero, M Matarazzo, and A Sanchez-Ferro. Computer keyboard interaction as an indicator of early parkinson's disease. *arXiv preprint arXiv:1604.08620*, 2016.

[52] Romain Giot, Mohamad El-Abed, Baptiste Hemery, and Christophe Rosenberger. Unconstrained keystroke dynamics authentication with shared secret. *computers & security*, 30(6):427–445, 2011.

[53] Romain Giot, Mohamad El-Abed, and Christophe Rosenberger. Greyc keystroke: a benchmark for keystroke dynamics biometric systems. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*, pages 1–6. IEEE, 2009.

[54] Daniele Gunetti and Claudia Picardi. Keystroke analysis of free text. *ACM Transactions on Information and System Security (TISSEC)*, 8(3):312–347, 2005.

[55] Aykut Guven and Ibrahim Sogukpinar. Understanding users' keystroke patterns for computer access security. *Computers & Security*, 22(8):695–706, 2003.

[56] Ozlem Guven, Selim Akyokus, Mitat Uysal, and Aykut Guven. Enhanced password authentication through keystroke typing characteristics. In *Proceedings of the 25th conference on Proceedings of the 25th IASTED International Multi-Conference: artificial intelligence and applications*, pages 317–322. ACTA Press, 2007.

[57] Alfred Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331–371, 1910.

[58] Sajjad Haider, Ahmed Abbas, and Abbas K Zaidi. A multi-technique approach for user identification through keystroke dynamics. In *Systems, Man, and Cybernetics, 2000 IEEE International Conference on*, volume 2, pages 1336–1341. IEEE, 2000.

[59] Jiawei Han, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.

[60] Khe Foon Hew and Wing Sum Cheung. Students' and instructors' use of massive open online courses (moocs): Motivations and challenges. *Educational research review*, 12:45–58, 2014.

[61] Reinhold Huber-Mörk, Sebastian Zambanini, Maia Zaharieva, and Martin Kampel. Identification of ancient coins based on fusion of shape and local features. *Machine vision and applications*, 22(6):983–994, 2011.

[62] F. Itakura. Minimum prediction residual principle applied to speech recognition. *IEEETrans. Acoustics, Speech, and Signal Processing*, pages 52–72, 1975.

[63] Plamen Ch Ivanov, Michael G Rosenblum, C-K Peng, Joseph Mietus, Shlomo Havlin, H Eugene Stanley, and Ary L Goldberger. Scaling behaviour of heartbeat intervals obtained by wavelet-based time-series analysis. *Nature*, 383(6598):323–327, 1996.

[64] Anil Jain, Ruud Bolle, and Sharath Pankanti. *Biometrics: personal identification in networked society*, volume 479. Springer Science & Business Media, 2006.

[65] Anil Jain, Lin Hong, and Sharath Pankanti. Biometric identification. *Communications of the ACM*, 43(2):90–98, 2000.

[66] Anil K Jain. Biometric recognition: how do i know who you are? In *International Conference on Image Analysis and Processing*, pages 19–26. Springer, 2005.

[67] Anil K Jain, Lin Hong, and Yatin Kulkarni. A multimodal biometric system using fingerprint, face and speech. In *Proceedings of 2nd Int'l Conference on Audio-and Video-based Biometric Person Authentication, Washington DC*, pages 182–187, 1999.

[68] Anil K Jain, Arun Ross, and Salil Prabhakar. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, 14(1):4–20, 2004.

[69] Gareth J Janacek, Anthony J Bagnall, and Michael Powell. A likelihood ratio distance measure for the similarity between the fourier transform of time series. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 737–743. Springer, 2005.

[70] Rajkumar Janakiraman and Terence Sim. Keystroke dynamics in a general setting. In *Advances in Biometrics*, pages 584–593. Springer, 2007.

[71] Jeffrey L Jenkins, Mark Grimes, Jeffrey Gainer Proudfoot, and Paul Benjamin Lowry. Improving password cybersecurity through inexpensive and minimally invasive means: Detecting and deterring password reuse through keystroke-dynamics monitoring and just-in-time fear appeals. *Information Technology for Development*, 20(2):196–213, 2014.

[72] Rick Joyce and Gopal Gupta. Identity authentication based on keystroke latencies. *Communications of the ACM*, 33(2):168–176, 1990.

[73] Tamer Kahveci, Ambuj Singh, and Aliekber Gurel. Similarity searching for multi-attribute sequences. In *Scientific and Statistical Database Management, 2002. Proceedings. 14th International Conference on*, pages 175–184. IEEE, 2002.

[74] M Karnan, M Akila, and N Krishnaraj. Biometric personal authentication using keystroke dynamics: A review. *Applied Soft Computing*, 11(2):1565–1573, 2011.

[75] Eamonn Keogh. A fast and robust method for pattern matching in time series databases. *Proceedings of WUSS*, 97(1):99, 1997.

[76] Eamonn Keogh. A decade of progress in indexing and mining large time series databases. In *Proceedings of the 32nd international conference on Very large data bases*, pages 1268–1268. VLDB Endowment, 2006.

[77] Eamonn Keogh, Kaushik Chakrabarti, Michael Pazzani, and Sharad Mehrotra. Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and information Systems*, 3(3):263–286, 2001.

[78] Eamonn Keogh, Kaushik Chakrabarti, Michael Pazzani, and Sharad Mehrotra. Locally adaptive dimensionality reduction for indexing large time series databases. *ACM Sigmod Record*, 30(2):151–162, 2001.

[79] Eamonn Keogh and Shruti Kasetty. On the need for time series data mining benchmarks: a survey and empirical demonstration. *Data Mining and knowledge discovery*, 7(4):349–371, 2003.

[80] Eamonn Keogh, Stefano Lonardi, and Bill'Yuan-chi' Chiu. Finding surprising patterns in a time series database in linear time and space. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 550–556. ACM, 2002.

[81] Eamonn Keogh, Li Wei, Xiaopeng Xi, Michail Vlachos, Sang-Hee Lee, and Pavlos Protopapas. Supporting exact indexing of arbitrarily rotated shapes and periodic time series under euclidean and warping distance measures. *The VLDB Journal—The International Journal on Very Large Data Bases*, 18(3):611–630, 2009.

[82] Kevin Killourhy and Roy Maxion. The effect of clock resolution on keystroke dynamics. In *RAID*, volume 5230, pages 331–350. Springer, 2008.

[83] Kevin S Killourhy and Roy A Maxion. Comparing anomaly-detection algorithms for keystroke dynamics. In *Dependable Systems & Networks, 2009. DSN'09. IEEE/IFIP International Conference on*, pages 125–134. IEEE, 2009.

[84] Kyoung-jae Kim. Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1):307–319, 2003.

[85] Ming Hsiao Ko, Geoff West, Svetha Venkatesh, and Mohan Kumar. Online context recognition in multisensor systems using dynamic time warping. In *Intelligent Sensors, Sensor Networks and Information Processing Conference, 2005. Proceedings of the 2005 International Conference on*, pages 283–288. IEEE, 2005.

[86] Zsolt Miklos Kovacs-Vajna. A fingerprint verification system based on triangular matching and dynamic time warping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1266–1276, 2000.

[87] Wei-Chi Ku and Shuai-Min Chen. Weaknesses and improvements of an efficient password based remote user authentication scheme using smart cards. *IEEE Transactions on Consumer Electronics*, 50(1):204–207, 2004.

[88] Christopher S Leberknight, George R Widmeyer, and Michael L Recce. An investigation into the efficacy of keystroke analysis for perimeter defense and facility access. In *Technologies for Homeland Security, 2008 IEEE Conference on*, pages 345–350. IEEE, 2008.

[89] John Leggett, Glen Williams, Mark Usnick, and Mike Longnecker. Dynamic identity verification via keystroke characteristics. *International Journal of Man-Machine Studies*, 35(6):859–870, 1991.

[90] Vladimir I Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710, 1966.

[91] M. Lichman. UCI machine learning repository, 2013.

[92] Jessica Lin, Eamonn Keogh, Stefano Lonardi, and Bill Chiu. A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, pages 2–11. ACM, 2003.

[93] Jessica Lin, Eamonn Keogh, Li Wei, and Stefano Lonardi. Experiencing sax: a novel symbolic representation of time series. *Data Mining and knowledge discovery*, 15(2):107–144, 2007.

[94] Rake& Agrawal King-lp Lin and Harpreet S Sawhney Kyuseok Shim. Fast similarity search in the presence of noise, scaling, and translation in time-series databases. In *Proceeding of the 21th International Conference on Very Large Data Bases*, pages 490–501. Citeseer, 1995.

[95] Jason Lines, Luke M Davis, Jon Hills, and Anthony Bagnall. A shapelet transform for time series classification. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 289–297. ACM, 2012.

[96] Simon Liu and Mark Silverman. A practical guide to biometric security technology. *IT Professional*, 3(1):27–32, 2001.

[97] Andrew Maas, Chris Heather, Chuong Tom Do, Relly Brandman, Daphne Koller, and Andrew Ng. Offering verified credentials in massive open online courses: Moocs and technology to advance learning and learning research (ubiquity symposium). *Ubiquity*, 2014(May):2, 2014.

[98] Davide Maltoni, Dario Maio, Anil Jain, and Salil Prabhakar. *Handbook of fingerprint recognition*. Springer Science & Business Media, 2009.

[99] Salvador Mandujano and Rogelio Soto. Deterring password sharing: User authentication via fuzzy c-means clustering applied to keystroke biometric data. In *Proceedings of the Fifth Mexican International Conference in Computer Science, 2004. ENC 2004.*, pages 181–187. IEEE, 2004.

[100] Jani Mantyjarvi, Mikko Lindholm, Elena Vildjiounaite, S-M Makela, and HA Ailisto. Identifying users of portable devices from gait pattern with accelerometers. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, volume 2, pages ii–973. IEEE, 2005.

[101] Stephen M Matyas and Jeff Stapleton. A biometric standard for information management and security. *Computers & Security*, 19(5):428–441, 2000.

[102] Vasileios Megalooikonomou, Qiang Wang, Guo Li, and Christos Faloutsos. A multiresolution symbolic representation of time series. In *Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on*, pages 668–679. IEEE, 2005.

[103] Benjamin Miller. Vital signs of identity [biometrics]. *IEEE spectrum*, 31(2):22–30, 1994.

[104] Fabian Moerchen, Ingo Mierswa, and Alfred Ultsch. Understandable models of music collections based on exhaustive feature generation with temporal statistics. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 882–891. ACM, 2006.

[105] Assad Moini and Azad M Madni. Leveraging biometrics for user authentication in online learning: a systems perspective. *IEEE Systems Journal*, 3(4):469–476, 2009.

[106] Fabian Monrose, Michael K Reiter, and Susanne Wetzel. Password hardening based on keystroke dynamics. *International journal of Information security*, 1(2):69–83, 2002.

[107] Fabian Monrose and Aviel Rubin. Authentication via keystroke dynamics. In *Proceedings of the 4th ACM conference on Computer and communications security*, pages 48–56. ACM, 1997.

[108] Fabian Monrose and Aviel D Rubin. Keystroke dynamics as a biometric for authentication. *Future Generation computer systems*, 16(4):351–359, 2000.

[109] Fabian Monrosea and Aviel D. Rubinb. Keystroke dynamics as a biometric for authentication. *Future Generation Computer Systems*, 16(4):351–359, 2000.

[110] Jugurta Montalvão, Eduardo O Freire, Murilo A Bezerra Jr, and Rodolfo Garcia. Contributions to empirical analysis of keystroke dynamics in passwords. *Pattern Recognition Letters*, 52:80–86, 2015.

[111] Jugurta R Montalvão Filho and Eduardo O Freire. On the equalization of keystroke timing histograms. *Pattern Recognition Letters*, 27(13):1440–1446, 2006.

[112] Masahiro Motoyoshi, Takao Miura, and Kohei Watanabe. Mining temporal classes from time series data. In *Proceedings of the eleventh international conference on Information and knowledge management*, pages 493–498. ACM, 2002.

[113] Meinard Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

[114] Meinard Müller. Analysis and retrieval techniques for motion and music data. In *Eurographics (Tutorials)*, pages 249–255, 2009.

[115] Ralph Niels, Louis Vuurpijl, et al. Using dynamic time warping for intuitive handwriting recognition. In *Proc. IGS*, pages 217–221, 2005.

[116] Vit Niennattrakul and Chotirat Ann Ratanamahatana. Shape averaging under time warping. In *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2009. ECTI-CON 2009. 6th International Conference on*, volume 2, pages 626–629. IEEE, 2009.

[117] Mohammad S Obaidat and Balqies Sadoun. Verification of computer users using keystroke dynamics. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 27(2):261–269, 1997.

[118] University of Liverpool. About university of liverpool online degrees and programmes. https://www.online.liverpool.ac.uk/about. Accessed: 2017-07-30.

[119] University of Liverpool. Global presence. https://www.online.liverpool.ac.uk/about/global-presence. Accessed: 2017-07-30.

[120] Akio Ogihara, H Matsumuar, and Akira Shiozaki. Biometric verification using keystroke motion and key press timing for atm user authentication. In *Intelligent Signal Processing and Communications, 2006. ISPACS'06. International Symposium on*, pages 223–226. IEEE, 2006.

[121] Lawrence O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, 2003.

[122] Kai Ou-Yang, Wenyan Jia, Pin Zhou, and Xin Meng. A new approach to transforming time series into symbolic sequences. In *[Engineering in Medicine and Biology, 1999. 21st Annual Conference and the 1999 Annual Fall Meetring of the Biomedical Engineering Society] BMES/EMBS Conference, 1999. Proceedings of the First Joint*, volume 2, pages 974–vol. IEEE, 1999.

[123] Chris Parr. Mooc creators criticise courses' lack of creativity. https://www.timeshighereducation.com/news/mooc-creators-criticise-courses-lack-of-creativity/2008180.article. Accessed: 2017-07-28.

[124] Alen Peacock, Xian Ke, and Matthew Wilkerson. Typing patterns: A key to user identification. *IEEE Security & Privacy*, 2(5):40–47, 2004.

[125] C-S Perng, Haixun Wang, Sylvia R Zhang, and D Stott Parker. Landmarks: a new model for similarity-based pattern querying in time series databases. In *Data Engineering, 2000. Proceedings. 16th International Conference on*, pages 33–42. IEEE, 2000.

[126] P Jonathon Phillips, Hyeonjoon Moon, Syed A Rizvi, and Patrick J Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on pattern analysis and machine intelligence*, 22(10):1090–1104, 2000.

[127] Terence Picton, Melvyn Hunt, Richard Mowrey, Rosendo Rodriguez, and Jagdish Maru. Evaluation of brain-stem auditory evoked potentials using dynamic time warping. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 71(3):212–225, 1988.

[128] Salil Prabhakar, Sharath Pankanti, and Anil K Jain. Biometric recognition: Security and privacy concerns. *IEEE security & privacy*, 99(2):33–42, 2003.

[129] William Pressly Jr. Tspad: a tablet-pc based application for annotation and collaboration on time series data. In *Proceedings of the 46th Annual Southeast Regional Conference on XX*, pages 166–171. ACM, 2008.

[130] Lawrence Rabiner and Biing-Hwang Juang. Fundamentals of speech recognition. *Prentice hall*, 1993.

[131] Davood Rafiei and Alberto Mendelzon. Similarity-based queries for time series data. In *ACM SIGMOD Record*, volume 26, pages 13–25. ACM, 1997.

[132] Muneeba Raja and Stephan Sigg. Applicability of rf-based methods for emotion recognition: A survey. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 1–6. IEEE, 2016.

[133] Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. Searching

and mining trillions of time series subsequences under dynamic time warping. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 262–270. ACM, 2012.

[134] Eric Rescorla. *SSL and TLS: designing and building secure systems*, volume 1. Addison-Wesley Reading, 2001.

[135] Kenneth Revett. A bioinformatics based approach to user authentication via keystroke dynamics. *International Journal of Control, Automation and Systems*, 7(1):7–15, 2009.

[136] John A Robinson, VW Liang, JA Michael Chambers, and Christine L MacKenzie. Computer user verification using login string keystroke dynamics. *IEEE transactions on systems, man, and cybernetics-part a: systems and humans*, 28(2):236–241, 1998.

[137] Arun Ross and Anil Jain. Information fusion in biometrics. *Pattern recognition letters*, 24(13):2115–2125, 2003.

[138] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization fro spoken word recognition. *IEEE Trans. Acoustics, Speech, and Signal Processing*, pages 43–49, 1978.

[139] Toshiharu Samura and Haruhiko Nishimura. Keystroke timing analysis for personal authentication in japanese long text input. In *SICE Annual Conference (SICE), 2011 Proceedings of*, pages 2121–2126. IEEE, 2011.

[140] Ravi Sandhu and Pierangela Samarati. Authentication, access control, and audit. *ACM Computing Surveys (CSUR)*, 28(1):241–243, 1996.

[141] Premkumar Saravanan, Samuel Clarke, Duen Horng Polo Chau, and Hongyuan Zha. Latentgesture: active user authentication through background touch analysis. In *Proceedings of the Second International Symposium of Chinese CHI*, pages 110–113. ACM, 2014.

[142] SJ Shepherd. Continuous authentication by analysis of keyboard typing characteristics. In *Security and Detection, 1995., European Convention on*, pages 111–114. IET, 1995.

[143] Amit Singhal. Modern information retrieval: A brief overview. *IEEE Data Eng. Bull.*, 24(4):35–43, 2001.

[144] Dawn Xiaodong Song, David Wagner, and Xuqing Tian. Timing analysis of keystrokes and timing attacks on ssh. In *USENIX Security Symposium*, volume 2001, 2001.

[145] R Spillane. Keyboard apparatus for personal identification. *IBM Technical Disclosure Bulletin*, 17(3346):3346, 1975.

[146] Wieslaw J Staszewski, Keith Worden, and Geof R Tomlinson. Time–frequency analysis in gearbox fault detection using the wigner–ville distribution and pattern recognition. *Mechanical systems and signal processing*, 11(5):673–692, 1997.

[147] Greg Stocksdale. Nsa glossary of terms used in security and intrusion detection. http://www. sans. org/newlook/resources/glossary. htm, 1998.

[148] Yan Sui, Xukai Zou, and Eliza Yingzi Du. Biometrics-based authentication: A new approach. In *Computer Communications and Networks (ICCCN), 2011 Proceedings of 20th International Conference on*, pages 1–6. IEEE, 2011.

[149] Zahid Ali Syed. *Keystroke and Touch-dynamics Based Authentication for Desktop and Mobile Devices*. PhD thesis, West Virginia University, 2014.

[150] Pin Shen Teh, Andrew Beng Jin Teoh, Thian Song Ong, and Han Foon Neo. Statistical fusion approach on keystroke dynamics. In *Signal-Image Technologies and Internet-Based System, 2007. SITIS'07. Third International IEEE Conference on*, pages 918–923. IEEE, 2007.

[151] Pin Shen Teh, Andrew Beng Jin Teoh, and Shigang Yue. A survey of keystroke dynamics biometrics. *The Scientific World Journal*, 2013, 2013.

[152] Dobromir Todorov. *Mechanics of user identification and authentication: Fundamentals of identity management*. CRC Press, 2007.

[153] Udacity. Scholarships at udacity. https://www.udacity.com/scholarships. Accessed: 2017-08-01.

[154] JA Unar, Woo Chaw Seng, and Almas Abbasi. A review of biometric technology along with trends and prospects. *Pattern recognition*, 47(8):2673–2688, 2014.

[155] Aleksandar Vakanski, Iraj Mantegh, Andrew Irish, and Farrokh Janabi-Sharifi. Trajectory learning for robot programming by demonstration using hidden markov model and dynamic time warping. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(4):1039–1052, 2012.

[156] Kristof Van Laerhoven, Eugen Berlin, and Bernt Schiele. Enabling efficient time series analysis for wearable activity data. In *Machine Learning and Applications, 2009. ICMLA'09. International Conference on*, pages 392–397. IEEE, 2009.

[157] Claus Vielhauer and Ralf Steinmetz. Handwriting: Feature correlation analysis for biometric hashes. *EURASIP Journal on Applied Signal Processing*, 2004:542–558, 2004.

[158] Mary Villani, Charles Tappert, Giang Ngo, Justin Simone, H St Fort, and Sung-Hyuk Cha. Keystroke biometric recognition studies on long-text input under ideal and application-oriented conditions. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 39–39. IEEE, 2006.

[159] Michail Vlachos, Christopher Meek, Zografoula Vagena, and Dimitrios Gunopulos. Identifying similarities, periodicities and bursts for online search queries. In *Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pages 131–142. ACM, 2004.

[160] Stein Emil Vollset. Confidence intervals for a binomial proportion. *Statistics in medicine*, 12(9):809–824, 1993.

[161] Esra Vural, Jiaju Huang, Daqing Hou, and Stephanie Schuckers. Shared research dataset to support development of keystroke authentication. In *Biometrics (IJCB), 2014 IEEE International Joint Conference on*, pages 1–8. IEEE, 2014.

[162] David Wagner, Bruce Schneier, et al. Analysis of the ssl 3.0 protocol. In *The Second USENIX Workshop on Electronic Commerce Proceedings*, volume 1, pages 29–40, 1996.

[163] Sean Wallis. Binomial confidence intervals and contingency tests: mathematical fundamentals and the evaluation of alternative methods. *Journal of Quantitative Linguistics*, 20(3):178–208, 2013.

[164] Xiaoyue Wang, Abdullah Mueen, Hui Ding, Goce Trajcevski, Peter Scheuermann, and Eamonn Keogh. Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, 26(2):275–309, 2013.

[165] James L Wayman. Fundamentals of biometric authentication technologies. *International Journal of Image and Graphics*, 1(01):93–113, 2001.

[166] Richard P Wildes. Iris recognition: an emerging biometric technology. *Proceedings of the IEEE*, 85(9):1348–1363, 1997.

[167] Pei-Yuan Wu, Chi-Chen Fang, J Morris Chang, Stephen B Gilbert, and SY Kung. Cost-effective kernel ridge regression implementation for keystroke-based active authentication system. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 6028–6032. IEEE, 2014.

[168] Zhijia Yang and G Zhao. Application of symbolic techniques in detecting determinism in time series. In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 20, pages 2670–2673, 1998.

[169] Lexiang Ye and Eamonn Keogh. Time series shapelets: a new primitive for data mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 947–956. ACM, 2009.

[170] Byoung-Kee Yi and Christos Faloutsos. Fast time sequence indexing for arbitrary lp norms. VLDB, 2000.

[171] Yaodong Zhang and James R Glass. An inner-product lower-bound estimate for dynamic time warping. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 5660–5663. IEEE, 2011.

[172] Yu Zhong, Yunbin Deng, and Anil K Jain. Keystroke dynamics for user authentication. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 117–123. IEEE, 2012.

Appendix A

Further Results Concerning OKCA Approach Presented in Chapter 5

A.1 Overview

This appendix presents some additional results concerning the experiments using the OKCA approach presented in Chapter 5. In particular, the ranked *Sim* matrix results, as introduced in Table 5.2, for all combination groups, across all data sets. For ease of understanding, Table A.1 shows the structure (interpretation) of *Sim* ranking tables as given in this appendix.

TABLE A.1: Structure of tables presented in the appendix.

Data set	Feature	Combination Group	Table
ACB	$\mathcal{F}^t$	$a \vee \{b, c\}$	Table A.2
		$b \vee \{a, c\}$	Table A.3
		$c \vee \{a, b\}$	Table A.4
	$\mathcal{KH}^t$	$a \vee \{b, c\}$	Table A.5
		$b \vee \{a, c\}$	Table A.6
		$c \vee \{a, b\}$	Table A.7
	Multi.	$a \vee \{b, c\}$	Table A.8
		$b \vee \{a, c\}$	Table A.9
		$c \vee \{a, b\}$	Table A.10
	FVR	$a \vee \{b, c\}$	Table A.11
		$b \vee \{a, c\}$	Table A.12
		$c \vee \{a, b\}$	Table A.13
GP	$\mathcal{F}^t$	$a \vee \{b, c\}$	Table A.14
		$b \vee \{a, c\}$	Table A.15

Continued on next page

		$c \vee \{a, b\}$	Table A.16
	FVR	$a \vee \{b, c\}$	Table A.17
		$b \vee \{a, c\}$	Table A.18
		$c \vee \{a, b\}$	Table A.19
VHHS	$\mathcal{F}^t$	$a \vee \{b, c\}$	Table A.20
		$b \vee \{a, c\}$	Table A.21
		$c \vee \{a, b\}$	Table A.22
	$\mathcal{KH}^t$	$a \vee \{b, c\}$	Table A.23
		$b \vee \{a, c\}$	Table A.24
		$c \vee \{a, b\}$	Table A.25
	Multi.	$a \vee \{b, c\}$	Table A.26
		$b \vee \{a, c\}$	Table A.27
		$c \vee \{a, b\}$	Table A.28
	FVR	$a \vee \{b, c\}$	Table A.29
		$b \vee \{a, c\}$	Table A.30
		$c \vee \{a, b\}$	Table A.31

Table A.2: Ranked *sim* values for $(a \vee \{b, c\})$ in ACB dataset using U-KTS $\mathcal{F}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group b and c	0.0528	0.0329	0.0635	0.0490	0.0725	0.0314	0.0782	0.0613	0.0467	0.0974
	0.0578	0.0374	0.0655	0.0520	0.0746	0.0413	0.0808	0.0614	0.0469	0.1097
	0.0590	0.0401	0.0665	0.0530	0.0750	0.0447	0.0823	0.0636	0.0469	0.1187
	0.0598	0.0484	0.0683	0.0557	0.0751	0.0479	0.0885	0.0683	0.0512	0.1198
	0.0621	0.0487	0.0688	0.0579	0.0753	0.0501	0.0889	0.0685	0.0524	0.1198
	0.0631	0.0527	0.0719	0.0592	0.0755	0.0515	0.0897	0.0700	0.0532	0.1207
	0.0635	0.0549	0.0740	0.0619	0.0798	0.0539	0.0927	0.0702	0.0587	0.1221
	0.0646	0.0550	0.0741	0.0668	0.0843	0.0558	0.0928	0.0702	0.0603	0.1233
	0.0652	0.0551	0.0772	0.0685	0.0858	0.0582	0.0957	0.0704	0.0606	0.1242
	0.0658	0.0575	0.0773	0.0691	0.0865	0.0583	0.0992	0.0719	0.0613	0.1262
	0.0660	0.0581	0.0775	0.0710	0.0889	0.0585	0.1020	0.0741	0.0644	0.1327
	0.0682	0.0583	0.0778	0.0712	0.0892	0.0613	0.1037	0.0745	0.0645	0.1346
	0.0698	0.0640	0.0780	0.0713	0.0908	0.0634	0.1039	0.0769	0.0705	0.1369
	0.0725	0.0721	0.0781	0.0730	0.0917	0.0686	0.1054	0.0770	0.0744	0.1386
	0.0739	0.0732	0.0790	0.0756	0.0918	0.0686	0.1061	0.0794	0.0756	0.1414
	0.0748	0.0750	0.0834	0.0767	0.0926	0.0720	0.1065	0.0815	0.0766	0.1436
	0.0749	0.0799	0.0836	0.0786	0.0926	0.0764	0.1066	0.0819	0.0770	0.1446
	0.0753	0.0839	0.0846	0.0789	0.0933	0.0807	0.1119	0.0863	0.0891	0.1449
	0.0756	0.0906	0.0846	0.0837	0.0970	0.0854	0.1122	0.0874	0.0898	0.1504
	0.0805	0.0910	0.0847	0.0850	0.0978	0.0915	0.1135	0.0881	0.0899	0.1559
	0.0838	0.0961	0.0851	0.0864	0.0986	0.0999	0.1136	0.0935	0.0931	0.1626
	0.0883	0.1055	0.0864	0.0882	0.1026	0.1012	0.1186	0.0949	0.0931	0.1641
	0.0883	0.1063	0.0864	0.0882	0.1035	0.1045	0.1221	0.0949	0.0935	0.1645
	0.0892	0.1063	0.0865	0.0894	0.1068	0.1076	0.1241	0.0956	0.0950	0.1652
	0.0902	0.1099	0.0865	0.0914	0.1138	0.1076	0.1258	0.0959	0.0977	0.1690
	0.0907	0.1123	0.0870	0.0924	0.1151	0.1087	0.1261	0.0961	0.1058	0.1726
	0.0939	0.1195	0.0924	0.0940	0.1182	0.1124	0.1261	0.0962	0.1065	0.1859
	0.0984	0.1232	0.1118	0.0947	0.1195	0.1214	0.1307	0.0974	0.1235	0.1864
	0.1137	0.1307	0.1302	0.0963	0.1211	0.1218	0.1336	0.0978	0.1330	0.1944
	0.1177	0.1415	0.1310	0.1002	0.1225	0.1235	0.1513	0.1011	0.1453	0.2155
r'	**1**	**2**	**1**	**1**	**2**	**1**	**5**	**1**	**6**	**2**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group b and c	0.0569	0.0500	0.0369	0.0343	0.0472	0.0396	0.0361	0.0470	0.0365	0.0588
	0.0613	0.0548	0.0391	0.0377	0.0517	0.0442	0.0384	0.0480	0.0387	0.0624
	0.0629	0.0551	0.0412	0.0451	0.0529	0.0456	0.0388	0.0490	0.0403	0.0670
	0.0634	0.0598	0.0423	0.0479	0.0539	0.0456	0.0446	0.0491	0.0410	0.0675
	0.0642	0.0670	0.0460	0.0482	0.0560	0.0461	0.0447	0.0528	0.0447	0.0678
	0.0650	0.0680	0.0477	0.0511	0.0561	0.0467	0.0480	0.0543	0.0489	0.0731
	0.0663	0.0693	0.0487	0.0520	0.0570	0.0472	0.0499	0.0572	0.0536	0.0755
	0.0667	0.0708	0.0509	0.0538	0.0573	0.0480	0.0504	0.0573	0.0554	0.0795
	0.0669	0.0719	0.0510	0.0550	0.0577	0.0502	0.0514	0.0575	0.0579	0.0797
	0.0703	0.0720	0.0532	0.0618	0.0578	0.0513	0.0551	0.0586	0.0587	0.0797
	0.0711	0.0759	0.0537	0.0647	0.0609	0.0518	0.0555	0.0605	0.0594	0.0802
	0.0719	0.0766	0.0545	0.0662	0.0615	0.0521	0.0601	0.0622	0.0628	0.0803
	0.0741	0.0772	0.0592	0.0681	0.0670	0.0607	0.0614	0.0630	0.0643	0.0804
	0.0751	0.0796	0.0595	0.0691	0.0676	0.0607	0.0633	0.0657	0.0655	0.0807
	0.0751	0.0805	0.0606	0.0745	0.0689	0.0609	0.0654	0.0685	0.0689	0.0817
	0.0762	0.0806	0.0656	0.0803	0.0691	0.0651	0.0740	0.0693	0.0719	0.0847
	0.0775	0.0806	0.0744	0.0827	0.0767	0.0746	0.0882	0.0713	0.0794	0.0884
	0.0778	0.0821	0.0806	0.0858	0.0775	0.0773	0.0929	0.0750	0.0812	0.0886
	0.0840	0.0853	0.0806	0.0904	0.0782	0.0820	0.0968	0.0765	0.0842	0.0893
	0.0889	0.0925	0.0834	0.0916	0.0793	0.0870	0.0973	0.0770	0.0920	0.0900
	0.0890	0.0932	0.0856	0.0930	0.0889	0.0879	0.0978	0.0821	0.0948	0.0907
	0.0893	0.0932	0.0932	0.0939	0.0891	0.0960	0.1001	0.0893	0.0961	0.0961
	0.0921	0.0938	0.0953	0.0956	0.0918	0.0960	0.1058	0.0899	0.0961	0.1035
	0.0921	0.0938	0.0992	0.0956	0.0972	0.0963	0.1065	0.0899	0.0970	0.1049
	0.0922	0.0954	0.0992	0.1018	0.1012	0.0981	0.1065	0.0950	0.0979	0.1053
	0.0922	0.0960	0.1009	0.1041	0.1033	0.1022	0.1088	0.1005	0.0991	0.1106
	0.0955	0.0975	0.1017	0.1059	0.1047	0.1034	0.1160	0.1008	0.1030	0.1121
	0.1010	0.0977	0.1164	0.1153	0.1068	0.1140	0.1281	0.1094	0.1167	0.1166
	0.1029	0.1000	0.1268	0.1224	0.1068	0.1261	0.1349	0.1194	0.1343	0.1166
	0.1085	0.1017	0.1492	0.1357	0.1187	0.1336	0.1530	0.1365	0.1590	0.1265
r'	**3**	**3**	**2**	**1**	**10**	**3**	**2**	**6**	**2**	**2**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group b and c	0.0580	0.0332	0.0622	0.0559	0.0519	0.0465	0.0580	0.0597	0.0385	0.0677
	0.0667	0.0387	0.0668	0.0598	0.0589	0.0533	0.0667	0.0616	0.0392	0.0697
	0.0668	0.0454	0.0687	0.0616	0.0645	0.0543	0.0668	0.0646	0.0435	0.0720
	0.0678	0.0467	0.0722	0.0649	0.0683	0.0549	0.0678	0.0699	0.0477	0.0724
	0.0679	0.0502	0.0748	0.0649	0.0718	0.0554	0.0679	0.0699	0.0490	0.0740
	0.0706	0.0511	0.0766	0.0672	0.0718	0.0558	0.0706	0.0705	0.0680	0.0751
	0.0706	0.0520	0.0767	0.0695	0.0766	0.0565	0.0706	0.0712	0.0714	0.0770
	0.0706	0.0523	0.0768	0.0713	0.0773	0.0565	0.0706	0.0732	0.0761	0.0777
	0.0707	0.0587	0.0779	0.0716	0.0792	0.0565	0.0707	0.0741	0.0765	0.0782
	0.0739	0.0595	0.0780	0.0723	0.0797	0.0579	0.0739	0.0742	0.0797	0.0786
	0.0741	0.0608	0.0792	0.0732	0.0803	0.0614	0.0741	0.0748	0.0814	0.0786
	0.0744	0.0644	0.0794	0.0745	0.0808	0.0615	0.0744	0.0751	0.0819	0.0787
	0.0763	0.0649	0.0803	0.0762	0.0810	0.0617	0.0763	0.0777	0.0896	0.0788
	0.0767	0.0677	0.0806	0.0775	0.0813	0.0654	0.0767	0.0786	0.0898	0.0811
	0.0796	0.0718	0.0810	0.0790	0.0817	0.0666	0.0796	0.0803	0.0920	0.0814
	0.0798	0.0723	0.0810	0.0803	0.0823	0.0674	0.0798	0.0832	0.0952	0.0817
	0.0798	0.0812	0.0811	0.0836	0.0887	0.0730	0.0798	0.0836	0.0990	0.0826
	0.0801	0.0921	0.0812	0.0872	0.0918	0.0741	0.0801	0.0845	0.1098	0.0832
	0.0810	0.0932	0.0824	0.0898	0.0919	0.0771	0.0810	0.0850	0.1134	0.0877
	0.0824	0.0951	0.0827	0.0898	0.0922	0.0835	0.0824	0.0855	0.1135	0.0901
	0.0826	0.0966	0.0827	0.0949	0.0929	0.0838	0.0826	0.0875	0.1135	0.0928
	0.0826	0.1037	0.0829	0.0963	0.0989	0.0854	0.0826	0.0883	0.1170	0.0928
	0.0838	0.1037	0.0875	0.0970	0.0999	0.0878	0.0838	0.0892	0.1200	0.0952
	0.0879	0.1057	0.0905	0.0978	0.1024	0.0956	0.0879	0.0898	0.1236	0.0955
	0.0896	0.1089	0.0971	0.0981	0.1024	0.0956	0.0896	0.1009	0.1243	0.0972
	0.0925	0.1095	0.0977	0.1022	0.1116	0.0986	0.0925	0.1033	0.1288	0.1001
	0.0946	0.1132	0.1001	0.1111	0.1191	0.0993	0.0946	0.1079	0.1345	0.1038
	0.0957	0.1254	0.1148	0.1147	0.1325	0.1060	0.0957	0.1083	0.1649	0.1041
	0.1079	0.1395	0.1203	0.1260	0.1605	0.1142	0.1079	0.1111	0.1677	0.1085
	0.1134	0.1617	0.1252	0.1278	0.1643	0.1291	0.1134	0.1177	0.1747	0.1205
r'	**5**	**1**	**3**	**1**	**7**	**1**	**4**	**1**	**1**	**6**

TABLE A.3: Ranked *sim* values for $(b \vee \{a, c\})$ in ACB dataset using U-KTS $\mathcal{F}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and c	0.0799	0.0769	0.0903	0.1340	0.0810	0.0602	0.0988	0.1046	0.0587	0.1531
	0.0803	0.0785	0.0944	0.1409	0.0905	0.0602	0.1057	0.1113	0.0588	0.1558
	0.0874	0.0830	0.0969	0.1425	0.0935	0.0609	0.1104	0.1133	0.0600	0.1643
	0.0882	0.0844	0.0971	0.1453	0.0958	0.0618	0.1113	0.1147	0.0602	0.1675
	0.0884	0.0858	0.0981	0.1485	0.0963	0.0621	0.1136	0.1156	0.0611	0.1677
	0.0886	0.0858	0.0994	0.1491	0.0977	0.0622	0.1136	0.1156	0.0613	0.1715
	0.0891	0.0866	0.1002	0.1491	0.1016	0.0622	0.1141	0.1163	0.0623	0.1735
	0.0891	0.0881	0.1002	0.1498	0.1019	0.0632	0.1171	0.1163	0.0623	0.1771
	0.0917	0.0885	0.1002	0.1523	0.1022	0.0634	0.1172	0.1189	0.0624	0.1781
	0.0939	0.0890	0.1015	0.1530	0.1031	0.0637	0.1184	0.1192	0.0633	0.1785
	0.0941	0.0898	0.1017	0.1531	0.1031	0.0642	0.1189	0.1193	0.0633	0.1787
	0.0948	0.0900	0.1057	0.1542	0.1033	0.0649	0.1197	0.1205	0.0641	0.1787
	0.0952	0.0904	0.1061	0.1556	0.1059	0.0653	0.1236	0.1207	0.0644	0.1787
	0.0957	0.0907	0.1062	0.1560	0.1064	0.0657	0.1267	0.1213	0.0644	0.1830
	0.0960	0.0912	0.1071	0.1575	0.1076	0.0663	0.1275	0.1221	0.0661	0.1873
	0.0966	0.0912	0.1089	0.1579	0.1102	0.0665	0.1276	0.1231	0.0666	0.1886
	0.0974	0.0918	0.1091	0.1580	0.1113	0.0667	0.1278	0.1232	0.0667	0.1889
	0.0982	0.0921	0.1094	0.1634	0.1127	0.0667	0.1284	0.1233	0.0669	0.1959
	0.0983	0.0924	0.1110	0.1642	0.1151	0.0673	0.1284	0.1241	0.0672	0.1972
	0.0988	0.0925	0.1111	0.1651	0.1151	0.0676	0.1325	0.1249	0.0679	0.2003
	0.0998	0.0931	0.1127	0.1658	0.1152	0.0701	0.1351	0.1268	0.0688	0.2035
	0.1009	0.0950	0.1127	0.1679	0.1166	0.0707	0.1363	0.1289	0.0705	0.2049
	0.1025	0.0954	0.1129	0.1696	0.1179	0.0728	0.1369	0.1298	0.0708	0.2059
	0.1033	0.0963	0.1142	0.1696	0.1189	0.0729	0.1373	0.1311	0.0709	0.2068
	0.1035	0.0991	0.1156	0.1706	0.1203	0.0730	0.1441	0.1315	0.0713	0.2071
	0.1046	0.0997	0.1181	0.1717	0.1210	0.0751	0.1455	0.1321	0.0734	0.2088
	0.1091	0.1061	0.1186	0.1721	0.1223	0.0754	0.1462	0.1342	0.0758	0.2133
	0.1123	0.1071	0.1187	0.1763	0.1225	0.0819	0.1487	0.1360	0.0803	0.2153
	0.1258	0.1194	0.1258	0.1763	0.1258	0.0848	0.1519	0.1372	0.1096	0.2154
	0.1266	0.1317	0.1278	0.1875	0.1381	0.0892	0.1520	0.1432	0.1144	0.2156
r'	**1**	**1**	**1**	**1**	**2**	**2**	**4**	**1**	**6**	**2**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and c	0.1043	0.0920	0.0542	0.0807	0.0795	0.0542	0.0924	0.0613	0.0529	0.1253
	0.1049	0.0982	0.0562	0.0860	0.0839	0.0551	0.0933	0.0615	0.0546	0.1254
	0.1068	0.0984	0.0583	0.0864	0.0843	0.0563	0.0935	0.0630	0.0565	0.1279
	0.1087	0.1010	0.0583	0.0871	0.0846	0.0576	0.0938	0.0635	0.0573	0.1280
	0.1120	0.1010	0.0586	0.0872	0.0879	0.0581	0.0954	0.0641	0.0581	0.1336
	0.1128	0.1010	0.0605	0.0872	0.0890	0.0585	0.0971	0.0642	0.0591	0.1336
	0.1128	0.1018	0.0605	0.0873	0.0893	0.0589	0.1004	0.0643	0.0594	0.1343
	0.1129	0.1028	0.0608	0.0881	0.0909	0.0592	0.1004	0.0650	0.0600	0.1346
	0.1150	0.1041	0.0613	0.0904	0.0911	0.0595	0.1004	0.0654	0.0607	0.1357
	0.1168	0.1042	0.0615	0.0918	0.0913	0.0609	0.1005	0.0666	0.0607	0.1365
	0.1168	0.1046	0.0620	0.0918	0.0918	0.0614	0.1027	0.0676	0.0607	0.1372
	0.1176	0.1046	0.0620	0.0927	0.0919	0.0614	0.1034	0.0677	0.0611	0.1379
	0.1202	0.1063	0.0621	0.0931	0.0919	0.0620	0.1044	0.0682	0.0620	0.1382
	0.1204	0.1082	0.0622	0.0943	0.0934	0.0622	0.1045	0.0695	0.0623	0.1392
	0.1212	0.1094	0.0626	0.0944	0.0948	0.0630	0.1048	0.0701	0.0627	0.1398
	0.1215	0.1096	0.0627	0.0957	0.0948	0.0630	0.1053	0.0735	0.0630	0.1401
	0.1222	0.1112	0.0628	0.0967	0.0971	0.0634	0.1054	0.0737	0.0632	0.1403
	0.1224	0.1115	0.0636	0.0973	0.0974	0.0634	0.1060	0.0738	0.0652	0.1444
	0.1230	0.1118	0.0641	0.0979	0.0974	0.0639	0.1071	0.0740	0.0657	0.1456
	0.1251	0.1119	0.0648	0.0991	0.0996	0.0652	0.1073	0.0742	0.0663	0.1463
	0.1286	0.1122	0.0648	0.1002	0.0999	0.0653	0.1078	0.0742	0.0664	0.1474
	0.1293	0.1147	0.0651	0.1012	0.1012	0.0653	0.1101	0.0749	0.0665	0.1474
	0.1300	0.1147	0.0660	0.1013	0.1014	0.0660	0.1102	0.0760	0.0686	0.1478
	0.1312	0.1159	0.0660	0.1018	0.1036	0.0669	0.1102	0.0771	0.0694	0.1482
	0.1336	0.1169	0.0663	0.1023	0.1037	0.0684	0.1108	0.0771	0.0727	0.1526
	0.1340	0.1177	0.0705	0.1031	0.1039	0.0707	0.1172	0.0778	0.0728	0.1536
	0.1342	0.1201	0.0747	0.1068	0.1041	0.0739	0.1193	0.0811	0.0741	0.1548
	0.1350	0.1236	0.0750	0.1089	0.1146	0.0754	0.1231	0.0879	0.0787	0.1554
	0.1373	0.1242	0.0989	0.1165	0.1147	0.0893	0.1239	0.1239	0.1164	0.1603
	0.1447	0.1319	0.1142	0.1171	0.1190	0.1105	0.1299	0.1338	0.1290	0.1670
r'	**3**	**3**	**2**	**1**	**9**	**4**	**2**	**6**	**2**	**3**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and c	0.1096	0.0712	0.1366	0.1235	0.1475	0.0910	0.1096	0.1115	0.1026	0.1080
	0.1110	0.0723	0.1422	0.1297	0.1489	0.0966	0.1110	0.1119	0.1027	0.1081
	0.1161	0.0751	0.1453	0.1354	0.1504	0.0982	0.1161	0.1148	0.1035	0.1105
	0.1176	0.0753	0.1453	0.1367	0.1507	0.0984	0.1176	0.1181	0.1054	0.1107
	0.1182	0.0757	0.1453	0.1372	0.1507	0.1005	0.1182	0.1184	0.1056	0.1113
	0.1189	0.0769	0.1471	0.1372	0.1531	0.1031	0.1189	0.1184	0.1065	0.1116
	0.1199	0.0769	0.1474	0.1375	0.1532	0.1031	0.1199	0.1197	0.1074	0.1116
	0.1199	0.0778	0.1482	0.1389	0.1538	0.1032	0.1199	0.1211	0.1077	0.1119
	0.1201	0.0784	0.1483	0.1401	0.1542	0.1044	0.1201	0.1216	0.1077	0.1143
	0.1203	0.0784	0.1483	0.1402	0.1570	0.1045	0.1203	0.1218	0.1085	0.1145
	0.1213	0.0787	0.1490	0.1431	0.1574	0.1047	0.1213	0.1223	0.1099	0.1146
	0.1222	0.0791	0.1543	0.1448	0.1576	0.1056	0.1222	0.1256	0.1103	0.1158
	0.1238	0.0793	0.1557	0.1486	0.1592	0.1076	0.1238	0.1264	0.1124	0.1159
	0.1259	0.0793	0.1567	0.1507	0.1594	0.1092	0.1259	0.1267	0.1132	0.1171
	0.1269	0.0796	0.1581	0.1515	0.1624	0.1097	0.1269	0.1269	0.1147	0.1173
	0.1271	0.0797	0.1585	0.1516	0.1627	0.1107	0.1271	0.1277	0.1155	0.1177
	0.1277	0.0800	0.1587	0.1529	0.1628	0.1129	0.1277	0.1309	0.1162	0.1185
	0.1287	0.0807	0.1603	0.1545	0.1672	0.1132	0.1287	0.1311	0.1163	0.1185
	0.1295	0.0812	0.1607	0.1553	0.1673	0.1146	0.1295	0.1324	0.1164	0.1190
	0.1299	0.0816	0.1609	0.1553	0.1681	0.1151	0.1299	0.1324	0.1169	0.1190
	0.1300	0.0818	0.1615	0.1557	0.1684	0.1157	0.1300	0.1328	0.1172	0.1229
	0.1311	0.0821	0.1625	0.1582	0.1688	0.1159	0.1311	0.1339	0.1193	0.1239
	0.1318	0.0824	0.1629	0.1605	0.1699	0.1161	0.1318	0.1355	0.1209	0.1241
	0.1336	0.0835	0.1638	0.1613	0.1701	0.1163	0.1336	0.1356	0.1223	0.1242
	0.1337	0.0842	0.1643	0.1614	0.1741	0.1171	0.1337	0.1364	0.1235	0.1272
	0.1367	0.0843	0.1663	0.1617	0.1748	0.1179	0.1367	0.1374	0.1242	0.1274
	0.1374	0.0868	0.1664	0.1633	0.1760	0.1211	0.1374	0.1402	0.1246	0.1290
	0.1394	0.0985	0.1678	0.1643	0.1764	0.1299	0.1394	0.1440	0.1270	0.1305
	0.1414	0.1267	0.1703	0.1669	0.1777	0.1318	0.1414	0.1450	0.1305	0.1315
	0.1416	0.1269	0.1723	0.1690	0.1802	0.1318	0.1416	0.1500	0.1432	0.1383
r'	**5**	**1**	**3**	**1**	**5**	**1**	**3**	**2**	**2**	**4**

Table A.4: Ranked *sim* values for $(c \vee \{a, b\})$ in ACB dataset using U-KTS $\mathcal{F}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and b	0.0829	0.0769	0.0941	0.1353	0.0869	0.0573	0.0913	0.1048	0.0594	0.1422
	0.0868	0.0812	0.0946	0.1358	0.0919	0.0600	0.0961	0.1099	0.0599	0.1438
	0.0870	0.0813	0.0960	0.1371	0.0924	0.0601	0.1070	0.1127	0.0608	0.1517
	0.0888	0.0841	0.0970	0.1396	0.0930	0.0621	0.1077	0.1130	0.0609	0.1546
	0.0889	0.0851	0.1003	0.1449	0.0966	0.0624	0.1094	0.1130	0.0615	0.1581
	0.0901	0.0855	0.1003	0.1480	0.0970	0.0633	0.1167	0.1136	0.0627	0.1680
	0.0903	0.0855	0.1004	0.1511	0.0971	0.0643	0.1178	0.1144	0.0642	0.1696
	0.0904	0.0859	0.1016	0.1516	0.0972	0.0648	0.1205	0.1149	0.0648	0.1711
	0.0915	0.0863	0.1033	0.1517	0.0981	0.0653	0.1217	0.1154	0.0648	0.1758
	0.0917	0.0863	0.1036	0.1522	0.0982	0.0659	0.1225	0.1164	0.0649	0.1773
	0.0920	0.0866	0.1043	0.1523	0.0986	0.0675	0.1225	0.1169	0.0651	0.1787
	0.0923	0.0877	0.1043	0.1546	0.1009	0.0683	0.1226	0.1177	0.0666	0.1816
	0.0925	0.0880	0.1047	0.1553	0.1045	0.0685	0.1238	0.1179	0.0680	0.1828
	0.0925	0.0886	0.1057	0.1555	0.1056	0.0686	0.1243	0.1183	0.0683	0.1828
	0.0938	0.0889	0.1061	0.1559	0.1063	0.0688	0.1251	0.1184	0.0683	0.1834
	0.0949	0.0899	0.1064	0.1559	0.1081	0.0689	0.1254	0.1196	0.0685	0.1834
	0.0951	0.0904	0.1064	0.1559	0.1083	0.0698	0.1267	0.1201	0.0685	0.1839
	0.0952	0.0916	0.1065	0.1569	0.1092	0.0700	0.1278	0.1202	0.0687	0.1845
	0.0953	0.0923	0.1084	0.1598	0.1092	0.0708	0.1291	0.1215	0.0693	0.1865
	0.0969	0.0924	0.1085	0.1599	0.1092	0.0711	0.1291	0.1223	0.0704	0.1876
	0.0972	0.0930	0.1096	0.1607	0.1102	0.0714	0.1312	0.1224	0.0714	0.1899
	0.0990	0.0933	0.1100	0.1610	0.1104	0.0738	0.1312	0.1233	0.0716	0.1904
	0.0999	0.0968	0.1101	0.1614	0.1114	0.0738	0.1335	0.1240	0.0720	0.1951
	0.1020	0.0983	0.1102	0.1616	0.1118	0.0740	0.1357	0.1261	0.0724	0.1956
	0.1028	0.0995	0.1104	0.1620	0.1192	0.0758	0.1375	0.1263	0.0726	0.1974
	0.1063	0.1005	0.1104	0.1640	0.1192	0.0758	0.1385	0.1285	0.0726	0.2008
	0.1070	0.1015	0.1123	0.1672	0.1253	0.0764	0.1435	0.1314	0.0837	0.2065
	0.1089	0.1063	0.1177	0.1772	0.1275	0.0785	0.1450	0.1321	0.0876	0.2078
	0.1107	0.1080	0.1178	0.1781	0.1280	0.0808	0.1470	0.1358	0.0909	0.2138
	0.1132	0.1235	0.1187	0.1852	0.1347	0.0829	0.1610	0.1371	0.1067	0.2186
r'	**2**	**2**	**1**	**2**	**2**	**1**	**2**	**1**	**4**	**2**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and b	0.1030	0.0895	0.0525	0.0837	0.0821	0.0561	0.0908	0.0616	0.0566	0.1175
	0.1070	0.0954	0.0571	0.0848	0.0832	0.0569	0.0929	0.0635	0.0567	0.1197
	0.1101	0.0956	0.0577	0.0870	0.0839	0.0577	0.0931	0.0636	0.0577	0.1234
	0.1119	0.0961	0.0579	0.0873	0.0877	0.0580	0.0950	0.0639	0.0579	0.1236
	0.1127	0.0981	0.0604	0.0876	0.0888	0.0593	0.0986	0.0649	0.0583	0.1249
	0.1133	0.0981	0.0608	0.0876	0.0888	0.0594	0.0986	0.0655	0.0586	0.1310
	0.1135	0.0989	0.0612	0.0891	0.0889	0.0605	0.0988	0.0675	0.0588	0.1349
	0.1136	0.0996	0.0613	0.0901	0.0891	0.0616	0.0989	0.0676	0.0592	0.1351
	0.1150	0.0997	0.0613	0.0922	0.0897	0.0618	0.0992	0.0677	0.0597	0.1359
	0.1150	0.1003	0.0613	0.0925	0.0906	0.0618	0.0996	0.0677	0.0603	0.1367
	0.1176	0.1018	0.0616	0.0931	0.0910	0.0623	0.1005	0.0680	0.0609	0.1371
	0.1181	0.1023	0.0624	0.0932	0.0910	0.0624	0.1011	0.0680	0.0609	0.1390
	0.1182	0.1025	0.0624	0.0932	0.0910	0.0638	0.1015	0.0691	0.0611	0.1396
	0.1183	0.1030	0.0629	0.0940	0.0924	0.0648	0.1018	0.0699	0.0619	0.1398
	0.1184	0.1057	0.0632	0.0940	0.0937	0.0648	0.1026	0.0701	0.0626	0.1405
	0.1184	0.1072	0.0637	0.0948	0.0943	0.0652	0.1033	0.0707	0.0633	0.1405
	0.1200	0.1075	0.0638	0.0955	0.0947	0.0656	0.1034	0.0708	0.0645	0.1408
	0.1202	0.1078	0.0652	0.0956	0.0958	0.0661	0.1041	0.0719	0.0645	0.1411
	0.1207	0.1094	0.0654	0.0968	0.0958	0.0676	0.1051	0.0736	0.0645	0.1413
	0.1210	0.1106	0.0656	0.0969	0.0964	0.0686	0.1070	0.0740	0.0670	0.1415
	0.1214	0.1107	0.0663	0.0986	0.0976	0.0700	0.1075	0.0741	0.0673	0.1423
	0.1214	0.1107	0.0663	0.0987	0.0978	0.0707	0.1082	0.0765	0.0677	0.1426
	0.1225	0.1117	0.0664	0.1012	0.0989	0.0717	0.1085	0.0771	0.0692	0.1426
	0.1256	0.1134	0.0672	0.1014	0.0992	0.0717	0.1085	0.0799	0.0699	0.1439
	0.1257	0.1138	0.0677	0.1037	0.1009	0.0743	0.1093	0.0799	0.0701	0.1461
	0.1269	0.1153	0.0709	0.1053	0.1033	0.0746	0.1130	0.0821	0.0704	0.1544
	0.1289	0.1181	0.0754	0.1057	0.1040	0.0781	0.1147	0.0899	0.0798	0.1563
	0.1300	0.1206	0.0824	0.1074	0.1076	0.0807	0.1155	0.0949	0.0835	0.1564
	0.1358	0.1215	0.0833	0.1082	0.1114	0.0834	0.1205	0.1023	0.0845	0.1577
	0.1434	0.1249	0.1026	0.1108	0.1137	0.0925	0.1212	0.1171	0.0946	0.1609
r'	**2**	**3**	**1**	**1**	**7**	**3**	**4**	**5**	**1**	**3**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and b	0.1088	0.0729	0.1407	0.1275	0.1425	0.0912	0.1088	0.1078	0.1041	0.1022
	0.1088	0.0750	0.1411	0.1304	0.1449	0.0979	0.1088	0.1144	0.1089	0.1068
	0.1120	0.0753	0.1438	0.1314	0.1503	0.0994	0.1120	0.1154	0.1091	0.1092
	0.1129	0.0756	0.1479	0.1337	0.1528	0.1000	0.1129	0.1165	0.1095	0.1095
	0.1153	0.0759	0.1490	0.1369	0.1556	0.1009	0.1153	0.1205	0.1099	0.1099
	0.1161	0.0759	0.1497	0.1369	0.1562	0.1021	0.1161	0.1212	0.1101	0.1110
	0.1168	0.0762	0.1498	0.1375	0.1565	0.1028	0.1168	0.1213	0.1103	0.1134
	0.1194	0.0764	0.1500	0.1395	0.1569	0.1045	0.1194	0.1216	0.1103	0.1135
	0.1204	0.0769	0.1500	0.1399	0.1571	0.1046	0.1204	0.1218	0.1103	0.1135
	0.1206	0.0769	0.1507	0.1402	0.1576	0.1046	0.1206	0.1220	0.1104	0.1135
	0.1225	0.0769	0.1521	0.1409	0.1590	0.1050	0.1225	0.1220	0.1105	0.1137
	0.1226	0.0781	0.1525	0.1416	0.1591	0.1055	0.1226	0.1225	0.1106	0.1140
	0.1227	0.0789	0.1547	0.1416	0.1597	0.1056	0.1227	0.1230	0.1112	0.1140
	0.1230	0.0789	0.1549	0.1467	0.1597	0.1061	0.1230	0.1237	0.1127	0.1147
	0.1233	0.0791	0.1551	0.1468	0.1614	0.1073	0.1233	0.1252	0.1129	0.1150
	0.1237	0.0795	0.1553	0.1476	0.1615	0.1086	0.1237	0.1252	0.1130	0.1164
	0.1249	0.0803	0.1555	0.1479	0.1625	0.1086	0.1249	0.1257	0.1138	0.1181
	0.1261	0.0803	0.1559	0.1480	0.1632	0.1090	0.1261	0.1259	0.1141	0.1185
	0.1263	0.0807	0.1568	0.1490	0.1638	0.1098	0.1263	0.1270	0.1163	0.1185
	0.1270	0.0812	0.1572	0.1495	0.1639	0.1106	0.1270	0.1273	0.1163	0.1185
	0.1280	0.0832	0.1574	0.1497	0.1642	0.1110	0.1280	0.1275	0.1179	0.1190
	0.1282	0.0838	0.1602	0.1498	0.1653	0.1111	0.1282	0.1278	0.1180	0.1190
	0.1296	0.0863	0.1604	0.1524	0.1657	0.1128	0.1296	0.1281	0.1185	0.1193
	0.1305	0.0872	0.1611	0.1529	0.1660	0.1137	0.1305	0.1282	0.1186	0.1202
	0.1312	0.0874	0.1614	0.1537	0.1660	0.1153	0.1312	0.1316	0.1188	0.1211
	0.1312	0.0875	0.1619	0.1576	0.1689	0.1175	0.1312	0.1317	0.1190	0.1213
	0.1314	0.0929	0.1637	0.1578	0.1707	0.1180	0.1314	0.1324	0.1243	0.1239
	0.1369	0.1016	0.1644	0.1625	0.1715	0.1213	0.1369	0.1334	0.1276	0.1263
	0.1401	0.1029	0.1672	0.1640	0.1756	0.1233	0.1401	0.1359	0.1278	0.1290
	0.1402	0.1211	0.1741	0.1702	0.1837	0.1245	0.1402	0.1464	0.1298	0.1305
r'	**4**	**1**	**4**	**1**	**8**	**2**	**3**	**1**	**2**	**5**

TABLE A.5: Ranked *sim* values for $(a \vee \{b, c\})$ in ACB dataset using U-KTS $\mathcal{KH}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group b and c	0.0115	0.1910	0.0129	0.0062	0.0125	0.0060	0.0174	0.0090	0.0110	0.0132
	0.0118	0.1953	0.0129	0.0067	0.0134	0.0115	0.0185	0.0090	0.0122	0.0135
	0.0136	0.2072	0.0129	0.0078	0.0135	0.0122	0.0221	0.0095	0.0126	0.0136
	0.0138	0.2146	0.0130	0.0079	0.0137	0.0123	0.0225	0.0097	0.0130	0.0137
	0.0138	0.2245	0.0133	0.0080	0.0138	0.0132	0.0226	0.0102	0.0130	0.0140
	0.0139	0.2259	0.0135	0.0084	0.0141	0.0141	0.0230	0.0103	0.0133	0.0149
	0.0148	0.2421	0.0136	0.0094	0.0141	0.0142	0.0238	0.0103	0.0135	0.0153
	0.0148	0.2451	0.0138	0.0095	0.0143	0.0143	0.0241	0.0104	0.0135	0.0154
	0.0154	0.2458	0.0139	0.0103	0.0146	0.0156	0.0247	0.0112	0.0141	0.0155
	0.0154	0.2459	0.0141	0.0105	0.0146	0.0173	0.0250	0.0114	0.0145	0.0157
	0.0154	0.2464	0.0142	0.0106	0.0148	0.0174	0.0250	0.0115	0.0147	0.0157
	0.0158	0.2478	0.0145	0.0110	0.0151	0.0185	0.0251	0.0115	0.0155	0.0166
	0.0158	0.2479	0.0145	0.0113	0.0151	0.0190	0.0252	0.0116	0.0156	0.0172
	0.0159	0.2481	0.0146	0.0114	0.0153	0.0194	0.0257	0.0119	0.0164	0.0174
	0.0161	0.2481	0.0147	0.0114	0.0157	0.0194	0.0258	0.0122	0.0167	0.0174
	0.0164	0.2482	0.0148	0.0120	0.0157	0.0194	0.0260	0.0127	0.0173	0.0175
	0.0166	0.2485	0.0149	0.0126	0.0158	0.0194	0.0267	0.0129	0.0173	0.0177
	0.0166	0.2499	0.0149	0.0138	0.0160	0.0198	0.0268	0.0131	0.0179	0.0180
	0.0177	0.2506	0.0151	0.0149	0.0160	0.0209	0.0274	0.0138	0.0185	0.0188
	0.0178	0.2519	0.0152	0.0150	0.0162	0.0212	0.0293	0.0140	0.0185	0.0189
	0.0183	0.2531	0.0158	0.0152	0.0163	0.0213	0.0293	0.0140	0.0188	0.0190
	0.0189	0.2545	0.0163	0.0152	0.0169	0.0219	0.0295	0.0145	0.0190	0.0196
	0.0199	0.2546	0.0165	0.0155	0.0177	0.0226	0.0296	0.0148	0.0199	0.0198
	0.0206	0.2552	0.0204	0.0157	0.0192	0.0226	0.0299	0.0157	0.0235	0.0198
	0.0210	0.2555	0.0206	0.0171	0.0195	0.0242	0.0303	0.0187	0.0265	0.0202
	0.0211	0.2559	0.0208	0.0181	0.0209	0.0291	0.0309	0.0188	0.0266	0.0206
	0.0214	0.2573	0.0217	0.0188	0.0234	0.0295	0.0313	0.0188	0.0290	0.0209
	0.0255	0.2592	0.0249	0.0195	0.0243	0.0383	0.0323	0.0222	0.0385	0.0259
	0.0302	0.2598	0.0259	0.0218	0.0266	0.0439	0.0389	0.0227	0.0391	0.0271
	0.4421	0.2653	0.4554	0.4548	0.4579	0.4771	0.4265	0.4541	0.4757	0.4397
r'	**1**	**2**	**4**	**2**	**3**	**5**	**10**	**5**	**2**	**1**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group b and c	0.0107	0.0125	0.0074	0.0062	0.0084	0.0074	0.0085	0.0066	0.0063	0.0134
	0.0112	0.0127	0.0076	0.0065	0.0086	0.0077	0.0091	0.0079	0.0068	0.0137
	0.0118	0.0132	0.0082	0.0067	0.0087	0.0081	0.0092	0.0085	0.0070	0.0139
	0.0119	0.0133	0.0086	0.0069	0.0092	0.0083	0.0097	0.0093	0.0073	0.0141
	0.0119	0.0133	0.0094	0.0075	0.0093	0.0087	0.0100	0.0094	0.0077	0.0144
	0.0128	0.0133	0.0095	0.0081	0.0097	0.0092	0.0109	0.0103	0.0081	0.0149
	0.0129	0.0137	0.0098	0.0081	0.0098	0.0097	0.0111	0.0106	0.0082	0.0154
	0.0132	0.0140	0.0100	0.0083	0.0109	0.0102	0.0117	0.0106	0.0082	0.0156
	0.0132	0.0143	0.0103	0.0083	0.0109	0.0102	0.0117	0.0106	0.0099	0.0158
	0.0133	0.0145	0.0105	0.0107	0.0111	0.0105	0.0117	0.0106	0.0101	0.0159
	0.0134	0.0146	0.0110	0.0111	0.0111	0.0108	0.0118	0.0107	0.0111	0.0159
	0.0135	0.0151	0.0121	0.0117	0.0114	0.0108	0.0121	0.0109	0.0118	0.0160
	0.0136	0.0151	0.0123	0.0117	0.0137	0.0114	0.0123	0.0112	0.0118	0.0163
	0.0140	0.0155	0.0139	0.0123	0.0138	0.0115	0.0129	0.0115	0.0120	0.0166
	0.0140	0.0157	0.0147	0.0124	0.0140	0.0123	0.0135	0.0121	0.0122	0.0168
	0.0144	0.0158	0.0147	0.0124	0.0140	0.0135	0.0136	0.0123	0.0125	0.0169
	0.0150	0.0160	0.0157	0.0138	0.0140	0.0138	0.0142	0.0134	0.0126	0.0173
	0.0152	0.0162	0.0159	0.0140	0.0153	0.0139	0.0144	0.0136	0.0136	0.0179
	0.0152	0.0168	0.0170	0.0144	0.0163	0.0139	0.0153	0.0142	0.0140	0.0182
	0.0153	0.0170	0.0170	0.0145	0.0170	0.0140	0.0156	0.0144	0.0150	0.0189
	0.0153	0.0170	0.0176	0.0150	0.0174	0.0142	0.0160	0.0145	0.0156	0.0189
	0.0157	0.0179	0.0176	0.0161	0.0176	0.0145	0.0167	0.0149	0.0156	0.0190
	0.0172	0.0179	0.0178	0.0165	0.0177	0.0150	0.0170	0.0155	0.0163	0.0193
	0.0199	0.0197	0.0179	0.0165	0.0180	0.0165	0.0181	0.0155	0.0163	0.0193
	0.0200	0.0228	0.0185	0.0167	0.0184	0.0186	0.0208	0.0166	0.0170	0.0196
	0.0205	0.0238	0.0211	0.0190	0.0216	0.0187	0.0238	0.0170	0.0187	0.0207
	0.0207	0.0250	0.0238	0.0206	0.0216	0.0195	0.0249	0.0174	0.0193	0.0210
	0.0209	0.0280	0.0328	0.0237	0.0304	0.0257	0.0322	0.0198	0.0262	0.0250
	0.0250	0.0304	0.0332	0.0279	0.0321	0.0274	0.0326	0.0218	0.0276	0.0252
	0.4529	0.4617	0.4692	0.4619	0.4686	0.4613	0.4695	0.4538	0.4627	0.4986
r'	**5**	**4**	**13**	**2**	**2**	**3**	**5**	**1**	**6**	**7**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group b and c	0.0105	0.0060	0.0125	0.0074	0.3251	0.0090	0.0105	0.0087	0.0119	0.0140
	0.0108	0.0064	0.0126	0.0088	0.4313	0.0099	0.0108	0.0093	0.0130	0.0142
	0.0121	0.0068	0.0127	0.0090	0.4317	0.0111	0.0121	0.0094	0.0135	0.0145
	0.0123	0.0071	0.0131	0.0093	0.4419	0.0119	0.0123	0.0095	0.0137	0.0152
	0.0124	0.0075	0.0131	0.0095	0.4431	0.0121	0.0124	0.0095	0.0148	0.0156
	0.0128	0.0076	0.0134	0.0097	0.4492	0.0121	0.0128	0.0095	0.0151	0.0156
	0.0134	0.0078	0.0134	0.0101	0.4522	0.0127	0.0134	0.0096	0.0152	0.0162
	0.0134	0.0081	0.0135	0.0103	0.4530	0.0131	0.0134	0.0097	0.0153	0.0164
	0.0135	0.0086	0.0136	0.0107	0.4531	0.0132	0.0135	0.0101	0.0158	0.0165
	0.0137	0.0100	0.0138	0.0109	0.4536	0.0132	0.0137	0.0119	0.0160	0.0169
	0.0139	0.0100	0.0156	0.0112	0.4549	0.0134	0.0139	0.0123	0.0172	0.0180
	0.0139	0.0105	0.0157	0.0113	0.4550	0.0135	0.0139	0.0123	0.0182	0.0185
	0.0139	0.0105	0.0160	0.0118	0.4552	0.0140	0.0139	0.0123	0.0183	0.0186
	0.0140	0.0113	0.0163	0.0128	0.4553	0.0141	0.0140	0.0126	0.0186	0.0189
	0.0145	0.0125	0.0163	0.0137	0.4553	0.0142	0.0145	0.0126	0.0188	0.0190
	0.0149	0.0128	0.0164	0.0141	0.4554	0.0148	0.0149	0.0126	0.0193	0.0193
	0.0150	0.0131	0.0164	0.0141	0.4556	0.0152	0.0150	0.0127	0.0193	0.0193
	0.0151	0.0131	0.0165	0.0147	0.4570	0.0152	0.0151	0.0133	0.0196	0.0198
	0.0156	0.0135	0.0171	0.0155	0.4578	0.0155	0.0156	0.0135	0.0200	0.0200
	0.0158	0.0148	0.0172	0.0159	0.4590	0.0156	0.0158	0.0138	0.0203	0.0204
	0.0159	0.0151	0.0172	0.0162	0.4603	0.0158	0.0159	0.0143	0.0205	0.0206
	0.0160	0.0157	0.0173	0.0162	0.4617	0.0162	0.0160	0.0148	0.0220	0.0211
	0.0162	0.0158	0.0175	0.0163	0.4618	0.0172	0.0162	0.0154	0.0221	0.0212
	0.0168	0.0160	0.0198	0.0173	0.4624	0.0177	0.0168	0.0156	0.0234	0.0221
	0.0171	0.0167	0.0200	0.0174	0.4627	0.0186	0.0171	0.0181	0.0269	0.0222
	0.0186	0.0185	0.0218	0.0211	0.4630	0.0187	0.0186	0.0189	0.0317	0.0245
	0.0188	0.0198	0.0226	0.0242	0.4644	0.0213	0.0188	0.0190	0.0334	0.0268
	0.0216	0.0247	0.0239	0.0303	0.4664	0.0222	0.0216	0.0191	0.0432	0.0270
	0.0232	0.0265	0.0245	0.0312	0.4669	0.0241	0.0232	0.0236	0.0435	0.0299
	0.4473	0.4621	0.4532	0.4682	0.4725	0.4443	0.4473	0.4586	0.4797	0.4458
r'	**5**	**6**	**2**	**2**	**2**	**3**	**1**	**1**	**2**	**1**

TABLE A.6: Ranked *sim* values for $(b \vee \{a, c\})$ in ACB dataset using U-KTS $\mathcal{KH}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and c	0.0180	0.0192	0.0170	0.0094	0.0144	0.0131	0.0265	0.0176	0.0162	0.0218
	0.0180	0.0200	0.0171	0.0107	0.0144	0.0136	0.0296	0.0177	0.0169	0.0221
	0.0189	0.0200	0.0172	0.0108	0.0145	0.0137	0.0304	0.0181	0.0172	0.0230
	0.0189	0.0206	0.0172	0.0112	0.0146	0.0144	0.0309	0.0182	0.0178	0.0231
	0.0191	0.0207	0.0174	0.0114	0.0146	0.0154	0.0318	0.0183	0.0179	0.0231
	0.0194	0.0209	0.0175	0.0117	0.0147	0.0155	0.0318	0.0184	0.0180	0.0231
	0.0196	0.0212	0.0175	0.0119	0.0151	0.0160	0.0323	0.0184	0.0181	0.0232
	0.0197	0.0213	0.0175	0.0120	0.0153	0.0163	0.0327	0.0185	0.0183	0.0236
	0.0199	0.0214	0.0175	0.0121	0.0154	0.0166	0.0328	0.0186	0.0185	0.0238
	0.0201	0.0214	0.0177	0.0124	0.0156	0.0168	0.0329	0.0187	0.0187	0.0239
	0.0204	0.0216	0.0179	0.0127	0.0157	0.0171	0.0331	0.0187	0.0187	0.0245
	0.0204	0.0217	0.0180	0.0127	0.0158	0.0172	0.0333	0.0189	0.0187	0.0246
	0.0206	0.0217	0.0181	0.0127	0.0158	0.0173	0.0333	0.0190	0.0188	0.0247
	0.0212	0.0217	0.0181	0.0129	0.0158	0.0173	0.0342	0.0191	0.0190	0.0248
	0.0213	0.0220	0.0184	0.0134	0.0159	0.0174	0.0348	0.0191	0.0192	0.0252
	0.0213	0.0220	0.0188	0.0134	0.0160	0.0175	0.0354	0.0195	0.0193	0.0254
	0.0224	0.0226	0.0189	0.0134	0.0161	0.0177	0.0362	0.0199	0.0196	0.0254
	0.0229	0.0226	0.0190	0.0136	0.0166	0.0183	0.0368	0.0204	0.0196	0.0260
	0.0230	0.0229	0.0191	0.0136	0.0166	0.0186	0.0377	0.0204	0.0199	0.0262
	0.0233	0.0230	0.0191	0.0137	0.0167	0.0188	0.0385	0.0208	0.0217	0.0267
	0.0233	0.0230	0.0192	0.0139	0.0169	0.0188	0.0395	0.0210	0.0217	0.0276
	0.0246	0.0235	0.0199	0.0140	0.0171	0.0188	0.0396	0.0213	0.0219	0.0279
	0.0254	0.0235	0.0201	0.0145	0.0171	0.0196	0.0398	0.0214	0.0219	0.0282
	0.0258	0.0240	0.0207	0.0147	0.0173	0.0204	0.0406	0.0216	0.0223	0.0288
	0.0279	0.0242	0.0208	0.0149	0.0180	0.0218	0.0426	0.0221	0.0230	0.0291
	0.0283	0.0269	0.0215	0.0154	0.0192	0.0223	0.0439	0.0222	0.0230	0.0298
	0.0305	0.0273	0.0232	0.0184	0.0203	0.0244	0.0457	0.0222	0.0236	0.0310
	0.0366	0.0298	0.0256	0.0188	0.0220	0.0280	0.0498	0.0250	0.0282	0.0393
	0.2072	0.2128	0.2286	0.2282	0.2295	0.2538	0.1857	0.2337	0.2462	0.1998
	0.9055	0.9171	0.9274	0.9269	0.9282	0.9522	0.8886	0.9354	0.9449	0.9054
r'	**1**	**3**	**3**	**4**	**2**	**4**	**8**	**7**	**1**	**2**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and c	0.0197	0.0252	0.0118	0.0112	0.0131	0.0124	0.0177	0.0112	0.0106	0.0215
	0.0203	0.0258	0.0122	0.0119	0.0135	0.0130	0.0180	0.0116	0.0109	0.0223
	0.0204	0.0258	0.0126	0.0120	0.0135	0.0131	0.0181	0.0118	0.0110	0.0224
	0.0207	0.0261	0.0126	0.0123	0.0137	0.0132	0.0182	0.0118	0.0115	0.0228
	0.0209	0.0266	0.0127	0.0128	0.0138	0.0138	0.0187	0.0123	0.0116	0.0228
	0.0213	0.0268	0.0128	0.0129	0.0139	0.0138	0.0189	0.0123	0.0118	0.0229
	0.0214	0.0280	0.0133	0.0130	0.0142	0.0139	0.0189	0.0126	0.0120	0.0230
	0.0214	0.0281	0.0141	0.0132	0.0150	0.0141	0.0189	0.0128	0.0122	0.0230
	0.0214	0.0287	0.0141	0.0135	0.0152	0.0143	0.0189	0.0129	0.0123	0.0231
	0.0217	0.0288	0.0143	0.0135	0.0153	0.0145	0.0189	0.0130	0.0125	0.0231
	0.0219	0.0288	0.0144	0.0137	0.0154	0.0145	0.0191	0.0132	0.0125	0.0232
	0.0220	0.0290	0.0150	0.0138	0.0157	0.0146	0.0191	0.0133	0.0127	0.0232
	0.0220	0.0290	0.0159	0.0141	0.0160	0.0146	0.0194	0.0134	0.0141	0.0235
	0.0221	0.0294	0.0159	0.0149	0.0165	0.0147	0.0195	0.0135	0.0144	0.0237
	0.0221	0.0294	0.0160	0.0151	0.0166	0.0148	0.0195	0.0141	0.0144	0.0240
	0.0233	0.0295	0.0161	0.0152	0.0170	0.0150	0.0197	0.0142	0.0145	0.0246
	0.0233	0.0298	0.0165	0.0152	0.0170	0.0150	0.0204	0.0143	0.0145	0.0246
	0.0234	0.0299	0.0171	0.0153	0.0175	0.0150	0.0207	0.0145	0.0146	0.0249
	0.0238	0.0302	0.0173	0.0156	0.0179	0.0152	0.0207	0.0145	0.0149	0.0251
	0.0240	0.0308	0.0173	0.0158	0.0185	0.0152	0.0208	0.0151	0.0150	0.0252
	0.0240	0.0309	0.0175	0.0162	0.0185	0.0152	0.0211	0.0156	0.0150	0.0259
	0.0240	0.0310	0.0175	0.0165	0.0186	0.0153	0.0211	0.0162	0.0155	0.0260
	0.0241	0.0310	0.0183	0.0167	0.0187	0.0158	0.0216	0.0162	0.0156	0.0263
	0.0245	0.0311	0.0188	0.0167	0.0188	0.0160	0.0231	0.0162	0.0160	0.0265
	0.0255	0.0313	0.0203	0.0172	0.0202	0.0165	0.0233	0.0163	0.0165	0.0276
	0.0259	0.0315	0.0204	0.0175	0.0217	0.0166	0.0241	0.0165	0.0167	0.0278
	0.0266	0.0337	0.0212	0.0176	0.0230	0.0195	0.0241	0.0171	0.0181	0.0288
	0.0323	0.0361	0.0240	0.0205	0.0254	0.0208	0.0264	0.0189	0.0205	0.0359
	0.2226	0.2249	0.2516	0.2382	0.2487	0.2311	0.2419	0.2357	0.2431	0.2156
	0.9213	0.9284	0.9503	0.9369	0.9475	0.9298	0.9442	0.9344	0.9419	0.9144
r'	**5**	**6**	**10**	**5**	**2**	**4**	**4**	**2**	**5**	**6**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and c	0.0170	0.0137	0.0200	0.0111	0.1846	0.0176	0.0170	0.0163	0.0215	0.0190
	0.0178	0.0138	0.0202	0.0115	0.2283	0.0180	0.0178	0.0168	0.0225	0.0196
	0.0181	0.0139	0.0202	0.0115	0.3049	0.0180	0.0181	0.0170	0.0228	0.0196
	0.0182	0.0143	0.0203	0.0115	0.3165	0.0187	0.0182	0.0172	0.0229	0.0199
	0.0183	0.0145	0.0204	0.0117	0.3187	0.0191	0.0183	0.0174	0.0232	0.0200
	0.0183	0.0145	0.0204	0.0118	0.3206	0.0191	0.0183	0.0174	0.0232	0.0201
	0.0184	0.0146	0.0205	0.0124	0.3208	0.0204	0.0184	0.0175	0.0232	0.0203
	0.0184	0.0149	0.0206	0.0129	0.3211	0.0207	0.0184	0.0175	0.0233	0.0203
	0.0186	0.0149	0.0208	0.0130	0.3246	0.0209	0.0186	0.0175	0.0236	0.0204
	0.0188	0.0150	0.0208	0.0132	0.3263	0.0209	0.0188	0.0176	0.0236	0.0206
	0.0189	0.0154	0.0208	0.0134	0.3263	0.0211	0.0189	0.0177	0.0237	0.0210
	0.0189	0.0157	0.0209	0.0137	0.3265	0.0218	0.0189	0.0178	0.0238	0.0211
	0.0190	0.0162	0.0209	0.0139	0.3294	0.0218	0.0190	0.0180	0.0238	0.0214
	0.0192	0.0169	0.0209	0.0144	0.3302	0.0220	0.0192	0.0181	0.0239	0.0215
	0.0193	0.0170	0.0212	0.0146	0.3318	0.0220	0.0193	0.0183	0.0240	0.0215
	0.0194	0.0172	0.0214	0.0146	0.3334	0.0221	0.0194	0.0183	0.0241	0.0217
	0.0196	0.0172	0.0215	0.0147	0.3353	0.0238	0.0196	0.0185	0.0242	0.0218
	0.0199	0.0173	0.0216	0.0155	0.3385	0.0239	0.0199	0.0189	0.0246	0.0222
	0.0202	0.0176	0.0217	0.0155	0.3406	0.0241	0.0202	0.0191	0.0251	0.0224
	0.0208	0.0176	0.0223	0.0157	0.3447	0.0242	0.0208	0.0198	0.0253	0.0231
	0.0208	0.0176	0.0226	0.0158	0.3476	0.0247	0.0208	0.0200	0.0254	0.0233
	0.0210	0.0177	0.0227	0.0160	0.3481	0.0262	0.0210	0.0202	0.0259	0.0234
	0.0212	0.0177	0.0230	0.0162	0.3491	0.0269	0.0212	0.0202	0.0266	0.0240
	0.0221	0.0185	0.0233	0.0169	0.3496	0.0270	0.0221	0.0205	0.0267	0.0241
	0.0222	0.0189	0.0233	0.0182	0.3500	0.0287	0.0222	0.0214	0.0275	0.0254
	0.0224	0.0190	0.0235	0.0184	0.3523	0.0301	0.0224	0.0218	0.0277	0.0256
	0.0234	0.0192	0.0261	0.0187	0.3551	0.0321	0.0234	0.0225	0.0287	0.0259
	0.0296	0.0234	0.0270	0.0220	0.3556	0.0425	0.0296	0.0268	0.0332	0.0319
	0.2225	0.2421	0.2310	0.2480	0.3558	0.1930	0.2225	0.2244	0.2241	0.2226
	0.9213	0.9408	0.9297	0.9467	0.5738	0.8970	0.9213	0.9232	0.9228	0.9214
r'	**2**	**5**	**1**	**1**	**1**	**1**	**2**	**1**	**3**	**1**

TABLE A.7: Ranked *sim* values for $(c \vee \{a, b\})$ in ACB dataset using U-KTS $\mathcal{KH}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and b	0.0188	0.0192	0.0164	0.0102	0.0135	0.0121	0.0282	0.0181	0.0172	0.0227
	0.0191	0.0199	0.0167	0.0108	0.0141	0.0133	0.0301	0.0182	0.0175	0.0231
	0.0191	0.0203	0.0170	0.0109	0.0141	0.0141	0.0313	0.0183	0.0177	0.0232
	0.0193	0.0204	0.0170	0.0109	0.0141	0.0145	0.0318	0.0184	0.0178	0.0232
	0.0199	0.0207	0.0170	0.0113	0.0144	0.0149	0.0320	0.0184	0.0179	0.0234
	0.0199	0.0208	0.0170	0.0113	0.0144	0.0150	0.0327	0.0184	0.0182	0.0235
	0.0203	0.0209	0.0173	0.0114	0.0148	0.0153	0.0332	0.0184	0.0182	0.0237
	0.0205	0.0209	0.0175	0.0114	0.0149	0.0158	0.0332	0.0185	0.0182	0.0238
	0.0205	0.0209	0.0176	0.0115	0.0151	0.0162	0.0335	0.0186	0.0182	0.0238
	0.0206	0.0210	0.0177	0.0117	0.0152	0.0163	0.0335	0.0186	0.0183	0.0239
	0.0207	0.0210	0.0177	0.0117	0.0152	0.0168	0.0336	0.0186	0.0183	0.0240
	0.0210	0.0211	0.0177	0.0120	0.0153	0.0169	0.0337	0.0188	0.0186	0.0246
	0.0215	0.0214	0.0178	0.0125	0.0154	0.0169	0.0341	0.0189	0.0188	0.0248
	0.0217	0.0214	0.0178	0.0126	0.0154	0.0170	0.0349	0.0190	0.0189	0.0249
	0.0217	0.0214	0.0179	0.0129	0.0155	0.0171	0.0349	0.0192	0.0190	0.0250
	0.0217	0.0217	0.0179	0.0131	0.0156	0.0181	0.0351	0.0193	0.0190	0.0251
	0.0217	0.0218	0.0179	0.0131	0.0156	0.0182	0.0352	0.0193	0.0191	0.0260
	0.0221	0.0219	0.0180	0.0132	0.0157	0.0182	0.0356	0.0194	0.0192	0.0262
	0.0221	0.0219	0.0182	0.0133	0.0158	0.0187	0.0360	0.0194	0.0196	0.0263
	0.0226	0.0219	0.0182	0.0134	0.0159	0.0189	0.0363	0.0195	0.0197	0.0266
	0.0230	0.0224	0.0183	0.0134	0.0164	0.0189	0.0374	0.0197	0.0200	0.0268
	0.0231	0.0227	0.0189	0.0140	0.0171	0.0196	0.0376	0.0199	0.0209	0.0269
	0.0232	0.0227	0.0195	0.0146	0.0175	0.0199	0.0376	0.0200	0.0212	0.0269
	0.0238	0.0229	0.0201	0.0148	0.0179	0.0200	0.0377	0.0202	0.0213	0.0270
	0.0239	0.0229	0.0213	0.0155	0.0183	0.0204	0.0381	0.0206	0.0216	0.0272
	0.0239	0.0236	0.0219	0.0159	0.0195	0.0213	0.0385	0.0224	0.0230	0.0273
	0.0240	0.0248	0.0222	0.0161	0.0196	0.0226	0.0387	0.0227	0.0237	0.0273
	0.0265	0.0253	0.0224	0.0168	0.0201	0.0229	0.0388	0.0238	0.0273	0.0279
	0.0315	0.0265	0.0227	0.0177	0.0201	0.0376	0.0430	0.0256	0.0297	0.0351
	0.9055	0.9171	0.9274	0.9269	0.9282	0.9522	0.8886	0.9354	0.9449	0.9054
r'	**2**	**2**	**1**	**3**	**3**	**6**	**8**	**6**	**1**	**1**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and b	0.0200	0.0301	0.0127	0.0116	0.0131	0.0129	0.0174	0.0108	0.0112	0.0218
	0.0200	0.0301	0.0132	0.0119	0.0135	0.0130	0.0178	0.0115	0.0112	0.0219
	0.0201	0.0302	0.0133	0.0120	0.0139	0.0132	0.0178	0.0117	0.0114	0.0223
	0.0203	0.0304	0.0135	0.0122	0.0143	0.0132	0.0182	0.0120	0.0114	0.0228
	0.0206	0.0306	0.0137	0.0125	0.0144	0.0132	0.0186	0.0120	0.0116	0.0228
	0.0207	0.0309	0.0137	0.0128	0.0145	0.0133	0.0188	0.0120	0.0116	0.0230
	0.0209	0.0310	0.0138	0.0129	0.0145	0.0136	0.0189	0.0121	0.0121	0.0230
	0.0212	0.0311	0.0147	0.0133	0.0146	0.0136	0.0189	0.0123	0.0122	0.0232
	0.0213	0.0312	0.0147	0.0134	0.0155	0.0138	0.0189	0.0124	0.0124	0.0233
	0.0215	0.0313	0.0150	0.0135	0.0157	0.0139	0.0190	0.0124	0.0125	0.0233
	0.0217	0.0314	0.0151	0.0135	0.0157	0.0140	0.0192	0.0126	0.0126	0.0234
	0.0222	0.0318	0.0152	0.0138	0.0158	0.0141	0.0192	0.0126	0.0128	0.0236
	0.0225	0.0320	0.0156	0.0142	0.0160	0.0143	0.0194	0.0127	0.0129	0.0237
	0.0225	0.0322	0.0157	0.0142	0.0160	0.0144	0.0194	0.0129	0.0131	0.0239
	0.0225	0.0324	0.0160	0.0143	0.0160	0.0144	0.0198	0.0136	0.0131	0.0240
	0.0226	0.0326	0.0162	0.0143	0.0161	0.0144	0.0198	0.0137	0.0133	0.0241
	0.0226	0.0327	0.0162	0.0147	0.0162	0.0145	0.0198	0.0138	0.0133	0.0244
	0.0227	0.0327	0.0164	0.0149	0.0174	0.0145	0.0198	0.0139	0.0137	0.0245
	0.0228	0.0329	0.0164	0.0151	0.0177	0.0146	0.0200	0.0140	0.0138	0.0247
	0.0229	0.0331	0.0164	0.0153	0.0177	0.0147	0.0202	0.0145	0.0141	0.0248
	0.0230	0.0331	0.0168	0.0154	0.0184	0.0153	0.0207	0.0145	0.0144	0.0250
	0.0233	0.0332	0.0174	0.0159	0.0184	0.0154	0.0210	0.0146	0.0144	0.0252
	0.0238	0.0334	0.0175	0.0159	0.0185	0.0158	0.0214	0.0149	0.0149	0.0253
	0.0249	0.0336	0.0176	0.0160	0.0186	0.0167	0.0216	0.0152	0.0151	0.0253
	0.0254	0.0341	0.0179	0.0167	0.0187	0.0175	0.0217	0.0157	0.0165	0.0255
	0.0259	0.0354	0.0197	0.0176	0.0191	0.0175	0.0217	0.0158	0.0167	0.0257
	0.0260	0.0363	0.0200	0.0181	0.0205	0.0183	0.0257	0.0159	0.0169	0.0263
	0.0270	0.0368	0.0235	0.0200	0.0243	0.0195	0.0270	0.0181	0.0194	0.0288
	0.0286	0.0381	0.0263	0.0212	0.0296	0.0199	0.0286	0.0191	0.0205	0.0306
	0.9213	0.9229	0.9503	0.9369	0.9475	0.9298	0.9442	0.9344	0.9419	0.9144
r'	**6**	**7**	**12**	**1**	**1**	**4**	**6**	**3**	**7**	**7**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and b	0.0177	0.0140	0.0195	0.0117	0.1774	0.0166	0.0177	0.0161	0.0219	0.0194
	0.0178	0.0140	0.0199	0.0117	0.1872	0.0188	0.0178	0.0167	0.0225	0.0194
	0.0180	0.0143	0.0199	0.0119	0.2301	0.0199	0.0180	0.0168	0.0228	0.0194
	0.0181	0.0144	0.0199	0.0121	0.2508	0.0203	0.0181	0.0169	0.0228	0.0197
	0.0183	0.0145	0.0203	0.0121	0.2950	0.0204	0.0183	0.0170	0.0231	0.0202
	0.0183	0.0145	0.0204	0.0122	0.2962	0.0206	0.0183	0.0170	0.0232	0.0202
	0.0184	0.0145	0.0204	0.0127	0.3188	0.0206	0.0184	0.0171	0.0232	0.0203
	0.0185	0.0146	0.0204	0.0128	0.3241	0.0207	0.0185	0.0171	0.0232	0.0205
	0.0186	0.0147	0.0205	0.0130	0.3268	0.0207	0.0186	0.0172	0.0233	0.0207
	0.0186	0.0151	0.0205	0.0132	0.3291	0.0208	0.0186	0.0172	0.0236	0.0208
	0.0189	0.0151	0.0208	0.0133	0.3295	0.0213	0.0189	0.0173	0.0236	0.0209
	0.0189	0.0152	0.0209	0.0134	0.3295	0.0213	0.0189	0.0175	0.0236	0.0210
	0.0189	0.0156	0.0210	0.0135	0.3302	0.0221	0.0189	0.0177	0.0236	0.0210
	0.0191	0.0157	0.0210	0.0135	0.3328	0.0222	0.0191	0.0177	0.0242	0.0211
	0.0195	0.0159	0.0211	0.0141	0.3332	0.0224	0.0195	0.0178	0.0242	0.0212
	0.0195	0.0159	0.0212	0.0142	0.3339	0.0224	0.0195	0.0180	0.0243	0.0214
	0.0196	0.0163	0.0212	0.0145	0.3341	0.0227	0.0196	0.0182	0.0243	0.0215
	0.0198	0.0164	0.0212	0.0145	0.3365	0.0227	0.0198	0.0185	0.0243	0.0215
	0.0199	0.0165	0.0213	0.0146	0.3378	0.0228	0.0199	0.0186	0.0245	0.0219
	0.0201	0.0165	0.0218	0.0154	0.3413	0.0237	0.0201	0.0186	0.0246	0.0222
	0.0202	0.0170	0.0219	0.0154	0.3416	0.0240	0.0202	0.0186	0.0249	0.0223
	0.0202	0.0172	0.0219	0.0156	0.3421	0.0243	0.0202	0.0194	0.0249	0.0225
	0.0204	0.0178	0.0220	0.0160	0.3421	0.0247	0.0204	0.0197	0.0254	0.0226
	0.0213	0.0182	0.0229	0.0161	0.3422	0.0251	0.0213	0.0200	0.0258	0.0230
	0.0217	0.0184	0.0246	0.0163	0.3437	0.0257	0.0217	0.0201	0.0260	0.0231
	0.0221	0.0199	0.0247	0.0177	0.3452	0.0258	0.0221	0.0206	0.0266	0.0234
	0.0224	0.0201	0.0250	0.0185	0.3461	0.0261	0.0224	0.0208	0.0273	0.0249
	0.0240	0.0228	0.0253	0.0223	0.3493	0.0285	0.0240	0.0229	0.0288	0.0271
	0.0258	0.0229	0.0253	0.0241	0.3557	0.0329	0.0258	0.0235	0.0298	0.0284
	0.9213	0.9408	0.9297	0.9467	0.5738	0.8970	0.9213	0.9232	0.9228	0.9214
r'	**4**	**4**	**1**	**5**	**3**	**1**	**1**	**1**	**2**	**2**

TABLE A.8: Ranked *sim* values for $(a \vee \{b, c\})$ in ACB dataset using M-KTS.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group b and c	0.0715	0.0580	0.0761	0.0809	0.1070	0.0403	0.1232	0.0564	0.0641	0.1059
	0.0754	0.0644	0.0773	0.0898	0.1088	0.0592	0.1291	0.0662	0.0664	0.1059
	0.0770	0.0717	0.0800	0.0921	0.1128	0.0649	0.1300	0.0719	0.0678	0.1148
	0.0775	0.0753	0.0808	0.0926	0.1171	0.0650	0.1315	0.0741	0.0726	0.1154
	0.0789	0.0761	0.0810	0.0928	0.1171	0.0658	0.1354	0.0744	0.0744	0.1182
	0.0808	0.0791	0.0817	0.0948	0.1179	0.0716	0.1402	0.0772	0.0746	0.1210
	0.0812	0.0835	0.0825	0.0983	0.1197	0.0732	0.1404	0.0793	0.0766	0.1267
	0.0839	0.0845	0.0825	0.0996	0.1241	0.0737	0.1477	0.0804	0.0769	0.1298
	0.0839	0.0858	0.0829	0.1001	0.1267	0.0760	0.1477	0.0806	0.0779	0.1300
	0.0880	0.0912	0.0839	0.1004	0.1303	0.0795	0.1486	0.0826	0.0779	0.1329
	0.0886	0.0916	0.0841	0.1007	0.1309	0.0815	0.1548	0.0847	0.0801	0.1333
	0.0934	0.0922	0.0842	0.1024	0.1310	0.0870	0.1549	0.0848	0.0804	0.1337
	0.0946	0.0976	0.0845	0.1026	0.1352	0.0898	0.1587	0.0864	0.0831	0.1341
	0.0971	0.0988	0.0849	0.1027	0.1377	0.0933	0.1614	0.0883	0.0863	0.1357
	0.0971	0.1082	0.0871	0.1035	0.1388	0.0979	0.1652	0.0915	0.0906	0.1380
	0.0971	0.1139	0.0902	0.1035	0.1393	0.1042	0.1672	0.0962	0.0913	0.1382
	0.0975	0.1166	0.0943	0.1037	0.1394	0.1049	0.1736	0.0962	0.0917	0.1384
	0.1014	0.1181	0.0943	0.1050	0.1432	0.1072	0.1746	0.0963	0.0936	0.1398
	0.1030	0.1192	0.0956	0.1072	0.1440	0.1072	0.1753	0.0969	0.0936	0.1403
	0.1031	0.1192	0.0972	0.1118	0.1470	0.1079	0.1763	0.0982	0.0938	0.1415
	0.1050	0.1232	0.0974	0.1140	0.1481	0.1102	0.1763	0.0990	0.0947	0.1448
	0.1057	0.1300	0.0977	0.1147	0.1499	0.1144	0.1791	0.1028	0.0973	0.1452
	0.1066	0.1310	0.0984	0.1158	0.1513	0.1156	0.1948	0.1125	0.0989	0.1467
	0.1107	0.1316	0.0990	0.1182	0.1553	0.1163	0.1954	0.1125	0.1019	0.1481
	0.1221	0.1348	0.1057	0.1194	0.1566	0.1207	0.2003	0.1186	0.1108	0.1484
	0.1233	0.1599	0.1342	0.1206	0.1569	0.1355	0.2038	0.1288	0.1297	0.1493
	0.1475	0.1980	0.1444	0.1580	0.1604	0.1526	0.2081	0.1477	0.1466	0.1578
	0.1831	0.2514	0.1753	0.1592	0.2088	0.2144	0.2430	0.2018	0.2007	0.1799
	0.3221	0.2549	0.3568	0.3663	0.4044	0.3455	0.4219	0.3320	0.3380	0.3845
	0.5823	0.5441	0.5992	0.6230	0.6417	0.6067	0.6414	0.6011	0.5784	0.5949
r'	**1**	**1**	**2**	**2**	**1**	**1**	**4**	**1**	**2**	**1**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group b and c	0.1045	0.0692	0.0628	0.0745	0.0841	0.0629	0.0960	0.0516	0.0533	0.1139
	0.1176	0.0756	0.0641	0.0762	0.0853	0.0662	0.0967	0.0522	0.0636	0.1224
	0.1187	0.0758	0.0665	0.0799	0.0888	0.0702	0.1036	0.0551	0.0650	0.1248
	0.1225	0.0788	0.0698	0.0808	0.0897	0.0759	0.1039	0.0596	0.0679	0.1258
	0.1231	0.0816	0.0699	0.0813	0.0898	0.0763	0.1054	0.0615	0.0684	0.1294
	0.1234	0.0817	0.0700	0.0836	0.0905	0.0813	0.1090	0.0625	0.0691	0.1329
	0.1234	0.0819	0.0722	0.0849	0.0914	0.0838	0.1108	0.0666	0.0725	0.1329
	0.1238	0.0849	0.0734	0.0861	0.0962	0.0851	0.1113	0.0697	0.0765	0.1347
	0.1258	0.0849	0.0743	0.0878	0.0967	0.0858	0.1116	0.0739	0.0795	0.1355
	0.1261	0.0854	0.0766	0.0888	0.0971	0.0869	0.1119	0.0759	0.0814	0.1366
	0.1264	0.0862	0.0778	0.0896	0.0974	0.0875	0.1121	0.0759	0.0822	0.1371
	0.1273	0.0894	0.0844	0.0942	0.0996	0.0888	0.1125	0.0783	0.0855	0.1423
	0.1275	0.0897	0.0870	0.0942	0.1018	0.0889	0.1126	0.0794	0.0870	0.1431
	0.1282	0.0936	0.0903	0.0964	0.1022	0.0919	0.1129	0.0905	0.0877	0.1436
	0.1297	0.0949	0.0943	0.0971	0.1028	0.0933	0.1129	0.0969	0.0899	0.1473
	0.1326	0.0949	0.0950	0.0977	0.1063	0.0947	0.1139	0.0990	0.0932	0.1482
	0.1327	0.0969	0.0973	0.0997	0.1122	0.0947	0.1142	0.1013	0.0945	0.1518
	0.1328	0.0969	0.1028	0.0999	0.1138	0.0968	0.1145	0.1013	0.0966	0.1528
	0.1356	0.0992	0.1028	0.1002	0.1138	0.0973	0.1146	0.1015	0.0966	0.1548
	0.1359	0.1006	0.1033	0.1033	0.1155	0.1011	0.1162	0.1040	0.0977	0.1565
	0.1362	0.1022	0.1033	0.1034	0.1161	0.1016	0.1167	0.1098	0.0993	0.1568
	0.1365	0.1037	0.1087	0.1036	0.1185	0.1028	0.1179	0.1121	0.1036	0.1580
	0.1400	0.1111	0.1100	0.1074	0.1215	0.1048	0.1184	0.1165	0.1082	0.1581
	0.1424	0.1112	0.1115	0.1106	0.1216	0.1067	0.1186	0.1184	0.1084	0.1666
	0.1432	0.1221	0.1133	0.1160	0.1302	0.1216	0.1330	0.1215	0.1132	0.1678
	0.1435	0.1231	0.1354	0.1246	0.1329	0.1270	0.1501	0.1488	0.1197	0.1689
	0.1547	0.1396	0.1515	0.1660	0.1672	0.1281	0.1560	0.1863	0.1570	0.1750
	0.1773	0.1996	0.1913	0.1775	0.1929	0.1728	0.1743	0.2280	0.1919	0.2237
	0.4027	0.3351	0.3341	0.3490	0.3443	0.3434	0.3791	0.3063	0.3304	0.4300
	0.6350	0.5958	0.6064	0.6147	0.6047	0.5918	0.6598	0.5521	0.5746	0.6766
r'	**3**	**2**	**6**	**2**	**7**	**2**	**2**	**3**	**2**	**1**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group b and c	0.0924	0.0469	0.0877	0.1027	0.0919	0.1006	0.0924	0.0933	0.0651	0.0714
	0.1052	0.0541	0.0906	0.1080	0.3118	0.1018	0.1052	0.0961	0.0715	0.0736
	0.1058	0.0550	0.0938	0.1095	0.5761	0.1062	0.1058	0.0974	0.0735	0.0775
	0.1071	0.0560	0.0973	0.1100	0.5804	0.1092	0.1071	0.0984	0.0735	0.0831
	0.1072	0.0591	0.0979	0.1135	0.5865	0.1092	0.1072	0.0990	0.0808	0.0847
	0.1072	0.0646	0.0980	0.1154	0.5893	0.1117	0.1072	0.1000	0.0852	0.0866
	0.1077	0.0660	0.0984	0.1159	0.5906	0.1118	0.1077	0.1001	0.0861	0.0903
	0.1078	0.0712	0.0986	0.1163	0.5910	0.1155	0.1078	0.1018	0.0878	0.0917
	0.1136	0.0736	0.1008	0.1176	0.5933	0.1215	0.1136	0.1028	0.0903	0.0932
	0.1151	0.0754	0.1019	0.1184	0.5965	0.1221	0.1151	0.1028	0.0913	0.0937
	0.1161	0.0769	0.1019	0.1191	0.5970	0.1223	0.1161	0.1046	0.0948	0.0944
	0.1164	0.0797	0.1031	0.1210	0.6009	0.1233	0.1164	0.1055	0.0999	0.0948
	0.1169	0.0822	0.1046	0.1218	0.6018	0.1247	0.1169	0.1072	0.1001	0.0950
	0.1184	0.0847	0.1047	0.1237	0.6045	0.1270	0.1184	0.1075	0.1027	0.0956
	0.1188	0.0902	0.1058	0.1238	0.6050	0.1288	0.1188	0.1139	0.1095	0.0963
	0.1188	0.0925	0.1059	0.1240	0.6056	0.1313	0.1188	0.1139	0.1095	0.1029
	0.1191	0.0967	0.1065	0.1241	0.6056	0.1365	0.1191	0.1141	0.1114	0.1083
	0.1214	0.1005	0.1074	0.1241	0.6083	0.1380	0.1214	0.1146	0.1121	0.1083
	0.1216	0.1006	0.1084	0.1262	0.6106	0.1386	0.1216	0.1155	0.1164	0.1083
	0.1219	0.1006	0.1105	0.1266	0.6137	0.1409	0.1219	0.1161	0.1166	0.1086
	0.1238	0.1018	0.1121	0.1266	0.6154	0.1436	0.1238	0.1182	0.1194	0.1105
	0.1311	0.1025	0.1121	0.1273	0.6155	0.1451	0.1311	0.1197	0.1261	0.1141
	0.1313	0.1101	0.1134	0.1291	0.6171	0.1467	0.1313	0.1241	0.1269	0.1198
	0.1367	0.1122	0.1157	0.1335	0.6180	0.1537	0.1367	0.1257	0.1270	0.1214
	0.1437	0.1131	0.1166	0.1388	0.6245	0.1547	0.1437	0.1262	0.1391	0.1280
	0.1501	0.1447	0.1254	0.1520	0.6257	0.1588	0.1501	0.1269	0.1558	0.1452
	0.1543	0.1800	0.1453	0.1613	0.6288	0.1693	0.1543	0.1393	0.1897	0.1671
	0.1662	0.2282	0.1658	0.1677	0.6313	0.1938	0.1662	0.1596	0.2275	0.2044
	0.3748	0.3181	0.3505	0.3843	0.6567	0.3519	0.3748	0.3631	0.3296	0.3302
	0.6169	0.5775	0.5980	0.6505	0.6579	0.6120	0.6169	0.6101	0.5720	0.5836
r'	**4**	**1**	**2**	**1**	**5**	**1**	**3**	**1**	**1**	**2**

Table A.9: Ranked *sim* values for $(b \vee \{a, c\})$ in ACB dataset using M-KTS.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and c	0.1124	0.1134	0.1202	0.1626	0.1137	0.0807	0.1440	0.1375	0.0866	0.1855
	0.1128	0.1142	0.1203	0.1658	0.1168	0.0882	0.1535	0.1426	0.0869	0.1925
	0.1167	0.1148	0.1233	0.1667	0.1180	0.0911	0.1544	0.1439	0.0875	0.2011
	0.1180	0.1170	0.1271	0.1678	0.1233	0.0912	0.1557	0.1441	0.0892	0.2059
	0.1187	0.1188	0.1271	0.1697	0.1239	0.0931	0.1557	0.1462	0.0897	0.2136
	0.1198	0.1188	0.1276	0.1722	0.1262	0.0934	0.1607	0.1468	0.0902	0.2136
	0.1211	0.1196	0.1285	0.1722	0.1277	0.0937	0.1609	0.1473	0.0903	0.2139
	0.1211	0.1204	0.1296	0.1735	0.1277	0.0940	0.1611	0.1489	0.0906	0.2145
	0.1224	0.1216	0.1300	0.1770	0.1309	0.0941	0.1645	0.1493	0.0911	0.2148
	0.1247	0.1227	0.1305	0.1798	0.1311	0.0949	0.1701	0.1495	0.0912	0.2151
	0.1257	0.1236	0.1312	0.1812	0.1314	0.0949	0.1726	0.1495	0.0947	0.2152
	0.1272	0.1237	0.1316	0.1814	0.1325	0.0951	0.1734	0.1507	0.0960	0.2180
	0.1289	0.1242	0.1317	0.1815	0.1340	0.0952	0.1734	0.1510	0.0967	0.2202
	0.1297	0.1242	0.1324	0.1853	0.1348	0.0954	0.1753	0.1511	0.0970	0.2231
	0.1297	0.1242	0.1332	0.1878	0.1350	0.0980	0.1763	0.1526	0.0972	0.2269
	0.1314	0.1257	0.1337	0.1881	0.1378	0.1009	0.1767	0.1527	0.0984	0.2343
	0.1327	0.1263	0.1346	0.1883	0.1421	0.1010	0.1802	0.1528	0.0987	0.2356
	0.1328	0.1269	0.1358	0.1883	0.1432	0.1011	0.1832	0.1530	0.1008	0.2408
	0.1366	0.1270	0.1386	0.1917	0.1433	0.1026	0.1833	0.1551	0.1008	0.2409
	0.1378	0.1299	0.1394	0.1932	0.1463	0.1038	0.1858	0.1560	0.1017	0.2467
	0.1402	0.1312	0.1411	0.1945	0.1486	0.1044	0.1957	0.1569	0.1030	0.2519
	0.1407	0.1332	0.1426	0.1957	0.1494	0.1051	0.1995	0.1580	0.1031	0.2563
	0.1425	0.1377	0.1445	0.1973	0.1502	0.1108	0.2003	0.1593	0.1048	0.2570
	0.1466	0.1393	0.1454	0.1974	0.1515	0.1158	0.2013	0.1595	0.1088	0.2586
	0.1505	0.1430	0.1471	0.1991	0.1521	0.1162	0.2022	0.1658	0.1097	0.2599
	0.1548	0.1566	0.1573	0.2006	0.1522	0.1171	0.2061	0.1682	0.1151	0.2639
	0.1605	0.1580	0.1575	0.2007	0.1523	0.1306	0.2244	0.1800	0.1597	0.2712
	0.1651	0.1866	0.1668	0.2340	0.1789	0.1365	0.2588	0.1891	0.1689	0.2938
	0.3361	0.3458	0.3769	0.4577	0.4012	0.3820	0.4256	0.4005	0.3375	0.4901
	0.8211	0.8205	0.8563	0.9119	0.8552	0.8487	0.8445	0.8888	0.8295	0.9184
r'	**1**	**1**	**2**	**1**	**1**	**2**	**3**	**1**	**1**	**2**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and c	0.1328	0.1366	0.0784	0.1075	0.1082	0.0767	0.1217	0.0807	0.0717	0.1618
	0.1413	0.1370	0.0800	0.1112	0.1084	0.0791	0.1272	0.0826	0.0742	0.1633
	0.1439	0.1404	0.0800	0.1130	0.1130	0.0792	0.1287	0.0830	0.0743	0.1638
	0.1455	0.1416	0.0808	0.1130	0.1174	0.0821	0.1296	0.0834	0.0773	0.1640
	0.1456	0.1416	0.0809	0.1131	0.1175	0.0822	0.1311	0.0836	0.0795	0.1688
	0.1456	0.1436	0.0813	0.1136	0.1185	0.0824	0.1322	0.0851	0.0797	0.1688
	0.1458	0.1447	0.0826	0.1137	0.1192	0.0828	0.1322	0.0880	0.0800	0.1689
	0.1487	0.1456	0.0829	0.1151	0.1193	0.0832	0.1322	0.0883	0.0800	0.1691
	0.1507	0.1461	0.0856	0.1161	0.1194	0.0845	0.1324	0.0904	0.0810	0.1703
	0.1512	0.1468	0.0861	0.1166	0.1198	0.0854	0.1327	0.0915	0.0826	0.1769
	0.1513	0.1474	0.0883	0.1171	0.1199	0.0856	0.1335	0.0917	0.0835	0.1775
	0.1524	0.1477	0.0891	0.1175	0.1218	0.0858	0.1341	0.0920	0.0846	0.1780
	0.1557	0.1500	0.0925	0.1181	0.1222	0.0858	0.1343	0.0923	0.0847	0.1784
	0.1572	0.1509	0.0931	0.1186	0.1233	0.0861	0.1353	0.0937	0.0856	0.1789
	0.1582	0.1511	0.0934	0.1196	0.1241	0.0866	0.1354	0.0974	0.0868	0.1807
	0.1622	0.1514	0.0937	0.1203	0.1248	0.0884	0.1354	0.0976	0.0872	0.1809
	0.1626	0.1523	0.0948	0.1215	0.1251	0.0884	0.1358	0.0982	0.0909	0.1822
	0.1633	0.1532	0.0954	0.1222	0.1251	0.0909	0.1372	0.0994	0.0920	0.1837
	0.1635	0.1550	0.0965	0.1231	0.1256	0.0909	0.1376	0.1015	0.0957	0.1840
	0.1650	0.1554	0.0965	0.1233	0.1259	0.0920	0.1377	0.1041	0.0961	0.1850
	0.1669	0.1581	0.0990	0.1234	0.1260	0.0921	0.1379	0.1044	0.0961	0.1891
	0.1676	0.1585	0.1025	0.1239	0.1267	0.0940	0.1389	0.1044	0.0967	0.1926
	0.1694	0.1588	0.1030	0.1251	0.1293	0.0947	0.1399	0.1044	0.0987	0.1932
	0.1703	0.1597	0.1030	0.1295	0.1307	0.0952	0.1441	0.1090	0.1023	0.1934
	0.1717	0.1604	0.1065	0.1328	0.1360	0.1008	0.1561	0.1117	0.1024	0.1937
	0.1717	0.1605	0.1080	0.1467	0.1416	0.1073	0.1570	0.1168	0.1048	0.1954
	0.1720	0.1813	0.1535	0.1503	0.1635	0.1340	0.1683	0.1709	0.1652	0.1955
	0.1962	0.1834	0.1671	0.1624	0.1737	0.1375	0.1932	0.1749	0.1716	0.2298
	0.4002	0.3861	0.3398	0.3758	0.3858	0.3241	0.3880	0.3215	0.3273	0.4340
	0.8633	0.8600	0.8331	0.8515	0.8645	0.8152	0.8691	0.8211	0.8200	0.8837
r'	**3**	**1**	**5**	**3**	**6**	**4**	**4**	**1**	**2**	**2**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and c	0.1432	0.0935	0.1715	0.1484	0.3574	0.1249	0.1432	0.1372	0.1346	0.1384
	0.1433	0.0977	0.1743	0.1596	0.3997	0.1254	0.1433	0.1395	0.1361	0.1399
	0.1443	0.0989	0.1743	0.1633	0.4297	0.1306	0.1443	0.1426	0.1374	0.1402
	0.1473	0.0993	0.1760	0.1662	0.4352	0.1307	0.1473	0.1432	0.1374	0.1412
	0.1474	0.0997	0.1762	0.1662	0.4392	0.1307	0.1474	0.1432	0.1375	0.1412
	0.1474	0.0998	0.1776	0.1675	0.4392	0.1318	0.1474	0.1443	0.1391	0.1413
	0.1474	0.1011	0.1805	0.1684	0.4410	0.1339	0.1474	0.1473	0.1391	0.1414
	0.1479	0.1015	0.1811	0.1698	0.4425	0.1349	0.1479	0.1489	0.1417	0.1419
	0.1503	0.1019	0.1827	0.1703	0.4434	0.1353	0.1503	0.1494	0.1419	0.1423
	0.1506	0.1026	0.1827	0.1703	0.4453	0.1388	0.1506	0.1495	0.1421	0.1432
	0.1512	0.1045	0.1850	0.1706	0.4460	0.1424	0.1512	0.1506	0.1429	0.1454
	0.1522	0.1053	0.1854	0.1711	0.4490	0.1430	0.1522	0.1510	0.1431	0.1455
	0.1548	0.1059	0.1858	0.1714	0.4494	0.1436	0.1548	0.1514	0.1439	0.1485
	0.1551	0.1066	0.1859	0.1724	0.4537	0.1471	0.1551	0.1523	0.1458	0.1500
	0.1573	0.1073	0.1869	0.1728	0.4568	0.1475	0.1573	0.1536	0.1483	0.1510
	0.1579	0.1081	0.1892	0.1730	0.4608	0.1487	0.1579	0.1556	0.1487	0.1510
	0.1582	0.1082	0.1908	0.1734	0.4608	0.1499	0.1582	0.1577	0.1489	0.1521
	0.1588	0.1082	0.1911	0.1734	0.4677	0.1522	0.1588	0.1585	0.1495	0.1531
	0.1593	0.1082	0.1925	0.1737	0.4685	0.1537	0.1593	0.1601	0.1514	0.1546
	0.1614	0.1084	0.1929	0.1739	0.4731	0.1545	0.1614	0.1610	0.1530	0.1568
	0.1623	0.1088	0.1938	0.1773	0.4738	0.1559	0.1623	0.1645	0.1544	0.1575
	0.1631	0.1104	0.1945	0.1782	0.4770	0.1586	0.1631	0.1664	0.1545	0.1586
	0.1640	0.1109	0.1951	0.1793	0.4778	0.1592	0.1640	0.1664	0.1559	0.1606
	0.1654	0.1111	0.1981	0.1808	0.4804	0.1617	0.1654	0.1678	0.1609	0.1613
	0.1672	0.1175	0.1997	0.1933	0.4810	0.1636	0.1672	0.1698	0.1609	0.1614
	0.1722	0.1261	0.2001	0.1943	0.4825	0.1684	0.1722	0.1728	0.1655	0.1703
	0.1800	0.1638	0.2114	0.2045	0.4867	0.1729	0.1800	0.1841	0.1700	0.1725
	0.1839	0.1857	0.2211	0.2187	0.4927	0.1790	0.1839	0.1931	0.1757	0.1739
	0.3914	0.3436	0.4286	0.4390	0.5007	0.3510	0.3914	0.3916	0.3773	0.3792
	0.8715	0.8358	0.9053	0.9150	0.5986	0.8288	0.8715	0.8688	0.8534	0.8607
r'	**5**	**1**	**1**	**2**	**4**	**2**	**2**	**1**	**1**	**1**

TABLE A.10: Ranked *sim* values for $(c \vee \{a, b\})$ in ACB dataset using M-KTS.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and b	0.1159	0.1110	0.1193	0.1597	0.1137	0.0799	0.1390	0.1388	0.0843	0.1840
	0.1168	0.1111	0.1208	0.1605	0.1156	0.0870	0.1414	0.1402	0.0883	0.1890
	0.1182	0.1150	0.1226	0.1620	0.1194	0.0927	0.1536	0.1404	0.0887	0.1904
	0.1187	0.1154	0.1233	0.1655	0.1246	0.0928	0.1571	0.1407	0.0893	0.1953
	0.1202	0.1183	0.1268	0.1704	0.1247	0.0930	0.1617	0.1408	0.0893	0.2034
	0.1215	0.1188	0.1268	0.1731	0.1247	0.0959	0.1668	0.1419	0.0907	0.2060
	0.1218	0.1191	0.1277	0.1735	0.1250	0.0967	0.1669	0.1419	0.0910	0.2100
	0.1230	0.1199	0.1284	0.1741	0.1252	0.0970	0.1687	0.1432	0.0931	0.2119
	0.1250	0.1202	0.1294	0.1749	0.1261	0.0972	0.1691	0.1473	0.0936	0.2143
	0.1250	0.1207	0.1302	0.1751	0.1289	0.0975	0.1702	0.1474	0.0947	0.2170
	0.1250	0.1213	0.1315	0.1777	0.1300	0.0982	0.1737	0.1476	0.0948	0.2218
	0.1253	0.1214	0.1315	0.1779	0.1301	0.0988	0.1744	0.1486	0.0948	0.2237
	0.1264	0.1219	0.1315	0.1791	0.1314	0.1001	0.1745	0.1486	0.0956	0.2248
	0.1265	0.1229	0.1319	0.1793	0.1322	0.1003	0.1752	0.1489	0.0963	0.2269
	0.1269	0.1237	0.1341	0.1801	0.1349	0.1004	0.1766	0.1514	0.0987	0.2278
	0.1274	0.1237	0.1344	0.1813	0.1355	0.1006	0.1833	0.1514	0.0992	0.2278
	0.1288	0.1249	0.1345	0.1821	0.1355	0.1030	0.1847	0.1514	0.1002	0.2278
	0.1324	0.1268	0.1359	0.1821	0.1355	0.1046	0.1848	0.1515	0.1002	0.2279
	0.1328	0.1269	0.1361	0.1832	0.1373	0.1055	0.1851	0.1520	0.1010	0.2291
	0.1347	0.1275	0.1361	0.1839	0.1377	0.1060	0.1901	0.1530	0.1035	0.2302
	0.1359	0.1310	0.1367	0.1849	0.1390	0.1060	0.1909	0.1535	0.1035	0.2335
	0.1378	0.1319	0.1389	0.1887	0.1418	0.1076	0.1916	0.1546	0.1038	0.2344
	0.1378	0.1340	0.1411	0.1899	0.1455	0.1087	0.1939	0.1552	0.1053	0.2392
	0.1379	0.1345	0.1411	0.1902	0.1457	0.1096	0.1952	0.1555	0.1060	0.2454
	0.1388	0.1355	0.1432	0.1921	0.1513	0.1109	0.1952	0.1602	0.1101	0.2491
	0.1416	0.1382	0.1454	0.1923	0.1529	0.1125	0.1973	0.1635	0.1226	0.2558
	0.1459	0.1414	0.1471	0.2028	0.1558	0.1177	0.2013	0.1644	0.1238	0.2585
	0.1474	0.1498	0.1486	0.2050	0.1590	0.1181	0.2027	0.1649	0.1285	0.2599
	0.1536	0.1616	0.1510	0.2159	0.1605	0.1216	0.2077	0.1705	0.1494	0.2795
	0.8041	0.8063	0.8398	0.8980	0.8438	0.8440	0.8652	0.8607	0.8226	0.8893
r'	**2**	**2**	**3**	**2**	**1**	**1**	**2**	**1**	**1**	**1**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and b	0.1376	0.1338	0.0777	0.1052	0.1044	0.0766	0.1196	0.0786	0.0724	0.1547
	0.1380	0.1366	0.0804	0.1076	0.1131	0.0802	0.1216	0.0822	0.0766	0.1558
	0.1430	0.1390	0.0809	0.1092	0.1140	0.0809	0.1255	0.0824	0.0782	0.1558
	0.1457	0.1390	0.0821	0.1105	0.1148	0.0810	0.1263	0.0825	0.0783	0.1595
	0.1460	0.1401	0.0822	0.1110	0.1155	0.0813	0.1266	0.0866	0.0784	0.1606
	0.1464	0.1401	0.0831	0.1124	0.1160	0.0816	0.1268	0.0874	0.0796	0.1705
	0.1473	0.1401	0.0834	0.1132	0.1165	0.0824	0.1290	0.0887	0.0811	0.1707
	0.1474	0.1405	0.0844	0.1144	0.1174	0.0838	0.1294	0.0894	0.0818	0.1715
	0.1504	0.1407	0.0857	0.1155	0.1178	0.0843	0.1304	0.0900	0.0818	0.1717
	0.1518	0.1408	0.0860	0.1162	0.1181	0.0860	0.1309	0.0904	0.0821	0.1731
	0.1518	0.1426	0.0875	0.1173	0.1194	0.0869	0.1309	0.0912	0.0823	0.1749
	0.1536	0.1432	0.0878	0.1175	0.1195	0.0874	0.1322	0.0913	0.0836	0.1761
	0.1554	0.1477	0.0884	0.1185	0.1198	0.0884	0.1326	0.0913	0.0836	0.1762
	0.1562	0.1479	0.0887	0.1193	0.1202	0.0890	0.1335	0.0928	0.0840	0.1770
	0.1562	0.1481	0.0889	0.1197	0.1206	0.0890	0.1344	0.0941	0.0876	0.1771
	0.1566	0.1495	0.0889	0.1206	0.1213	0.0895	0.1349	0.0952	0.0882	0.1777
	0.1571	0.1526	0.0939	0.1208	0.1218	0.0905	0.1351	0.0957	0.0905	0.1800
	0.1573	0.1532	0.0944	0.1210	0.1231	0.0906	0.1399	0.0971	0.0905	0.1822
	0.1580	0.1537	0.0944	0.1218	0.1231	0.0924	0.1399	0.0973	0.0911	0.1833
	0.1590	0.1542	0.0965	0.1259	0.1247	0.0936	0.1405	0.1013	0.0915	0.1840
	0.1590	0.1544	0.0966	0.1259	0.1272	0.0936	0.1412	0.1051	0.0945	0.1842
	0.1592	0.1558	0.0985	0.1276	0.1284	0.0961	0.1446	0.1051	0.0954	0.1859
	0.1613	0.1578	0.1007	0.1289	0.1303	0.0975	0.1451	0.1062	0.0977	0.1859
	0.1637	0.1593	0.1082	0.1317	0.1328	0.0983	0.1452	0.1099	0.0977	0.1872
	0.1641	0.1613	0.1090	0.1320	0.1354	0.0989	0.1453	0.1167	0.0997	0.1878
	0.1691	0.1614	0.1157	0.1326	0.1377	0.1077	0.1483	0.1228	0.1104	0.1942
	0.1709	0.1633	0.1177	0.1349	0.1380	0.1089	0.1517	0.1272	0.1105	0.1944
	0.1719	0.1645	0.1210	0.1363	0.1458	0.1090	0.1529	0.1312	0.1192	0.1987
	0.1836	0.1698	0.1387	0.1431	0.1578	0.1284	0.1615	0.1582	0.1276	0.2061
	0.8606	0.8366	0.8219	0.8368	0.8484	0.8093	0.8490	0.8077	0.8060	0.8819
r'	**2**	**4**	**4**	**1**	**6**	**3**	**1**	**2**	**3**	**2**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and b	0.1405	0.0937	0.1719	0.1487	0.3383	0.1237	0.1405	0.1350	0.1381	0.1322
	0.1409	0.0945	0.1727	0.1576	0.3940	0.1324	0.1409	0.1368	0.1395	0.1358
	0.1409	0.0947	0.1767	0.1598	0.4250	0.1327	0.1409	0.1424	0.1402	0.1364
	0.1438	0.0972	0.1769	0.1608	0.4257	0.1340	0.1438	0.1438	0.1402	0.1379
	0.1447	0.0978	0.1781	0.1624	0.4278	0.1352	0.1447	0.1450	0.1404	0.1422
	0.1457	0.0994	0.1786	0.1629	0.4356	0.1366	0.1457	0.1460	0.1428	0.1428
	0.1472	0.0995	0.1802	0.1631	0.4374	0.1374	0.1472	0.1465	0.1430	0.1430
	0.1477	0.1002	0.1831	0.1636	0.4434	0.1378	0.1477	0.1465	0.1430	0.1436
	0.1480	0.1013	0.1831	0.1647	0.4472	0.1378	0.1480	0.1477	0.1442	0.1443
	0.1511	0.1016	0.1837	0.1647	0.4472	0.1389	0.1511	0.1487	0.1448	0.1445
	0.1516	0.1025	0.1844	0.1664	0.4511	0.1398	0.1516	0.1492	0.1448	0.1448
	0.1525	0.1025	0.1849	0.1677	0.4512	0.1420	0.1525	0.1493	0.1448	0.1448
	0.1534	0.1031	0.1855	0.1684	0.4520	0.1425	0.1534	0.1503	0.1449	0.1457
	0.1538	0.1039	0.1862	0.1694	0.4558	0.1438	0.1538	0.1506	0.1462	0.1469
	0.1553	0.1052	0.1862	0.1701	0.4577	0.1440	0.1553	0.1525	0.1466	0.1471
	0.1556	0.1067	0.1869	0.1705	0.4598	0.1469	0.1556	0.1529	0.1469	0.1477
	0.1557	0.1067	0.1870	0.1714	0.4606	0.1479	0.1557	0.1532	0.1483	0.1480
	0.1576	0.1090	0.1887	0.1745	0.4610	0.1489	0.1576	0.1533	0.1494	0.1491
	0.1579	0.1090	0.1888	0.1760	0.4652	0.1490	0.1579	0.1540	0.1498	0.1498
	0.1581	0.1091	0.1894	0.1761	0.4667	0.1492	0.1581	0.1545	0.1502	0.1517
	0.1591	0.1116	0.1900	0.1773	0.4667	0.1501	0.1591	0.1545	0.1525	0.1518
	0.1600	0.1125	0.1901	0.1781	0.4675	0.1505	0.1600	0.1561	0.1527	0.1543
	0.1612	0.1149	0.1914	0.1814	0.4676	0.1517	0.1612	0.1570	0.1540	0.1550
	0.1613	0.1154	0.1930	0.1815	0.4684	0.1520	0.1613	0.1601	0.1553	0.1556
	0.1667	0.1167	0.1954	0.1821	0.4699	0.1577	0.1667	0.1602	0.1566	0.1579
	0.1680	0.1311	0.1987	0.1827	0.4721	0.1587	0.1680	0.1638	0.1588	0.1596
	0.1688	0.1318	0.1997	0.1920	0.4739	0.1593	0.1688	0.1700	0.1591	0.1616
	0.1689	0.1362	0.2019	0.1950	0.4754	0.1598	0.1689	0.1704	0.1657	0.1618
	0.1696	0.1590	0.2072	0.1950	0.4903	0.1605	0.1696	0.1726	0.1701	0.1631
	0.8523	0.8224	0.8869	0.8911	0.5983	0.8056	0.8523	0.8492	0.8439	0.8475
r'	**3**	**1**	**3**	**1**	**4**	**1**	**2**	**1**	**2**	**1**

Table A.11: Ranked *sim* values for $(a \vee \{b, c\})$ in ACB dataset using FVR.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group b and c	0.6821	0.6409	0.7964	0.7776	0.7892	0.7711	0.6540	0.8789	0.7948	0.7663
	0.6690	0.6403	0.7307	0.7667	0.7566	0.7440	0.6503	0.8600	0.7484	0.6755
	0.6680	0.6207	0.7241	0.7634	0.7428	0.7371	0.6262	0.8173	0.7223	0.6494
	0.6567	0.6171	0.7063	0.7093	0.7425	0.7148	0.6117	0.7834	0.7202	0.6039
	0.6137	0.6146	0.6725	0.6727	0.7242	0.7085	0.6115	0.7772	0.7132	0.6010
	0.5881	0.5967	0.6443	0.6683	0.6949	0.7045	0.5817	0.7741	0.7120	0.6003
	0.5689	0.5957	0.6327	0.6651	0.6946	0.7010	0.5748	0.7667	0.7081	0.5838
	0.5530	0.5934	0.6134	0.6595	0.6937	0.6903	0.5688	0.7489	0.7010	0.5786
	0.5487	0.5926	0.5880	0.6587	0.6925	0.6834	0.5660	0.7422	0.7008	0.5699
	0.5038	0.5895	0.5826	0.6480	0.6877	0.6740	0.5632	0.7389	0.6814	0.5600
	0.5032	0.5860	0.5809	0.6463	0.6572	0.6661	0.5400	0.7347	0.6801	0.5596
	0.4953	0.5741	0.5805	0.6463	0.6541	0.6659	0.5307	0.7159	0.6563	0.5456
	0.4906	0.5726	0.5804	0.6391	0.6458	0.6607	0.5302	0.7146	0.6562	0.5448
	0.4830	0.5655	0.5597	0.6379	0.6441	0.6577	0.5283	0.6985	0.6550	0.5353
	0.4824	0.5568	0.5516	0.6229	0.6405	0.6509	0.5183	0.6966	0.6538	0.5227
	0.4770	0.5316	0.5505	0.6043	0.6224	0.6499	0.5088	0.6741	0.6348	0.5184
	0.4763	0.5310	0.5434	0.5964	0.6156	0.6450	0.4900	0.6738	0.6298	0.5126
	0.4580	0.5291	0.5388	0.5944	0.5867	0.6171	0.4870	0.6717	0.6283	0.5126
	0.4575	0.5239	0.5351	0.5895	0.5856	0.6060	0.4837	0.6707	0.6271	0.5096
	0.4530	0.5171	0.5212	0.5882	0.5756	0.6060	0.4761	0.6680	0.6120	0.5042
	0.4509	0.5063	0.5137	0.5816	0.5753	0.5937	0.4702	0.6582	0.6043	0.4740
	0.4384	0.4980	0.5094	0.5782	0.5682	0.5893	0.4686	0.6582	0.5999	0.4719
	0.4371	0.4964	0.5013	0.5698	0.5536	0.5849	0.4686	0.6540	0.5897	0.4634
	0.4236	0.4804	0.4897	0.5695	0.5430	0.5791	0.4390	0.6502	0.5849	0.4591
	0.4236	0.4804	0.4750	0.5325	0.5429	0.5278	0.4290	0.6470	0.5849	0.4471
	0.4180	0.4691	0.4333	0.5269	0.5401	0.5234	0.4264	0.5934	0.5849	0.4061
	0.3945	0.4655	0.4247	0.5114	0.5203	0.5234	0.4189	0.5933	0.5808	0.3787
	0.3897	0.4443	0.4064	0.5103	0.4595	0.5143	0.4183	0.5854	0.5772	0.3774
	0.3821	0.4355	0.4064	0.5058	0.4595	0.4453	0.4145	0.5541	0.5715	0.3657
	0.3449	0.3492	0.3898	0.4493	0.4299	0.4270	0.3938	0.4885	0.5685	0.3449
r'	**7**	**10**	**12**	**14**	**11**	**5**	**7**	**1**	**3**	**24**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group b and c	0.6913	0.8079	0.6951	0.5682	0.7006	0.6595	0.7980	0.6595	0.7199	0.6627
	0.6791	0.7262	0.6545	0.5216	0.7006	0.6532	0.7432	0.6532	0.6737	0.6484
	0.6736	0.6832	0.6474	0.5156	0.6703	0.6510	0.7355	0.6510	0.6673	0.6412
	0.6577	0.6817	0.6436	0.4868	0.6659	0.6316	0.7281	0.6316	0.6544	0.6383
	0.6524	0.6815	0.6318	0.4808	0.6332	0.6163	0.7234	0.6163	0.6242	0.6337
	0.6462	0.6814	0.6226	0.4757	0.6152	0.6144	0.7004	0.6144	0.6186	0.6269
	0.6443	0.6525	0.6193	0.4755	0.5910	0.6009	0.7002	0.6009	0.6077	0.6164
	0.6353	0.6511	0.6066	0.4751	0.5830	0.5973	0.6937	0.5973	0.5677	0.6020
	0.6233	0.6440	0.6058	0.4513	0.5567	0.5898	0.6833	0.5898	0.5530	0.6019
	0.6233	0.6327	0.6020	0.4438	0.5350	0.5868	0.6793	0.5868	0.5471	0.5884
	0.6204	0.6233	0.5993	0.4410	0.5312	0.5853	0.6793	0.5853	0.5456	0.5870
	0.6203	0.6232	0.5910	0.4338	0.4884	0.5727	0.6639	0.5727	0.5255	0.5452
	0.6149	0.6217	0.5782	0.4273	0.4812	0.5718	0.6417	0.5718	0.5244	0.5443
	0.6085	0.6167	0.5771	0.4218	0.4721	0.5664	0.6390	0.5664	0.5244	0.5443
	0.6065	0.6004	0.5758	0.4200	0.4703	0.5599	0.6338	0.5599	0.5153	0.5253
	0.5965	0.5964	0.5650	0.4179	0.4605	0.5593	0.6298	0.5593	0.5057	0.5180
	0.5885	0.5938	0.5406	0.4116	0.4581	0.5418	0.6280	0.5418	0.4991	0.5169
	0.5715	0.5862	0.5349	0.4058	0.4520	0.5218	0.6275	0.5218	0.4887	0.5088
	0.5715	0.5836	0.5248	0.3921	0.4486	0.5006	0.6167	0.5006	0.4864	0.5017
	0.5451	0.5745	0.5112	0.3836	0.4453	0.4974	0.6016	0.4974	0.4803	0.5006
	0.5406	0.5623	0.5037	0.3821	0.4408	0.4938	0.5988	0.4938	0.4794	0.4999
	0.5304	0.5540	0.5010	0.3698	0.4358	0.4936	0.5978	0.4936	0.4733	0.4759
	0.5226	0.5337	0.4879	0.3685	0.4286	0.4796	0.5900	0.4796	0.4732	0.4614
	0.5215	0.5226	0.4729	0.3590	0.4227	0.4772	0.5891	0.4772	0.4681	0.4395
	0.5203	0.5190	0.4664	0.3587	0.4030	0.4764	0.5784	0.4764	0.4635	0.4314
	0.5031	0.4961	0.4342	0.3549	0.3964	0.4648	0.5562	0.4648	0.4634	0.4079
	0.5029	0.4942	0.4244	0.3284	0.3918	0.4518	0.4965	0.4518	0.4457	0.3950
	0.4690	0.4556	0.3847	0.2991	0.3761	0.4518	0.4920	0.4518	0.4413	0.3903
	0.4666	0.4556	0.3847	0.2991	0.3706	0.4156	0.4714	0.4156	0.3982	0.3674
	0.4327	0.4254	0.3720	0.2764	0.3474	0.3827	0.4335	0.3827	0.3981	0.3260
r'	**12**	**16**	**24**	**8**	**25**	**13**	**24**	**8**	**9**	**20**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group b and c	0.6859	0.6611	0.7769	0.6735	0.7167	0.7353	0.6875	0.8155	0.7323	0.7654
	0.6264	0.6119	0.7560	0.6081	0.6477	0.7128	0.6229	0.7202	0.6495	0.7525
	0.6249	0.5769	0.7538	0.5906	0.6095	0.7112	0.6051	0.6852	0.5867	0.7506
	0.6007	0.5609	0.7178	0.5709	0.6008	0.7036	0.5985	0.6731	0.5537	0.7018
	0.6000	0.5538	0.7142	0.5542	0.5938	0.6972	0.5713	0.6669	0.5249	0.6986
	0.5979	0.5440	0.6805	0.5475	0.5789	0.6961	0.5696	0.6556	0.5234	0.6936
	0.5670	0.5437	0.6698	0.5414	0.5786	0.6825	0.5610	0.6491	0.5233	0.6890
	0.5659	0.5428	0.6613	0.5279	0.5634	0.6824	0.5553	0.6471	0.5146	0.6778
	0.5579	0.5399	0.6574	0.5238	0.5621	0.6761	0.5481	0.6439	0.5044	0.6732
	0.5570	0.5344	0.6556	0.5221	0.5373	0.6456	0.5374	0.6243	0.4677	0.6690
	0.5262	0.5306	0.6285	0.5178	0.5235	0.6054	0.5244	0.6221	0.4672	0.6518
	0.5262	0.5263	0.6204	0.5053	0.5220	0.6007	0.5199	0.5876	0.4666	0.6432
	0.5168	0.5136	0.6004	0.4947	0.5204	0.6007	0.5134	0.5840	0.4662	0.6316
	0.5149	0.5109	0.5956	0.4827	0.5179	0.6000	0.5110	0.5807	0.4598	0.6264
	0.4884	0.4830	0.5956	0.4728	0.5172	0.5883	0.5047	0.5797	0.4576	0.6161
	0.4836	0.4795	0.5948	0.4707	0.5073	0.5862	0.4931	0.5760	0.4498	0.6143
	0.4774	0.4735	0.5918	0.4403	0.5062	0.5646	0.4676	0.5614	0.4491	0.5897
	0.4760	0.4666	0.5904	0.4338	0.5050	0.5611	0.4635	0.5582	0.4416	0.5842
	0.4725	0.4560	0.5822	0.4269	0.4984	0.5507	0.4577	0.5582	0.4386	0.5766
	0.4611	0.4481	0.5710	0.4269	0.4916	0.5495	0.4530	0.5560	0.3948	0.5751
	0.4610	0.4476	0.5708	0.4260	0.4910	0.5475	0.4444	0.5462	0.3937	0.5737
	0.4607	0.3959	0.5631	0.4239	0.4906	0.5374	0.4374	0.5451	0.3908	0.5546
	0.4606	0.3801	0.5619	0.4133	0.4488	0.5239	0.4333	0.5368	0.3908	0.5369
	0.4567	0.3756	0.5453	0.4104	0.4463	0.5233	0.4273	0.5282	0.3898	0.5306
	0.4525	0.3754	0.5377	0.4059	0.4439	0.5202	0.4013	0.5249	0.3893	0.5284
	0.4293	0.3669	0.5253	0.3920	0.4340	0.5023	0.4013	0.4999	0.3845	0.5127
	0.4020	0.3669	0.5023	0.3903	0.4340	0.4887	0.3967	0.4992	0.3770	0.4799
	0.3617	0.3646	0.4986	0.3857	0.4230	0.4887	0.3780	0.4888	0.3700	0.4675
	0.3617	0.3492	0.4725	0.3740	0.4155	0.4832	0.3521	0.4830	0.3667	0.4385
	0.2945	0.3478	0.4725	0.3660	0.4036	0.4247	0.3463	0.4554	0.2727	0.4385
r'	**13**	**16**	**17**	**10**	**17**	**10**	**14**	**6**	**2**	**2**

Table A.12: Ranked *sim* values for $(b \vee \{a, c\})$ in ACB dataset using FVR.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and c	0.7043	0.7343	0.7270	0.8098	0.7575	0.7667	0.7979	0.8475	0.8219	0.7128
	0.6302	0.6930	0.6960	0.7422	0.7122	0.7538	0.7600	0.8239	0.8085	0.7107
	0.6127	0.6762	0.6679	0.7411	0.6785	0.7075	0.6968	0.8129	0.7567	0.6244
	0.6123	0.6577	0.6675	0.7283	0.6696	0.7015	0.6936	0.8087	0.7546	0.6185
	0.5107	0.6504	0.6640	0.7170	0.6517	0.6995	0.6929	0.7495	0.7355	0.6183
	0.4983	0.6439	0.6403	0.7040	0.6394	0.6829	0.6769	0.7219	0.7272	0.5817
	0.4963	0.6326	0.6122	0.7039	0.6339	0.6656	0.6743	0.7176	0.7243	0.5738
	0.4872	0.6313	0.6094	0.6943	0.6295	0.6512	0.6730	0.7173	0.7186	0.5399
	0.4866	0.6267	0.5872	0.6884	0.6256	0.6450	0.6730	0.7123	0.7122	0.5280
	0.4864	0.6199	0.5862	0.6766	0.6060	0.6130	0.6609	0.7028	0.7116	0.5178
	0.4853	0.6065	0.5683	0.6747	0.5994	0.6042	0.6190	0.7011	0.6827	0.5144
	0.4724	0.6030	0.5650	0.6524	0.5983	0.6007	0.6163	0.7006	0.6661	0.5079
	0.4684	0.5977	0.5627	0.6502	0.5983	0.5928	0.6136	0.6927	0.6636	0.5045
	0.4640	0.5861	0.5533	0.6499	0.5972	0.5824	0.6136	0.6898	0.6503	0.5042
	0.4639	0.5859	0.5518	0.6416	0.5888	0.5765	0.6088	0.6874	0.6457	0.4969
	0.4611	0.5770	0.5459	0.6354	0.5847	0.5760	0.6036	0.6869	0.6451	0.4955
	0.4324	0.5561	0.5456	0.6241	0.5832	0.5748	0.5919	0.6810	0.6405	0.4955
	0.4287	0.5535	0.5347	0.6180	0.5832	0.5692	0.5814	0.6678	0.6389	0.4898
	0.4202	0.5436	0.5347	0.6085	0.5781	0.5637	0.5792	0.6673	0.6336	0.4885
	0.4176	0.5409	0.5312	0.5988	0.5756	0.5637	0.5776	0.6575	0.6270	0.4847
	0.4137	0.5387	0.5308	0.5980	0.5536	0.5583	0.5589	0.6563	0.6262	0.4809
	0.4076	0.5387	0.5195	0.5948	0.5500	0.5497	0.5381	0.6419	0.6217	0.4803
	0.3946	0.5384	0.5169	0.5948	0.5494	0.5327	0.5293	0.6256	0.6217	0.4796
	0.3938	0.5254	0.4981	0.5923	0.5449	0.5305	0.5244	0.6077	0.6202	0.4675
	0.3936	0.5039	0.4922	0.5909	0.5342	0.5273	0.5217	0.6077	0.5851	0.4599
	0.3900	0.5014	0.4666	0.5530	0.5243	0.5207	0.5197	0.5978	0.5841	0.4459
	0.3900	0.5000	0.4666	0.5500	0.5059	0.4697	0.5041	0.5697	0.5807	0.4429
	0.3900	0.4983	0.4400	0.5450	0.5052	0.4535	0.4993	0.5692	0.5786	0.4283
	0.3628	0.4972	0.4286	0.5418	0.5030	0.4524	0.4967	0.5432	0.5365	0.4240
	0.3564	0.3419	0.3818	0.5023	0.5010	0.4141	0.4921	0.5399	0.5208	0.4185
r'	**9**	**12**	**13**	**19**	**10**	**6**	**4**	**1**	**2**	**20**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and c	0.7548	0.8412	0.8149	0.7257	0.7247	0.8425	0.7910	0.8425	0.7194	0.7269
	0.7398	0.8192	0.6918	0.6956	0.6788	0.7403	0.7398	0.7403	0.6826	0.6968
	0.6918	0.7540	0.6557	0.6722	0.6786	0.6881	0.6878	0.6881	0.6756	0.6792
	0.6828	0.7515	0.6471	0.6393	0.6085	0.6696	0.6830	0.6696	0.6166	0.6316
	0.6800	0.7512	0.6295	0.6288	0.6045	0.6649	0.6729	0.6649	0.6125	0.6205
	0.6800	0.7206	0.6070	0.6269	0.5789	0.6562	0.6644	0.6562	0.6115	0.6117
	0.6666	0.7111	0.5803	0.6215	0.5653	0.6501	0.6535	0.6501	0.6025	0.6091
	0.6629	0.7106	0.5799	0.6074	0.5628	0.6428	0.6523	0.6428	0.6006	0.5872
	0.6628	0.6911	0.5751	0.6074	0.5451	0.6315	0.6504	0.6315	0.5880	0.5742
	0.6534	0.6789	0.5681	0.5982	0.5430	0.6290	0.6504	0.6290	0.5854	0.5475
	0.6175	0.6707	0.5634	0.5879	0.5392	0.6235	0.6164	0.6235	0.5818	0.5428
	0.6140	0.6689	0.5626	0.5866	0.5275	0.6156	0.6107	0.6156	0.5735	0.5376
	0.6088	0.6668	0.5571	0.5847	0.5275	0.5989	0.6086	0.5989	0.5735	0.5350
	0.5973	0.6522	0.5424	0.5838	0.5089	0.5950	0.5997	0.5950	0.5670	0.5314
	0.5939	0.6521	0.5406	0.5808	0.5067	0.5919	0.5977	0.5919	0.5626	0.5190
	0.5873	0.6479	0.5376	0.5725	0.5058	0.5873	0.5944	0.5873	0.5569	0.5033
	0.5688	0.6441	0.5368	0.5554	0.5038	0.5786	0.5903	0.5786	0.5437	0.5032
	0.5667	0.6298	0.5346	0.5503	0.4914	0.5765	0.5824	0.5765	0.5369	0.5007
	0.5662	0.6236	0.5034	0.5373	0.4897	0.5684	0.5778	0.5684	0.5294	0.4989
	0.5627	0.6226	0.4773	0.5207	0.4881	0.5684	0.5753	0.5684	0.5214	0.4983
	0.5478	0.6124	0.4773	0.5142	0.4806	0.5621	0.5673	0.5621	0.5140	0.4983
	0.5405	0.6106	0.4767	0.5057	0.4715	0.5617	0.5658	0.5617	0.4923	0.4968
	0.5385	0.6093	0.4757	0.4950	0.4569	0.5268	0.5623	0.5268	0.4830	0.4927
	0.5360	0.6059	0.4669	0.4819	0.4524	0.5216	0.5554	0.5216	0.4762	0.4879
	0.5120	0.6059	0.4624	0.4749	0.4510	0.5125	0.5517	0.5125	0.4725	0.4763
	0.5071	0.5915	0.4591	0.4747	0.4340	0.5101	0.5476	0.5101	0.4610	0.4685
	0.4854	0.5879	0.4570	0.4598	0.4096	0.4951	0.5442	0.4951	0.4555	0.4566
	0.4806	0.5874	0.4536	0.4585	0.3951	0.4918	0.5238	0.4918	0.4424	0.4445
	0.4713	0.4982	0.4431	0.4490	0.3777	0.4870	0.5145	0.4870	0.4162	0.4103
	0.4271	0.4867	0.4356	0.4116	0.2968	0.4836	0.3964	0.4836	0.4057	0.3987
r'	**10**	**14**	**23**	**8**	**24**	**12**	**20**	**6**	**10**	**21**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and c	0.7019	0.6450	0.6966	0.7608	0.7055	0.7622	0.7787	0.7124	0.7311	0.8668
	0.6897	0.5914	0.6544	0.7586	0.6701	0.7236	0.6717	0.7124	0.7227	0.8212
	0.6761	0.5864	0.6456	0.7356	0.6555	0.7114	0.6496	0.7122	0.7073	0.7900
	0.6707	0.5864	0.6404	0.7185	0.6485	0.6918	0.6243	0.6887	0.6865	0.7680
	0.6577	0.5820	0.6282	0.6689	0.6325	0.6909	0.5891	0.6725	0.6292	0.7631
	0.6209	0.5819	0.5973	0.6655	0.6276	0.6780	0.5889	0.6491	0.6290	0.7619
	0.6142	0.5700	0.5973	0.6402	0.6157	0.6654	0.5761	0.6487	0.6256	0.7453
	0.6042	0.5663	0.5968	0.6307	0.6032	0.6524	0.5701	0.6458	0.6070	0.7359
	0.5991	0.5640	0.5897	0.5880	0.6012	0.6514	0.5617	0.6080	0.6050	0.7162
	0.5852	0.5600	0.5878	0.5433	0.6012	0.6430	0.5589	0.5962	0.6038	0.7052
	0.5725	0.5527	0.5766	0.5367	0.5944	0.6336	0.5328	0.5944	0.5854	0.7004
	0.5713	0.5305	0.5755	0.5321	0.5903	0.6309	0.5153	0.5844	0.5819	0.7003
	0.5669	0.5077	0.5666	0.5311	0.5855	0.6253	0.5016	0.5806	0.5650	0.6948
	0.5598	0.5051	0.5566	0.5296	0.5771	0.6088	0.4938	0.5799	0.5548	0.6926
	0.5484	0.5013	0.5524	0.5128	0.5753	0.5999	0.4862	0.5710	0.5500	0.6924
	0.5419	0.4968	0.5183	0.5034	0.5619	0.5994	0.4856	0.5639	0.5400	0.6891
	0.5387	0.4920	0.5180	0.5023	0.5557	0.5901	0.4820	0.5628	0.5400	0.6842
	0.5267	0.4786	0.5048	0.4837	0.5552	0.5875	0.4781	0.5481	0.5249	0.6807
	0.5187	0.4748	0.4960	0.4837	0.5453	0.5813	0.4681	0.5419	0.5224	0.6780
	0.5074	0.4748	0.4960	0.4808	0.5170	0.5786	0.4562	0.5405	0.5098	0.6776
	0.5015	0.4712	0.4719	0.4753	0.5161	0.5754	0.4554	0.5345	0.5066	0.6726
	0.4998	0.4704	0.4687	0.4680	0.5083	0.5747	0.4554	0.5341	0.5054	0.6702
	0.4872	0.4664	0.4673	0.4544	0.4971	0.5679	0.4461	0.5334	0.5033	0.6589
	0.4739	0.4629	0.4531	0.4451	0.4916	0.5641	0.4394	0.5238	0.5007	0.6260
	0.4731	0.4560	0.4466	0.4448	0.4769	0.5337	0.4264	0.5223	0.4922	0.6222
	0.4658	0.4555	0.4263	0.4418	0.4750	0.5228	0.4249	0.5181	0.4874	0.6136
	0.4658	0.4463	0.4170	0.3820	0.4702	0.5224	0.4098	0.4903	0.4784	0.6135
	0.4082	0.4427	0.4127	0.3635	0.4528	0.5224	0.3994	0.4646	0.4447	0.6135
	0.3941	0.4336	0.3840	0.3624	0.4499	0.5140	0.3925	0.4585	0.4441	0.6086
	0.3468	0.4056	0.3553	0.3192	0.4333	0.4914	0.3281	0.4063	0.3867	0.5455
r'	**12**	**16**	**20**	**11**	**18**	**10**	**12**	**7**	**4**	**2**

TABLE A.13: Ranked *sim* values for $(c \vee \{a, b\})$ in ACB dataset using FVR.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and b	0.7492	0.7819	0.6994	0.8142	0.8011	0.7921	0.7508	0.8877	0.8435	0.6990
	0.7082	0.7360	0.6718	0.7398	0.7499	0.7803	0.7278	0.8605	0.8234	0.6773
	0.6655	0.7171	0.6688	0.7362	0.7377	0.7251	0.6870	0.8073	0.7523	0.6597
	0.6450	0.6945	0.6559	0.7170	0.7245	0.6951	0.6726	0.7914	0.7456	0.6510
	0.6242	0.6775	0.6445	0.6980	0.7030	0.6951	0.6544	0.7702	0.7244	0.6398
	0.5844	0.6696	0.6242	0.6978	0.6908	0.6812	0.6273	0.7366	0.7211	0.6286
	0.5469	0.6514	0.6211	0.6938	0.6852	0.6753	0.6203	0.7341	0.7175	0.6192
	0.5413	0.6465	0.6179	0.6860	0.6770	0.6675	0.6140	0.7190	0.7144	0.5879
	0.5364	0.6223	0.6112	0.6860	0.6752	0.6672	0.6132	0.7160	0.7112	0.5869
	0.5345	0.6201	0.6064	0.6778	0.6621	0.6322	0.6066	0.7047	0.7100	0.5783
	0.5341	0.5865	0.5970	0.6729	0.6606	0.6287	0.6016	0.7033	0.7076	0.5734
	0.5210	0.5808	0.5880	0.6474	0.6273	0.6273	0.5996	0.7008	0.6814	0.5592
	0.5125	0.5776	0.5706	0.6449	0.6237	0.6258	0.5693	0.6989	0.6814	0.5502
	0.5114	0.5592	0.5338	0.6440	0.6233	0.6105	0.5593	0.6946	0.6728	0.5396
	0.5114	0.5587	0.5336	0.6435	0.5985	0.6061	0.5569	0.6840	0.6718	0.5321
	0.5065	0.5433	0.5324	0.6191	0.5917	0.5975	0.5511	0.6820	0.6668	0.5314
	0.4999	0.5429	0.5294	0.6052	0.5898	0.5974	0.5446	0.6764	0.6529	0.5241
	0.4961	0.5386	0.5226	0.6051	0.5809	0.5930	0.5414	0.6710	0.6395	0.5204
	0.4961	0.5291	0.5135	0.6010	0.5715	0.5926	0.5241	0.6682	0.6391	0.5037
	0.4935	0.5279	0.5014	0.5982	0.5604	0.5829	0.5239	0.6678	0.6365	0.4986
	0.4853	0.5235	0.5013	0.5847	0.5604	0.5787	0.5194	0.6667	0.6365	0.4901
	0.4734	0.5132	0.4998	0.5782	0.5576	0.5688	0.5122	0.6536	0.6264	0.4826
	0.4723	0.5017	0.4872	0.5734	0.5333	0.5620	0.5064	0.6494	0.6218	0.4815
	0.4700	0.4940	0.4872	0.5638	0.5325	0.5615	0.5062	0.6494	0.6216	0.4775
	0.4695	0.4811	0.4862	0.5528	0.5272	0.5458	0.4975	0.6379	0.6159	0.4709
	0.4665	0.4754	0.4597	0.5349	0.5262	0.5255	0.4970	0.6309	0.6060	0.4709
	0.4655	0.4360	0.4486	0.5116	0.5163	0.5217	0.4593	0.6122	0.5995	0.4530
	0.4521	0.4271	0.4461	0.5049	0.5105	0.5054	0.4405	0.5878	0.5886	0.4372
	0.3795	0.4271	0.4146	0.4953	0.4812	0.4727	0.4240	0.5862	0.5696	0.4246
	0.3109	0.4089	0.3289	0.4777	0.3795	0.3938	0.4240	0.5853	0.5278	0.3211
r'	**8**	**13**	**15**	**19**	**7**	**8**	**4**	**1**	**5**	**19**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and b	0.7455	0.8508	0.6893	0.5369	0.6962	0.7134	0.8160	0.7134	0.6324	0.6335
	0.7228	0.7713	0.6874	0.5168	0.6568	0.6925	0.7854	0.6925	0.6316	0.5784
	0.7224	0.7560	0.6872	0.5140	0.6468	0.6430	0.7535	0.6430	0.6061	0.5767
	0.6973	0.7257	0.6708	0.4944	0.6039	0.6192	0.7254	0.6192	0.6000	0.5598
	0.6850	0.7069	0.6294	0.4941	0.5941	0.6142	0.7094	0.6142	0.6000	0.5385
	0.6823	0.6994	0.6113	0.4899	0.5890	0.6116	0.6753	0.6116	0.5904	0.5383
	0.6698	0.6899	0.6003	0.4895	0.5851	0.6014	0.6744	0.6014	0.5716	0.5244
	0.6696	0.6859	0.5998	0.4859	0.5573	0.6014	0.6727	0.6014	0.5596	0.5169
	0.6521	0.6831	0.5955	0.4751	0.5402	0.6008	0.6727	0.6008	0.5578	0.5162
	0.6431	0.6773	0.5913	0.4682	0.5378	0.5911	0.6606	0.5911	0.5537	0.5131
	0.6320	0.6744	0.5910	0.4617	0.5241	0.5899	0.6564	0.5899	0.5340	0.4931
	0.6268	0.6560	0.5866	0.4606	0.5240	0.5639	0.6450	0.5639	0.5334	0.4923
	0.6243	0.6523	0.5754	0.4592	0.5152	0.5545	0.6449	0.5545	0.5318	0.4917
	0.6057	0.6487	0.5675	0.4486	0.5151	0.5515	0.6436	0.5515	0.5137	0.4806
	0.5954	0.6365	0.5574	0.4147	0.5111	0.5462	0.6385	0.5462	0.5001	0.4700
	0.5771	0.6360	0.5421	0.4019	0.4993	0.5316	0.6327	0.5316	0.4978	0.4656
	0.5721	0.6339	0.5351	0.3890	0.4991	0.5266	0.6071	0.5266	0.4903	0.4460
	0.5699	0.6300	0.5229	0.3811	0.4857	0.5264	0.6068	0.5264	0.4838	0.4460
	0.5687	0.6237	0.5213	0.3811	0.4857	0.5240	0.6048	0.5240	0.4775	0.4349
	0.5660	0.6234	0.5052	0.3775	0.4822	0.5201	0.6010	0.5201	0.4763	0.4345
	0.5660	0.6036	0.4810	0.3703	0.4613	0.5135	0.5961	0.5135	0.4678	0.4298
	0.5649	0.5957	0.4810	0.3703	0.4450	0.5077	0.5718	0.5077	0.4643	0.4281
	0.5596	0.5927	0.4735	0.3666	0.4364	0.5065	0.5717	0.5065	0.4432	0.4166
	0.5459	0.5879	0.4728	0.3643	0.4286	0.4900	0.5449	0.4900	0.4307	0.4160
	0.5453	0.5834	0.4697	0.3558	0.4151	0.4830	0.5408	0.4830	0.4277	0.4041
	0.5359	0.5741	0.4631	0.3445	0.4046	0.4771	0.5383	0.4771	0.4251	0.3982
	0.5168	0.5687	0.4574	0.3344	0.4046	0.4765	0.5254	0.4765	0.3946	0.3879
	0.4839	0.5509	0.4462	0.3131	0.3976	0.4648	0.5243	0.4648	0.3775	0.3659
	0.4790	0.5268	0.4439	0.3119	0.3964	0.4427	0.5072	0.4427	0.3660	0.3607
	0.4610	0.5268	0.4297	0.2879	0.3502	0.4400	0.4887	0.4400	0.3229	0.3606
r'	**8**	**14**	**20**	**10**	**28**	**12**	**21**	**8**	**8**	**23**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and b	0.7200	0.7189	0.7830	0.6295	0.7346	0.7748	0.7055	0.7760	0.6387	0.7759
	0.6844	0.6556	0.7308	0.6014	0.6726	0.7402	0.6565	0.6889	0.6188	0.7607
	0.6506	0.6153	0.7270	0.5914	0.6289	0.7266	0.6501	0.6785	0.6122	0.7443
	0.6385	0.6024	0.7239	0.5665	0.6152	0.7020	0.6476	0.6783	0.5958	0.7419
	0.6269	0.5901	0.7001	0.5546	0.6121	0.6852	0.6446	0.6750	0.5870	0.7163
	0.6108	0.5443	0.6637	0.5517	0.6050	0.6650	0.6379	0.6210	0.5865	0.7150
	0.6065	0.5351	0.6600	0.5459	0.6020	0.6551	0.6206	0.6134	0.5471	0.7052
	0.6008	0.4971	0.6562	0.5345	0.5782	0.6505	0.6199	0.5988	0.5015	0.6956
	0.5635	0.4934	0.6453	0.5233	0.5483	0.6462	0.6038	0.5950	0.4850	0.6794
	0.5602	0.4921	0.6397	0.5083	0.5481	0.6452	0.5932	0.5827	0.4708	0.6748
	0.5532	0.4815	0.6344	0.5039	0.5361	0.6434	0.5796	0.5826	0.4680	0.6563
	0.5353	0.4805	0.6296	0.5027	0.5343	0.6344	0.5628	0.5781	0.4645	0.6513
	0.5294	0.4732	0.6144	0.5002	0.5338	0.6206	0.5236	0.5767	0.4559	0.6421
	0.5163	0.4614	0.6122	0.4869	0.5279	0.6116	0.5208	0.5737	0.4514	0.6278
	0.5150	0.4607	0.6113	0.4664	0.5264	0.5942	0.5132	0.5634	0.4443	0.6266
	0.5118	0.4605	0.5897	0.4610	0.5243	0.5844	0.5081	0.5589	0.4436	0.6099
	0.5013	0.4564	0.5852	0.4589	0.5242	0.5602	0.5034	0.5508	0.4205	0.6079
	0.4989	0.4551	0.5831	0.4509	0.5220	0.5524	0.5034	0.5483	0.4169	0.6004
	0.4966	0.4460	0.5755	0.4464	0.5177	0.5517	0.4953	0.5480	0.4169	0.5984
	0.4936	0.4460	0.5708	0.4298	0.4937	0.5479	0.4902	0.5374	0.4141	0.5912
	0.4756	0.4359	0.5672	0.4228	0.4745	0.5350	0.4895	0.5297	0.3990	0.5895
	0.4539	0.4347	0.5672	0.4141	0.4646	0.5350	0.4886	0.5290	0.3930	0.5786
	0.4537	0.4338	0.5638	0.4133	0.4630	0.5344	0.4812	0.5185	0.3930	0.5747
	0.4452	0.4290	0.5584	0.4093	0.4618	0.5264	0.4752	0.5113	0.3914	0.5673
	0.4446	0.4205	0.5566	0.3942	0.4559	0.5263	0.4748	0.4985	0.3721	0.5599
	0.4154	0.4143	0.5509	0.3842	0.4460	0.5252	0.4626	0.4917	0.3626	0.5599
	0.4084	0.3896	0.5425	0.3801	0.4440	0.5127	0.4505	0.4817	0.3383	0.5577
	0.4084	0.3873	0.5320	0.3591	0.4241	0.5061	0.4282	0.4817	0.3122	0.5252
	0.4069	0.3466	0.5246	0.3537	0.4241	0.4733	0.4273	0.4640	0.3085	0.4889
	0.4039	0.3394	0.4839	0.3537	0.4195	0.4700	0.3548	0.4560	0.2935	0.4197
r'	**12**	**18**	**21**	**10**	**17**	**12**	**13**	**9**	**5**	**3**

TABLE A.14: Ranked *sim* values for $(a \vee \{b, c\})$ in GP dataset using U-KTS $\mathcal{F}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group b and c	0.0325	0.0311	0.0373	0.0440	0.0572	0.0466	0.0383	0.0426	0.0490	0.0389
	0.0358	0.0340	0.0398	0.0445	0.0604	0.0519	0.0430	0.0444	0.0502	0.0405
	0.0376	0.0341	0.0445	0.0462	0.0619	0.0534	0.0473	0.0455	0.0503	0.0407
	0.0386	0.0347	0.0455	0.0468	0.0628	0.0537	0.0496	0.0455	0.0504	0.0435
	0.0404	0.0355	0.0459	0.0471	0.0628	0.0545	0.0498	0.0468	0.0522	0.0436
	0.0409	0.0366	0.0475	0.0474	0.0628	0.0561	0.0502	0.0488	0.0522	0.0443
	0.0413	0.0367	0.0476	0.0489	0.0647	0.0579	0.0513	0.0492	0.0524	0.0448
	0.0423	0.0383	0.0476	0.0497	0.0667	0.0604	0.0530	0.0494	0.0533	0.0449
	0.0448	0.0388	0.0492	0.0504	0.0689	0.0618	0.0531	0.0499	0.0536	0.0466
	0.0459	0.0390	0.0493	0.0507	0.0707	0.0649	0.0531	0.0500	0.0544	0.0469
	0.0472	0.0399	0.0493	0.0520	0.0717	0.0649	0.0551	0.0503	0.0554	0.0480
	0.0492	0.0423	0.0509	0.0523	0.0726	0.0667	0.0557	0.0513	0.0557	0.0510
	0.0503	0.0434	0.0521	0.0538	0.0735	0.0692	0.0560	0.0517	0.0562	0.0512
	0.0504	0.0437	0.0523	0.0552	0.0737	0.0693	0.0574	0.0525	0.0563	0.0523
	0.0507	0.0437	0.0527	0.0552	0.0744	0.0699	0.0584	0.0529	0.0567	0.0543
	0.0514	0.0440	0.0537	0.0553	0.0752	0.0702	0.0612	0.0547	0.0574	0.0560
	0.0524	0.0467	0.0549	0.0579	0.0760	0.0708	0.0640	0.0559	0.0576	0.0561
	0.0529	0.0484	0.0562	0.0581	0.0766	0.0728	0.0645	0.0561	0.0582	0.0576
	0.0537	0.0487	0.0586	0.0583	0.0766	0.0750	0.0655	0.0564	0.0586	0.0577
	0.0567	0.0526	0.0591	0.0586	0.0771	0.0755	0.0655	0.0564	0.0604	0.0578
	0.0625	0.0528	0.0604	0.0588	0.0777	0.0767	0.0674	0.0581	0.0619	0.0597
	0.0634	0.0532	0.0618	0.0588	0.0783	0.0786	0.0678	0.0582	0.0633	0.0598
	0.0644	0.0565	0.0625	0.0589	0.0786	0.0823	0.0712	0.0596	0.0634	0.0619
	0.0674	0.0576	0.0636	0.0603	0.0791	0.0831	0.0717	0.0598	0.0661	0.0632
	0.0698	0.0600	0.0639	0.0609	0.0823	0.0831	0.0766	0.0602	0.0687	0.0634
	0.0709	0.0638	0.0687	0.0619	0.0852	0.0908	0.0769	0.0620	0.0688	0.0643
	0.0716	0.0641	0.0703	0.0645	0.0857	0.0919	0.0801	0.0629	0.0699	0.0679
	0.0780	0.0664	0.0705	0.0645	0.0940	0.0929	0.0829	0.0673	0.0702	0.0763
	0.0808	0.0750	0.0714	0.0682	0.0955	0.0956	0.0892	0.0679	0.0711	0.0838
	0.0947	0.0819	0.0863	0.0695	0.0973	0.0999	0.0962	0.0750	0.0793	0.0844
	0.0957	0.0896	0.0952	0.0743	0.0986	0.1087	0.0985	0.0753	0.0802	0.0858
r'	**1**	**1**	**2**	**3**	**2**	**3**	**2**	**5**	**4**	**3**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group b and c	0.0311	0.0763	0.0618	0.0432	0.0458	0.0330	0.0418	0.0491	0.0257	0.0378
	0.0389	0.0822	0.0630	0.0457	0.0488	0.0365	0.0421	0.0587	0.0288	0.0412
	0.0390	0.0855	0.0657	0.0486	0.0498	0.0390	0.0433	0.0600	0.0349	0.0450
	0.0421	0.0867	0.0672	0.0507	0.0508	0.0395	0.0442	0.0615	0.0398	0.0485
	0.0426	0.0877	0.0705	0.0522	0.0509	0.0412	0.0455	0.0619	0.0405	0.0487
	0.0427	0.0879	0.0715	0.0534	0.0517	0.0422	0.0459	0.0622	0.0446	0.0494
	0.0434	0.0896	0.0721	0.0538	0.0529	0.0425	0.0484	0.0625	0.0453	0.0496
	0.0441	0.0906	0.0731	0.0538	0.0536	0.0439	0.0486	0.0657	0.0473	0.0513
	0.0446	0.0920	0.0752	0.0546	0.0540	0.0455	0.0505	0.0673	0.0477	0.0514
	0.0464	0.0937	0.0762	0.0548	0.0545	0.0468	0.0508	0.0682	0.0485	0.0523
	0.0467	0.0952	0.0767	0.0549	0.0564	0.0480	0.0512	0.0682	0.0494	0.0537
	0.0472	0.0960	0.0781	0.0558	0.0566	0.0484	0.0537	0.0684	0.0497	0.0538
	0.0478	0.0973	0.0783	0.0561	0.0569	0.0488	0.0540	0.0685	0.0504	0.0544
	0.0480	0.0973	0.0784	0.0564	0.0573	0.0511	0.0541	0.0689	0.0505	0.0548
	0.0506	0.0977	0.0791	0.0566	0.0576	0.0512	0.0548	0.0693	0.0513	0.0550
	0.0516	0.0986	0.0797	0.0573	0.0577	0.0515	0.0549	0.0697	0.0516	0.0557
	0.0518	0.0994	0.0811	0.0574	0.0582	0.0517	0.0553	0.0698	0.0518	0.0562
	0.0534	0.1004	0.0812	0.0578	0.0588	0.0521	0.0575	0.0699	0.0521	0.0567
	0.0544	0.1012	0.0813	0.0581	0.0593	0.0527	0.0576	0.0711	0.0534	0.0573
	0.0552	0.1020	0.0822	0.0588	0.0609	0.0535	0.0588	0.0713	0.0537	0.0574
	0.0559	0.1021	0.0839	0.0595	0.0612	0.0537	0.0595	0.0750	0.0539	0.0576
	0.0570	0.1048	0.0843	0.0597	0.0643	0.0564	0.0600	0.0759	0.0549	0.0583
	0.0574	0.1063	0.0848	0.0604	0.0644	0.0628	0.0602	0.0773	0.0553	0.0599
	0.0587	0.1104	0.0850	0.0624	0.0654	0.0652	0.0624	0.0783	0.0585	0.0611
	0.0591	0.1118	0.0876	0.0630	0.0657	0.0657	0.0627	0.0806	0.0589	0.0615
	0.0609	0.1123	0.0876	0.0637	0.0658	0.0668	0.0629	0.0808	0.0596	0.0616
	0.0633	0.1143	0.0886	0.0647	0.0683	0.0682	0.0630	0.0816	0.0619	0.0621
	0.0689	0.1195	0.0929	0.0688	0.0695	0.0736	0.0642	0.0817	0.0642	0.0701
	0.0694	0.1235	0.0930	0.0700	0.0706	0.0808	0.0659	0.0820	0.0678	0.0713
	0.0725	0.1241	0.0982	0.0866	0.0725	0.0884	0.0826	0.0852	0.0725	0.0901
	0.0781	0.1242	0.1036	0.0891	0.0808	0.0946	0.0832	0.0860	0.0800	0.0934
r'	**2**	**8**	**5**	**2**	**7**	**1**	**4**	**1**	**4**	**1**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}	S_{31}
All samples in the group b and c	0.0314	0.0437	0.0623	0.0643	0.0557	0.0550	0.0311	0.0362	0.0376	0.0377	0.0271
	0.0356	0.0444	0.0643	0.0667	0.0566	0.0551	0.0330	0.0389	0.0380	0.0390	0.0373
	0.0369	0.0474	0.0658	0.0680	0.0572	0.0557	0.0365	0.0390	0.0400	0.0395	0.0374
	0.0370	0.0476	0.0659	0.0684	0.0588	0.0563	0.0371	0.0405	0.0402	0.0444	0.0377
	0.0386	0.0478	0.0673	0.0704	0.0597	0.0598	0.0385	0.0407	0.0404	0.0452	0.0382
	0.0393	0.0481	0.0683	0.0723	0.0613	0.0630	0.0400	0.0408	0.0406	0.0458	0.0391
	0.0405	0.0483	0.0702	0.0726	0.0615	0.0652	0.0412	0.0408	0.0412	0.0473	0.0426
	0.0407	0.0499	0.0710	0.0733	0.0620	0.0655	0.0427	0.0424	0.0425	0.0479	0.0450
	0.0413	0.0509	0.0712	0.0740	0.0621	0.0663	0.0434	0.0424	0.0434	0.0495	0.0452
	0.0416	0.0512	0.0712	0.0745	0.0626	0.0678	0.0441	0.0426	0.0487	0.0497	0.0454
	0.0421	0.0516	0.0718	0.0754	0.0641	0.0684	0.0458	0.0429	0.0491	0.0500	0.0459
	0.0426	0.0517	0.0727	0.0758	0.0651	0.0698	0.0470	0.0434	0.0494	0.0510	0.0462
	0.0430	0.0522	0.0732	0.0759	0.0660	0.0715	0.0480	0.0439	0.0497	0.0511	0.0479
	0.0432	0.0559	0.0734	0.0775	0.0670	0.0718	0.0492	0.0455	0.0499	0.0512	0.0483
	0.0435	0.0569	0.0736	0.0777	0.0679	0.0726	0.0497	0.0460	0.0506	0.0522	0.0490
	0.0458	0.0575	0.0739	0.0782	0.0681	0.0731	0.0511	0.0460	0.0513	0.0551	0.0496
	0.0504	0.0588	0.0743	0.0788	0.0700	0.0741	0.0513	0.0484	0.0513	0.0551	0.0506
	0.0518	0.0593	0.0750	0.0793	0.0702	0.0741	0.0515	0.0489	0.0524	0.0553	0.0510
	0.0536	0.0598	0.0752	0.0799	0.0704	0.0747	0.0516	0.0500	0.0556	0.0555	0.0515
	0.0594	0.0604	0.0779	0.0799	0.0718	0.0762	0.0525	0.0508	0.0574	0.0556	0.0528
	0.0596	0.0635	0.0793	0.0803	0.0721	0.0792	0.0543	0.0515	0.0596	0.0558	0.0535
	0.0625	0.0683	0.0793	0.0808	0.0724	0.0810	0.0552	0.0535	0.0608	0.0568	0.0583
	0.0633	0.0698	0.0797	0.0822	0.0738	0.0814	0.0558	0.0555	0.0623	0.0571	0.0608
	0.0688	0.0716	0.0802	0.0833	0.0748	0.0814	0.0561	0.0569	0.0625	0.0574	0.0609
	0.0696	0.0721	0.0806	0.0848	0.0772	0.0819	0.0577	0.0586	0.0659	0.0576	0.0616
	0.0702	0.0728	0.0816	0.0859	0.0790	0.0828	0.0579	0.0608	0.0676	0.0582	0.0647
	0.0705	0.0746	0.0825	0.0875	0.0823	0.0830	0.0611	0.0624	0.0693	0.0604	0.0692
	0.0721	0.0748	0.0864	0.0878	0.0836	0.0847	0.0624	0.0629	0.0751	0.0616	0.0743
	0.0876	0.0849	0.0887	0.0898	0.0845	0.0892	0.0720	0.0708	0.0791	0.0672	0.0807
	0.0942	0.0874	0.0907	0.0918	0.0853	0.0903	0.0843	0.0794	0.0879	0.0778	0.0890
	0.1000	0.0982	0.0943	0.1001	0.0862	0.0932	0.0951	0.0801	0.0890	0.0813	0.0997
r'	**5**	**2**	**1**	**8**	**6**	**2**	**4**	**1**	**2**	**2**	**1**

Table A.15: Ranked *sim* values for $(b \vee \{a, b\})$ in GP dataset using U-KTS $\mathcal{F}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and c	0.0428	0.0409	0.0524	0.0664	0.1065	0.0664	0.0587	0.0840	0.0810	0.0351
	0.0432	0.0423	0.0553	0.0666	0.1066	0.0682	0.0592	0.0888	0.0815	0.0362
	0.0443	0.0429	0.0554	0.0670	0.1071	0.0697	0.0602	0.0891	0.0816	0.0367
	0.0447	0.0430	0.0557	0.0687	0.1086	0.0722	0.0612	0.0894	0.0838	0.0369
	0.0449	0.0433	0.0562	0.0691	0.1116	0.0726	0.0613	0.0902	0.0840	0.0371
	0.0455	0.0434	0.0582	0.0699	0.1117	0.0728	0.0618	0.0911	0.0841	0.0378
	0.0460	0.0448	0.0587	0.0702	0.1124	0.0742	0.0622	0.0924	0.0843	0.0380
	0.0472	0.0449	0.0587	0.0717	0.1127	0.0743	0.0622	0.0926	0.0851	0.0389
	0.0475	0.0450	0.0598	0.0720	0.1141	0.0764	0.0622	0.0927	0.0855	0.0397
	0.0475	0.0453	0.0599	0.0726	0.1144	0.0765	0.0625	0.0928	0.0856	0.0400
	0.0476	0.0455	0.0600	0.0727	0.1149	0.0765	0.0633	0.0942	0.0857	0.0406
	0.0477	0.0455	0.0601	0.0727	0.1159	0.0769	0.0638	0.0944	0.0860	0.0410
	0.0478	0.0457	0.0602	0.0736	0.1163	0.0773	0.0644	0.0944	0.0865	0.0415
	0.0478	0.0461	0.0602	0.0736	0.1167	0.0784	0.0645	0.0947	0.0871	0.0420
	0.0479	0.0465	0.0604	0.0737	0.1171	0.0792	0.0647	0.0950	0.0877	0.0422
	0.0482	0.0468	0.0608	0.0739	0.1200	0.0793	0.0650	0.0951	0.0879	0.0425
	0.0482	0.0475	0.0614	0.0742	0.1200	0.0795	0.0651	0.0955	0.0879	0.0426
	0.0487	0.0490	0.0620	0.0747	0.1213	0.0799	0.0652	0.0960	0.0884	0.0427
	0.0497	0.0492	0.0623	0.0748	0.1220	0.0799	0.0654	0.0967	0.0895	0.0430
	0.0513	0.0493	0.0623	0.0749	0.1222	0.0801	0.0655	0.0967	0.0899	0.0447
	0.0522	0.0496	0.0624	0.0750	0.1227	0.0808	0.0656	0.0974	0.0904	0.0451
	0.0523	0.0498	0.0627	0.0750	0.1250	0.0810	0.0659	0.0976	0.0905	0.0463
	0.0537	0.0501	0.0632	0.0751	0.1251	0.0816	0.0659	0.0985	0.0911	0.0463
	0.0551	0.0513	0.0644	0.0753	0.1263	0.0827	0.0661	0.0989	0.0912	0.0508
	0.0561	0.0517	0.0679	0.0778	0.1281	0.0828	0.0664	0.0990	0.0921	0.0513
	0.0573	0.0517	0.0680	0.0780	0.1282	0.0844	0.0670	0.0996	0.0922	0.0534
	0.0578	0.0537	0.0683	0.0783	0.1307	0.0849	0.0686	0.1003	0.0953	0.0553
	0.0585	0.0540	0.0710	0.0790	0.1312	0.0883	0.0696	0.1007	0.0955	0.0567
	0.0622	0.0545	0.0726	0.0790	0.1317	0.0886	0.0718	0.1008	0.0956	0.0598
	0.0622	0.0554	0.0744	0.0798	0.1321	0.0887	0.0724	0.1011	0.0991	0.0626
	0.0636	0.0566	0.0750	0.0808	0.1326	0.0923	0.0738	0.1046	0.1010	0.0673
r'	**2**	**2**	**2**	**2**	**3**	**3**	**1**	**4**	**1**	**3**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and c	0.0791	0.1027	0.1080	0.0731	0.0697	0.0677	0.0508	0.1048	0.0634	0.0602
	0.0793	0.1042	0.1086	0.0747	0.0719	0.0712	0.0539	0.1063	0.0648	0.0609
	0.0795	0.1057	0.1091	0.0758	0.0724	0.0718	0.0557	0.1065	0.0680	0.0616
	0.0797	0.1070	0.1113	0.0758	0.0729	0.0724	0.0558	0.1068	0.0706	0.0631
	0.0798	0.1079	0.1118	0.0759	0.0733	0.0726	0.0563	0.1088	0.0709	0.0641
	0.0807	0.1079	0.1148	0.0767	0.0735	0.0728	0.0582	0.1097	0.0713	0.0645
	0.0807	0.1081	0.1182	0.0786	0.0737	0.0729	0.0584	0.1111	0.0715	0.0656
	0.0815	0.1096	0.1187	0.0793	0.0742	0.0730	0.0587	0.1115	0.0716	0.0660
	0.0816	0.1097	0.1187	0.0804	0.0746	0.0733	0.0604	0.1123	0.0718	0.0669
	0.0822	0.1104	0.1188	0.0805	0.0749	0.0740	0.0613	0.1124	0.0720	0.0679
	0.0828	0.1116	0.1195	0.0815	0.0758	0.0745	0.0613	0.1132	0.0725	0.0684
	0.0839	0.1118	0.1207	0.0816	0.0763	0.0746	0.0624	0.1149	0.0735	0.0686
	0.0842	0.1139	0.1212	0.0819	0.0765	0.0746	0.0626	0.1156	0.0735	0.0688
	0.0867	0.1144	0.1219	0.0825	0.0767	0.0748	0.0627	0.1161	0.0740	0.0688
	0.0871	0.1166	0.1223	0.0833	0.0769	0.0751	0.0636	0.1166	0.0746	0.0690
	0.0879	0.1177	0.1232	0.0834	0.0775	0.0752	0.0643	0.1166	0.0757	0.0691
	0.0884	0.1181	0.1233	0.0837	0.0781	0.0755	0.0646	0.1171	0.0761	0.0696
	0.0886	0.1192	0.1237	0.0840	0.0784	0.0757	0.0650	0.1172	0.0766	0.0698
	0.0889	0.1194	0.1239	0.0844	0.0785	0.0758	0.0651	0.1180	0.0767	0.0703
	0.0894	0.1196	0.1251	0.0845	0.0786	0.0771	0.0652	0.1186	0.0768	0.0710
	0.0898	0.1199	0.1261	0.0852	0.0787	0.0773	0.0652	0.1197	0.0774	0.0713
	0.0913	0.1207	0.1279	0.0854	0.0790	0.0774	0.0661	0.1212	0.0775	0.0714
	0.0920	0.1234	0.1292	0.0858	0.0792	0.0779	0.0663	0.1216	0.0777	0.0714
	0.0930	0.1235	0.1306	0.0865	0.0799	0.0787	0.0667	0.1217	0.0780	0.0718
	0.0944	0.1245	0.1319	0.0872	0.0801	0.0794	0.0669	0.1225	0.0800	0.0731
	0.0966	0.1252	0.1340	0.0875	0.0801	0.0804	0.0676	0.1225	0.0804	0.0747
	0.0968	0.1265	0.1342	0.0875	0.0806	0.0806	0.0679	0.1234	0.0805	0.0759
	0.0983	0.1265	0.1346	0.0880	0.0813	0.0806	0.0690	0.1245	0.0813	0.0759
	0.1022	0.1300	0.1362	0.0891	0.0818	0.0807	0.0696	0.1251	0.0815	0.0769
	0.1045	0.1348	0.1373	0.0903	0.0830	0.0811	0.0712	0.1273	0.0830	0.0771
	0.1046	0.1363	0.1379	0.0909	0.0836	0.0812	0.0713	0.1309	0.0864	0.0801
r'	**2**	**4**	**3**	**4**	**5**	**1**	**5**	**4**	**3**	**2**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}	S_{31}
All samples in the group a and c	0.0498	0.0545	0.0876	0.0897	0.0614	0.0951	0.0521	0.0500	0.0707	0.0565	0.0660
	0.0500	0.0599	0.0926	0.0906	0.0630	0.0982	0.0578	0.0501	0.0722	0.0594	0.0672
	0.0515	0.0630	0.0928	0.0908	0.0637	0.1024	0.0579	0.0508	0.0727	0.0602	0.0674
	0.0516	0.0632	0.0937	0.0928	0.0639	0.1028	0.0580	0.0515	0.0738	0.0614	0.0677
	0.0528	0.0644	0.0938	0.0930	0.0643	0.1029	0.0587	0.0542	0.0740	0.0617	0.0681
	0.0528	0.0654	0.0940	0.0936	0.0648	0.1030	0.0590	0.0543	0.0742	0.0629	0.0707
	0.0537	0.0667	0.0942	0.0937	0.0651	0.1036	0.0595	0.0544	0.0743	0.0636	0.0710
	0.0538	0.0668	0.0958	0.0943	0.0655	0.1052	0.0597	0.0545	0.0744	0.0639	0.0716
	0.0546	0.0670	0.0964	0.0944	0.0656	0.1053	0.0600	0.0551	0.0756	0.0642	0.0719
	0.0550	0.0677	0.0968	0.0954	0.0662	0.1057	0.0601	0.0552	0.0756	0.0655	0.0719
	0.0551	0.0681	0.0968	0.0961	0.0664	0.1060	0.0606	0.0561	0.0759	0.0658	0.0722
	0.0551	0.0683	0.0971	0.0971	0.0666	0.1065	0.0608	0.0562	0.0760	0.0660	0.0723
	0.0551	0.0692	0.0986	0.0979	0.0669	0.1104	0.0610	0.0565	0.0763	0.0662	0.0725
	0.0553	0.0693	0.0987	0.0980	0.0670	0.1126	0.0611	0.0566	0.0774	0.0663	0.0734
	0.0560	0.0710	0.0994	0.0990	0.0671	0.1126	0.0611	0.0571	0.0774	0.0666	0.0734
	0.0564	0.0710	0.0998	0.0993	0.0679	0.1137	0.0620	0.0573	0.0784	0.0671	0.0748
	0.0565	0.0718	0.1017	0.0994	0.0679	0.1141	0.0622	0.0575	0.0792	0.0671	0.0749
	0.0567	0.0724	0.1019	0.0998	0.0680	0.1141	0.0624	0.0579	0.0794	0.0676	0.0751
	0.0568	0.0730	0.1020	0.1005	0.0680	0.1145	0.0626	0.0589	0.0798	0.0677	0.0752
	0.0571	0.0731	0.1028	0.1006	0.0683	0.1160	0.0628	0.0592	0.0799	0.0691	0.0756
	0.0578	0.0733	0.1035	0.1014	0.0684	0.1164	0.0631	0.0593	0.0807	0.0695	0.0759
	0.0583	0.0735	0.1051	0.1021	0.0693	0.1174	0.0641	0.0595	0.0815	0.0700	0.0772
	0.0590	0.0741	0.1057	0.1037	0.0694	0.1177	0.0647	0.0597	0.0818	0.0705	0.0774
	0.0599	0.0741	0.1069	0.1040	0.0696	0.1183	0.0649	0.0602	0.0831	0.0706	0.0782
	0.0600	0.0743	0.1069	0.1051	0.0702	0.1184	0.0651	0.0602	0.0835	0.0715	0.0782
	0.0601	0.0755	0.1085	0.1062	0.0708	0.1188	0.0655	0.0629	0.0846	0.0725	0.0782
	0.0605	0.0757	0.1106	0.1068	0.0715	0.1190	0.0673	0.0635	0.0846	0.0726	0.0788
	0.0609	0.0769	0.1108	0.1080	0.0724	0.1206	0.0677	0.0647	0.0852	0.0729	0.0796
	0.0622	0.0774	0.1111	0.1083	0.0729	0.1240	0.0677	0.0686	0.0859	0.0731	0.0803
	0.0645	0.0776	0.1117	0.1086	0.0734	0.1254	0.0682	0.0714	0.0883	0.0733	0.0804
	0.0693	0.0793	0.1151	0.1139	0.0737	0.1336	0.0685	0.0751	0.0886	0.0774	0.0811
r'	**4**	**1**	**4**	**8**	**3**	**2**	**4**	**1**	**4**	**1**	**1**

Table A.16: Ranked *sim* values for $(c \vee \{a, b\})$ in GP dataset using U-KTS $\mathcal{F}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and b	0.0402	0.0423	0.0531	0.0619	0.0999	0.0668	0.0582	0.0828	0.0794	0.0342
	0.0426	0.0425	0.0540	0.0655	0.1044	0.0684	0.0597	0.0881	0.0795	0.0354
	0.0455	0.0436	0.0546	0.0664	0.1071	0.0724	0.0606	0.0884	0.0796	0.0368
	0.0456	0.0439	0.0559	0.0671	0.1085	0.0736	0.0613	0.0904	0.0825	0.0370
	0.0456	0.0439	0.0568	0.0673	0.1085	0.0740	0.0626	0.0906	0.0833	0.0380
	0.0463	0.0441	0.0591	0.0683	0.1092	0.0742	0.0630	0.0911	0.0842	0.0389
	0.0470	0.0442	0.0592	0.0688	0.1105	0.0748	0.0630	0.0911	0.0846	0.0389
	0.0472	0.0444	0.0597	0.0692	0.1107	0.0749	0.0632	0.0912	0.0847	0.0391
	0.0473	0.0463	0.0598	0.0694	0.1134	0.0752	0.0632	0.0912	0.0850	0.0396
	0.0475	0.0466	0.0601	0.0712	0.1146	0.0757	0.0633	0.0916	0.0851	0.0397
	0.0485	0.0467	0.0615	0.0712	0.1152	0.0759	0.0633	0.0917	0.0852	0.0401
	0.0485	0.0467	0.0619	0.0712	0.1155	0.0766	0.0634	0.0918	0.0861	0.0414
	0.0486	0.0472	0.0620	0.0713	0.1159	0.0766	0.0634	0.0920	0.0862	0.0416
	0.0491	0.0473	0.0620	0.0716	0.1171	0.0780	0.0638	0.0932	0.0866	0.0418
	0.0491	0.0475	0.0622	0.0721	0.1189	0.0783	0.0638	0.0933	0.0875	0.0423
	0.0492	0.0475	0.0624	0.0724	0.1193	0.0788	0.0642	0.0934	0.0878	0.0425
	0.0495	0.0476	0.0630	0.0735	0.1201	0.0791	0.0644	0.0937	0.0883	0.0425
	0.0503	0.0481	0.0634	0.0736	0.1209	0.0796	0.0645	0.0939	0.0883	0.0427
	0.0506	0.0487	0.0639	0.0737	0.1224	0.0803	0.0647	0.0949	0.0884	0.0429
	0.0511	0.0489	0.0641	0.0742	0.1226	0.0805	0.0649	0.0954	0.0893	0.0429
	0.0518	0.0497	0.0644	0.0744	0.1232	0.0818	0.0659	0.0964	0.0896	0.0449
	0.0518	0.0497	0.0645	0.0748	0.1235	0.0839	0.0662	0.0968	0.0898	0.0455
	0.0528	0.0507	0.0655	0.0749	0.1238	0.0840	0.0674	0.0974	0.0898	0.0469
	0.0532	0.0512	0.0660	0.0763	0.1240	0.0843	0.0684	0.0975	0.0913	0.0486
	0.0548	0.0525	0.0663	0.0776	0.1245	0.0843	0.0688	0.0976	0.0924	0.0505
	0.0552	0.0526	0.0667	0.0778	0.1253	0.0850	0.0695	0.0977	0.0926	0.0513
	0.0553	0.0532	0.0674	0.0792	0.1296	0.0866	0.0702	0.0999	0.0928	0.0518
	0.0560	0.0536	0.0678	0.0794	0.1323	0.0867	0.0709	0.1009	0.0937	0.0520
	0.0575	0.0544	0.0684	0.0795	0.1326	0.0878	0.0709	0.1010	0.0939	0.0550
	0.0608	0.0560	0.0735	0.0801	0.1342	0.0911	0.0719	0.1032	0.0965	0.0628
	0.0636	0.0572	0.0763	0.0834	0.1398	0.0926	0.0749	0.1063	0.0971	0.0659
r'	**2**	**1**	**1**	**2**	**2**	**3**	**1**	**4**	**2**	**3**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and b	0.0739	0.1001	0.1067	0.0762	0.0680	0.0603	0.0563	0.1028	0.0660	0.0566
	0.0759	0.1030	0.1079	0.0765	0.0704	0.0664	0.0570	0.1067	0.0670	0.0624
	0.0763	0.1030	0.1123	0.0766	0.0706	0.0711	0.0580	0.1084	0.0676	0.0624
	0.0778	0.1064	0.1132	0.0781	0.0708	0.0723	0.0584	0.1086	0.0679	0.0636
	0.0782	0.1075	0.1138	0.0782	0.0714	0.0728	0.0587	0.1093	0.0680	0.0638
	0.0785	0.1077	0.1143	0.0784	0.0717	0.0729	0.0596	0.1099	0.0684	0.0640
	0.0804	0.1088	0.1144	0.0795	0.0719	0.0732	0.0598	0.1106	0.0708	0.0644
	0.0810	0.1096	0.1154	0.0797	0.0723	0.0733	0.0614	0.1118	0.0710	0.0654
	0.0813	0.1108	0.1166	0.0802	0.0724	0.0735	0.0619	0.1123	0.0718	0.0654
	0.0823	0.1108	0.1172	0.0803	0.0741	0.0737	0.0622	0.1134	0.0719	0.0661
	0.0828	0.1113	0.1174	0.0804	0.0747	0.0740	0.0622	0.1150	0.0733	0.0664
	0.0840	0.1133	0.1184	0.0810	0.0763	0.0740	0.0626	0.1151	0.0745	0.0669
	0.0857	0.1149	0.1221	0.0814	0.0763	0.0740	0.0630	0.1152	0.0747	0.0678
	0.0868	0.1153	0.1228	0.0826	0.0771	0.0745	0.0648	0.1154	0.0756	0.0690
	0.0872	0.1171	0.1230	0.0827	0.0772	0.0750	0.0650	0.1169	0.0770	0.0690
	0.0874	0.1174	0.1237	0.0828	0.0783	0.0750	0.0651	0.1174	0.0770	0.0695
	0.0877	0.1176	0.1243	0.0832	0.0786	0.0751	0.0664	0.1175	0.0771	0.0696
	0.0898	0.1177	0.1257	0.0836	0.0787	0.0757	0.0665	0.1185	0.0772	0.0704
	0.0900	0.1179	0.1257	0.0839	0.0788	0.0759	0.0665	0.1193	0.0774	0.0713
	0.0903	0.1190	0.1262	0.0849	0.0790	0.0761	0.0666	0.1201	0.0776	0.0730
	0.0905	0.1198	0.1280	0.0851	0.0791	0.0766	0.0668	0.1208	0.0779	0.0734
	0.0906	0.1213	0.1282	0.0854	0.0793	0.0772	0.0674	0.1210	0.0785	0.0736
	0.0908	0.1223	0.1291	0.0854	0.0794	0.0775	0.0691	0.1218	0.0787	0.0737
	0.0912	0.1225	0.1314	0.0858	0.0796	0.0781	0.0696	0.1219	0.0790	0.0742
	0.0932	0.1236	0.1316	0.0866	0.0797	0.0787	0.0707	0.1226	0.0795	0.0747
	0.0950	0.1252	0.1322	0.0881	0.0800	0.0788	0.0710	0.1236	0.0795	0.0747
	0.0968	0.1262	0.1324	0.0889	0.0804	0.0788	0.0712	0.1239	0.0808	0.0760
	0.0983	0.1263	0.1328	0.0891	0.0805	0.0792	0.0715	0.1264	0.0817	0.0771
	0.1021	0.1311	0.1356	0.0898	0.0806	0.0820	0.0740	0.1274	0.0832	0.0806
	0.1024	0.1331	0.1386	0.0901	0.0842	0.0826	0.0746	0.1304	0.0851	0.0846
	0.1076	0.1370	0.1470	0.0934	0.0859	0.0846	0.0789	0.1321	0.0899	0.0859
r'	**2**	**6**	**5**	**4**	**6**	**1**	**5**	**3**	**3**	**1**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}	S_{31}
All samples in the group a and b	0.0491	0.0628	0.0880	0.0877	0.0608	0.0976	0.0559	0.0513	0.0673	0.0611	0.0687
	0.0500	0.0638	0.0914	0.0909	0.0610	0.0977	0.0560	0.0527	0.0721	0.0613	0.0690
	0.0502	0.0642	0.0926	0.0913	0.0627	0.1023	0.0568	0.0530	0.0731	0.0613	0.0693
	0.0515	0.0642	0.0941	0.0922	0.0632	0.1030	0.0569	0.0532	0.0735	0.0620	0.0699
	0.0516	0.0652	0.0945	0.0932	0.0633	0.1057	0.0571	0.0540	0.0736	0.0623	0.0701
	0.0530	0.0663	0.0946	0.0945	0.0636	0.1059	0.0571	0.0541	0.0743	0.0626	0.0707
	0.0533	0.0664	0.0951	0.0946	0.0640	0.1063	0.0576	0.0545	0.0762	0.0638	0.0708
	0.0537	0.0673	0.0951	0.0947	0.0645	0.1066	0.0576	0.0550	0.0770	0.0645	0.0710
	0.0541	0.0677	0.0953	0.0952	0.0654	0.1076	0.0578	0.0552	0.0770	0.0649	0.0714
	0.0541	0.0684	0.0964	0.0958	0.0656	0.1081	0.0579	0.0555	0.0774	0.0649	0.0716
	0.0545	0.0684	0.0980	0.0975	0.0657	0.1082	0.0593	0.0564	0.0774	0.0650	0.0727
	0.0553	0.0686	0.0980	0.0981	0.0663	0.1107	0.0600	0.0565	0.0775	0.0650	0.0728
	0.0557	0.0696	0.0984	0.0983	0.0663	0.1111	0.0603	0.0566	0.0775	0.0652	0.0737
	0.0558	0.0700	0.0995	0.0986	0.0665	0.1113	0.0609	0.0567	0.0776	0.0652	0.0739
	0.0558	0.0710	0.0996	0.0990	0.0666	0.1114	0.0624	0.0576	0.0778	0.0665	0.0742
	0.0560	0.0714	0.1003	0.0993	0.0667	0.1121	0.0626	0.0577	0.0778	0.0673	0.0748
	0.0563	0.0715	0.1005	0.0993	0.0678	0.1124	0.0638	0.0578	0.0779	0.0678	0.0751
	0.0584	0.0722	0.1012	0.0994	0.0682	0.1130	0.0639	0.0583	0.0781	0.0691	0.0759
	0.0587	0.0724	0.1023	0.0997	0.0688	0.1131	0.0641	0.0591	0.0784	0.0698	0.0769
	0.0588	0.0725	0.1034	0.0998	0.0688	0.1142	0.0643	0.0593	0.0791	0.0705	0.0769
	0.0591	0.0728	0.1035	0.1018	0.0689	0.1161	0.0645	0.0595	0.0793	0.0710	0.0783
	0.0592	0.0729	0.1038	0.1023	0.0692	0.1168	0.0646	0.0598	0.0807	0.0713	0.0783
	0.0592	0.0744	0.1044	0.1024	0.0697	0.1175	0.0655	0.0600	0.0808	0.0718	0.0784
	0.0604	0.0749	0.1048	0.1030	0.0704	0.1184	0.0655	0.0607	0.0809	0.0720	0.0784
	0.0608	0.0756	0.1066	0.1038	0.0714	0.1187	0.0658	0.0612	0.0813	0.0721	0.0793
	0.0610	0.0787	0.1073	0.1048	0.0719	0.1200	0.0665	0.0614	0.0821	0.0725	0.0793
	0.0615	0.0788	0.1096	0.1051	0.0723	0.1211	0.0672	0.0623	0.0831	0.0727	0.0796
	0.0621	0.0789	0.1098	0.1058	0.0723	0.1224	0.0685	0.0660	0.0841	0.0738	0.0803
	0.0621	0.0795	0.1102	0.1103	0.0723	0.1246	0.0715	0.0668	0.0848	0.0747	0.0806
	0.0672	0.0810	0.1123	0.1144	0.0760	0.1283	0.0716	0.0670	0.0872	0.0772	0.0813
	0.0677	0.0877	0.1154	0.1179	0.0760	0.1345	0.0737	0.0747	0.0909	0.0814	0.0847
r'	**3**	**2**	**1**	**7**	**5**	**2**	**4**	**2**	**1**	**3**	**1**

TABLE A.17: Ranked *sim* values for $(a \vee \{b, c\})$ in GP dataset using FVR.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group b and c	0.6507	0.7093	0.6458	0.7854	0.7037	0.6907	0.7341	0.6695	0.7345	0.8100
	0.6472	0.6099	0.6228	0.7662	0.6659	0.6784	0.6488	0.6142	0.7236	0.6611
	0.6230	0.6035	0.5996	0.7446	0.6551	0.6179	0.6440	0.5967	0.6576	0.6551
	0.6182	0.6034	0.5985	0.7175	0.6424	0.6164	0.6232	0.5959	0.6494	0.6540
	0.6083	0.5827	0.5816	0.7045	0.6403	0.6058	0.6197	0.5816	0.6416	0.6497
	0.5874	0.5800	0.5814	0.6752	0.6307	0.6024	0.6133	0.5802	0.6331	0.6320
	0.5627	0.5691	0.5803	0.6729	0.6023	0.6003	0.6055	0.5680	0.6104	0.6235
	0.5570	0.5582	0.5625	0.6710	0.5928	0.5794	0.5931	0.5674	0.6070	0.6124
	0.5558	0.5576	0.5470	0.6685	0.5925	0.5686	0.5907	0.5661	0.6064	0.6121
	0.5329	0.5570	0.5459	0.6675	0.5844	0.5659	0.5748	0.5643	0.6003	0.5836
	0.5222	0.5487	0.5434	0.6667	0.5706	0.5653	0.5685	0.5632	0.5907	0.5831
	0.5202	0.5384	0.5196	0.6655	0.5634	0.5566	0.5644	0.5591	0.5871	0.5802
	0.5187	0.5326	0.5039	0.6613	0.5629	0.5477	0.5391	0.5591	0.5858	0.5780
	0.5180	0.5308	0.4958	0.6486	0.5595	0.5466	0.5303	0.5385	0.5669	0.5709
	0.5134	0.5243	0.4866	0.6471	0.5551	0.5428	0.5221	0.5365	0.5603	0.5693
	0.5131	0.5215	0.4852	0.6308	0.5501	0.5403	0.5151	0.5296	0.5586	0.5610
	0.5126	0.5045	0.4697	0.6302	0.5460	0.5375	0.5028	0.5259	0.5465	0.5375
	0.5044	0.5023	0.4597	0.6267	0.5320	0.5222	0.4895	0.5237	0.5439	0.5372
	0.5035	0.4969	0.4489	0.6259	0.5279	0.5200	0.4785	0.5184	0.5429	0.5356
	0.4990	0.4837	0.4446	0.6251	0.5207	0.5193	0.4782	0.5180	0.5416	0.5069
	0.4959	0.4826	0.4438	0.6108	0.5185	0.5187	0.4764	0.5170	0.5354	0.5044
	0.4952	0.4782	0.4438	0.6073	0.5172	0.5068	0.4670	0.5170	0.5319	0.4977
	0.4837	0.4685	0.4436	0.6069	0.4957	0.5039	0.4371	0.5088	0.5311	0.4908
	0.4742	0.4618	0.4429	0.6025	0.4939	0.4929	0.4353	0.4986	0.5168	0.4876
	0.4709	0.4488	0.4412	0.5977	0.4872	0.4905	0.4350	0.4873	0.5160	0.4821
	0.4691	0.4480	0.4409	0.5880	0.4845	0.4834	0.4173	0.4847	0.5050	0.4775
	0.4674	0.4382	0.4203	0.5758	0.4618	0.4491	0.4149	0.4703	0.4846	0.4759
	0.4494	0.4333	0.4123	0.5696	0.4544	0.4435	0.4121	0.4680	0.4832	0.4695
	0.4322	0.4176	0.4114	0.5679	0.4531	0.4260	0.4076	0.4571	0.4793	0.4602
	0.4239	0.4013	0.3859	0.5674	0.4104	0.4180	0.4047	0.4294	0.4698	0.4320
	0.4149	0.3777	0.3596	0.5558	0.3731	0.4053	0.4012	0.3965	0.4333	0.3973
r'	**24**	**11**	**4**	**9**	**19**	**3**	**15**	**23**	**7**	**9**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group b and c	0.7526	0.6270	0.7064	0.7030	0.8217	0.6963	0.7275	0.7139	0.7492	0.7192
	0.6929	0.5955	0.6416	0.6767	0.7981	0.6878	0.6816	0.6802	0.6245	0.6989
	0.6873	0.5834	0.6031	0.6626	0.7736	0.6775	0.6772	0.6465	0.6039	0.6846
	0.6834	0.5784	0.5928	0.6157	0.7033	0.6774	0.6539	0.6461	0.5976	0.6422
	0.6717	0.5745	0.5703	0.5891	0.7014	0.6688	0.6468	0.6446	0.5905	0.6281
	0.6626	0.5509	0.5618	0.5880	0.6878	0.6429	0.6430	0.6329	0.5887	0.6256
	0.6585	0.5489	0.5589	0.5668	0.6783	0.6249	0.6171	0.6250	0.5645	0.6254
	0.6478	0.5480	0.5588	0.5405	0.6771	0.6201	0.6169	0.6223	0.5608	0.6000
	0.6443	0.5296	0.5505	0.5336	0.6675	0.6172	0.6142	0.6105	0.5585	0.5846
	0.6440	0.5247	0.5473	0.5283	0.6669	0.6130	0.6111	0.6086	0.5582	0.5683
	0.6414	0.5240	0.5327	0.5235	0.6491	0.6092	0.6092	0.6060	0.5525	0.5676
	0.6407	0.5147	0.5308	0.5219	0.6405	0.5885	0.6068	0.6032	0.5515	0.5618
	0.6300	0.5035	0.5261	0.5219	0.6349	0.5871	0.6016	0.5992	0.5396	0.5560
	0.6266	0.5016	0.5109	0.5152	0.6196	0.5737	0.5762	0.5847	0.5335	0.5556
	0.6219	0.4979	0.5102	0.5150	0.6129	0.5709	0.5730	0.5830	0.5254	0.5518
	0.6215	0.4974	0.5055	0.5088	0.6003	0.5604	0.5695	0.5792	0.4993	0.5498
	0.6166	0.4923	0.4894	0.4965	0.5995	0.5603	0.5555	0.5782	0.4955	0.5467
	0.6147	0.4857	0.4888	0.4904	0.5982	0.5570	0.5432	0.5674	0.4946	0.5451
	0.6098	0.4854	0.4838	0.4793	0.5832	0.5569	0.5432	0.5618	0.4931	0.5416
	0.5911	0.4734	0.4822	0.4772	0.5831	0.5539	0.5392	0.5519	0.4863	0.5320
	0.5855	0.4705	0.4756	0.4765	0.5810	0.5378	0.5374	0.5486	0.4770	0.5286
	0.5838	0.4698	0.4735	0.4719	0.5747	0.5341	0.5371	0.5328	0.4639	0.5265
	0.5751	0.4616	0.4719	0.4705	0.5726	0.5156	0.5237	0.5305	0.4546	0.5242
	0.5728	0.4607	0.4697	0.4704	0.5658	0.5143	0.5217	0.5109	0.4239	0.5090
	0.5712	0.4363	0.4568	0.4610	0.5509	0.5115	0.5185	0.5107	0.4180	0.5038
	0.5611	0.4228	0.4558	0.4587	0.5491	0.5089	0.5159	0.5048	0.4113	0.4989
	0.5488	0.4179	0.4546	0.4281	0.5401	0.4822	0.5134	0.5020	0.4066	0.4937
	0.5354	0.4174	0.4496	0.4167	0.5304	0.4757	0.4858	0.4825	0.3975	0.4885
	0.5132	0.3967	0.4483	0.4099	0.5259	0.4693	0.4839	0.4822	0.3846	0.4522
	0.5057	0.3871	0.4200	0.4028	0.5093	0.4588	0.4575	0.4609	0.3818	0.4296
	0.5031	0.3753	0.4050	0.3821	0.5059	0.4373	0.4496	0.4503	0.3671	0.4271
r'	**12**	**1**	**2**	**13**	**13**	**12**	**5**	**12**	**10**	**16**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}	S_{31}
All samples in the group b and c	0.6610	0.6912	0.6917	0.7121	0.7494	0.7290	0.7506	0.7530	0.6617	0.6999	0.8437
	0.6402	0.6656	0.6831	0.6463	0.6938	0.6677	0.7238	0.7098	0.6448	0.6784	0.8258
	0.6240	0.6302	0.6564	0.6220	0.6802	0.6271	0.7155	0.7008	0.6063	0.6611	0.8160
	0.6093	0.6234	0.6346	0.6111	0.6438	0.6112	0.7131	0.6916	0.6004	0.6557	0.7605
	0.5974	0.5891	0.6170	0.6035	0.6104	0.6104	0.6787	0.6544	0.5799	0.6269	0.7566
	0.5936	0.5688	0.6047	0.5862	0.5991	0.5679	0.6554	0.6528	0.5414	0.5771	0.7523
	0.5782	0.5639	0.6008	0.5809	0.5974	0.5659	0.6429	0.6429	0.5368	0.5587	0.7517
	0.5635	0.5551	0.5807	0.5779	0.5910	0.5638	0.6429	0.6405	0.5312	0.5584	0.7496
	0.5595	0.5512	0.5802	0.5720	0.5667	0.5590	0.6419	0.6268	0.5260	0.5444	0.7484
	0.5497	0.5366	0.5790	0.5670	0.5643	0.5556	0.6347	0.6262	0.5236	0.5152	0.7357
	0.5493	0.5263	0.5769	0.5638	0.5442	0.5543	0.6326	0.6205	0.5184	0.5046	0.7227
	0.5483	0.5246	0.5740	0.5530	0.5426	0.5496	0.6211	0.6176	0.5142	0.5025	0.7219
	0.5380	0.5142	0.5652	0.5486	0.5409	0.5478	0.6192	0.6146	0.5105	0.4974	0.7196
	0.5351	0.5140	0.5275	0.5383	0.5277	0.5447	0.6142	0.6067	0.5054	0.4927	0.7055
	0.5338	0.5125	0.5184	0.5351	0.5258	0.5405	0.6106	0.6061	0.4994	0.4912	0.6985
	0.5168	0.5040	0.5148	0.5325	0.5219	0.5338	0.5918	0.6029	0.4958	0.4724	0.6873
	0.5148	0.5035	0.5122	0.5307	0.5139	0.5280	0.5772	0.6001	0.4898	0.4641	0.6843
	0.5136	0.5022	0.5068	0.5259	0.5120	0.5118	0.5745	0.5850	0.4812	0.4385	0.6758
	0.5104	0.4995	0.5058	0.5129	0.5027	0.5016	0.5505	0.5810	0.4764	0.4322	0.6732
	0.5099	0.4974	0.5012	0.5122	0.4961	0.4997	0.5421	0.5757	0.4753	0.4282	0.6695
	0.5090	0.4871	0.4895	0.5059	0.4953	0.4970	0.5357	0.5738	0.4702	0.4247	0.6525
	0.5046	0.4847	0.4869	0.4801	0.4880	0.4954	0.5141	0.5703	0.4587	0.4236	0.6515
	0.5039	0.4759	0.4599	0.4800	0.4636	0.4941	0.5040	0.5692	0.4582	0.4210	0.6494
	0.4971	0.4351	0.4555	0.4727	0.4617	0.4900	0.5009	0.5691	0.4359	0.4188	0.6323
	0.4832	0.4267	0.4487	0.4703	0.4442	0.4718	0.4987	0.5685	0.4356	0.4164	0.6171
	0.4732	0.4220	0.4371	0.4594	0.4437	0.4689	0.4874	0.5617	0.4351	0.3847	0.6155
	0.4645	0.4210	0.4258	0.4567	0.4366	0.4681	0.4641	0.5100	0.4321	0.3815	0.6145
	0.4620	0.4101	0.4207	0.4450	0.4359	0.4561	0.4593	0.5047	0.4302	0.3802	0.6141
	0.4537	0.4089	0.4184	0.4209	0.4321	0.4387	0.4591	0.4988	0.4189	0.3779	0.6085
	0.4328	0.4070	0.4127	0.4052	0.4193	0.4351	0.4417	0.4818	0.4097	0.3597	0.6074
	0.4327	0.3767	0.3600	0.3423	0.4078	0.4176	0.4400	0.4716	0.3949	0.3381	0.5680
r'	**12**	**14**	**18**	**3**	**5**	**8**	**1**	**3**	**11**	**22**	**1**

Table A.18: Ranked *sim* values for $(b \vee \{a, b\})$ in GP dataset using FVR.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and c	0.7103	0.7198	0.8280	0.6657	0.6801	0.7302	0.8288	0.6885	0.7204	0.7919
	0.6860	0.6781	0.8001	0.6485	0.6592	0.7018	0.8170	0.6872	0.7193	0.7576
	0.6470	0.6771	0.7796	0.6086	0.6228	0.6995	0.7476	0.6813	0.7069	0.7487
	0.6297	0.6752	0.7756	0.5948	0.5898	0.6701	0.7435	0.6620	0.6885	0.7324
	0.6248	0.6579	0.7714	0.5767	0.5813	0.6630	0.7118	0.6546	0.6665	0.7241
	0.6092	0.6137	0.7474	0.5651	0.5692	0.6576	0.6934	0.6463	0.6403	0.6932
	0.6085	0.6115	0.7316	0.5497	0.5638	0.6570	0.6890	0.6271	0.6251	0.6866
	0.5830	0.6017	0.7224	0.5469	0.5560	0.6482	0.6756	0.6223	0.6219	0.6814
	0.5827	0.5914	0.7108	0.5379	0.5445	0.6473	0.6746	0.6179	0.5961	0.6807
	0.5661	0.5887	0.7072	0.5344	0.5373	0.6339	0.6744	0.6102	0.5854	0.6787
	0.5605	0.5863	0.6943	0.5336	0.5152	0.6336	0.6685	0.6070	0.5777	0.6694
	0.5597	0.5832	0.6914	0.5268	0.5103	0.6168	0.6681	0.5704	0.5774	0.6665
	0.5530	0.5708	0.6811	0.5225	0.4999	0.6041	0.6630	0.5668	0.5739	0.6633
	0.5438	0.5699	0.6687	0.5217	0.4980	0.5971	0.6596	0.5522	0.5718	0.6594
	0.5436	0.5646	0.6635	0.5152	0.4926	0.5947	0.6594	0.5518	0.5661	0.6581
	0.5423	0.5549	0.6612	0.5089	0.4926	0.5916	0.6510	0.5503	0.5630	0.6556
	0.5362	0.5540	0.6511	0.5034	0.4913	0.5900	0.6465	0.5451	0.5517	0.6539
	0.5339	0.5447	0.6508	0.5025	0.4853	0.5795	0.6427	0.5423	0.5481	0.6529
	0.5316	0.5320	0.6373	0.5020	0.4852	0.5754	0.6340	0.5374	0.5393	0.6441
	0.5194	0.5278	0.6340	0.5012	0.4841	0.5648	0.6283	0.5132	0.5267	0.6415
	0.5056	0.5278	0.6317	0.4929	0.4708	0.5636	0.6240	0.5121	0.5242	0.6309
	0.5001	0.5203	0.6280	0.4854	0.4610	0.5629	0.6183	0.5050	0.5212	0.6271
	0.4996	0.5105	0.6221	0.4646	0.4595	0.5510	0.6081	0.5021	0.5210	0.6250
	0.4993	0.5077	0.6164	0.4628	0.4553	0.5276	0.6022	0.5013	0.5192	0.6180
	0.4864	0.4926	0.6111	0.4469	0.4527	0.5239	0.5961	0.4667	0.5001	0.6031
	0.4834	0.4898	0.6050	0.4357	0.4155	0.5077	0.5868	0.4647	0.4994	0.5959
	0.4822	0.4892	0.6045	0.4334	0.4140	0.5055	0.5705	0.4638	0.4929	0.5577
	0.4784	0.4837	0.6020	0.4299	0.3998	0.5038	0.5501	0.4570	0.4920	0.5559
	0.4571	0.4786	0.5680	0.4235	0.3965	0.4805	0.4997	0.4492	0.4669	0.5349
	0.4213	0.4663	0.5443	0.4185	0.3689	0.4746	0.4980	0.4461	0.4150	0.5182
	0.3992	0.4396	0.5367	0.4121	0.3605	0.4286	0.4852	0.3983	0.4088	0.4902
r'	**20**	**9**	**5**	**8**	**22**	**2**	**14**	**24**	**7**	**9**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and c	0.8654	0.7280	0.6692	0.7149	0.6889	0.7579	0.7389	0.6747	0.7658	0.7462
	0.7773	0.7172	0.5651	0.6791	0.6504	0.7311	0.7107	0.6388	0.7221	0.6599
	0.7672	0.7106	0.5589	0.6400	0.6181	0.7070	0.6999	0.6112	0.6961	0.6585
	0.7549	0.6927	0.5489	0.6310	0.5782	0.6933	0.6997	0.5979	0.6942	0.6578
	0.7354	0.6920	0.5397	0.6230	0.5593	0.6893	0.6633	0.5828	0.6767	0.6406
	0.7238	0.6906	0.5371	0.6183	0.5347	0.6890	0.6602	0.5810	0.6687	0.6304
	0.7173	0.6209	0.5238	0.5739	0.5345	0.6577	0.6487	0.5777	0.6460	0.6148
	0.6984	0.6133	0.5176	0.5585	0.5256	0.6354	0.6436	0.5738	0.6337	0.6122
	0.6984	0.5942	0.5152	0.5532	0.5217	0.6334	0.6320	0.5706	0.6330	0.6036
	0.6828	0.5785	0.5087	0.5470	0.5069	0.6283	0.6291	0.5658	0.6302	0.5877
	0.6729	0.5734	0.5064	0.5350	0.5061	0.6266	0.6194	0.5597	0.6161	0.5860
	0.6672	0.5733	0.4977	0.5340	0.5046	0.6040	0.6122	0.5545	0.6128	0.5860
	0.6617	0.5340	0.4952	0.5245	0.5021	0.6020	0.6093	0.5506	0.6094	0.5844
	0.6435	0.5287	0.4913	0.5236	0.4884	0.6008	0.6075	0.5320	0.6092	0.5644
	0.6365	0.5278	0.4793	0.5168	0.4880	0.5990	0.5989	0.5315	0.6089	0.5612
	0.6357	0.5271	0.4706	0.5124	0.4870	0.5845	0.5964	0.5194	0.6056	0.5579
	0.6260	0.5227	0.4703	0.5105	0.4827	0.5839	0.5933	0.5026	0.6045	0.5567
	0.6188	0.5116	0.4666	0.5062	0.4818	0.5778	0.5808	0.5025	0.5985	0.5542
	0.6141	0.5036	0.4640	0.5049	0.4753	0.5747	0.5632	0.4950	0.5968	0.5430
	0.5839	0.5019	0.4570	0.4981	0.4707	0.5733	0.5613	0.4855	0.5967	0.5332
	0.5722	0.4991	0.4408	0.4869	0.4624	0.5662	0.5581	0.4804	0.5956	0.5241
	0.5699	0.4959	0.4401	0.4758	0.4500	0.5522	0.5460	0.4759	0.5886	0.5230
	0.5623	0.4921	0.4399	0.4727	0.4450	0.5283	0.5424	0.4755	0.5666	0.4873
	0.5603	0.4755	0.4392	0.4624	0.4443	0.5276	0.5365	0.4612	0.5619	0.4770
	0.5559	0.4632	0.4303	0.4497	0.4364	0.5271	0.5182	0.4573	0.5539	0.4692
	0.5532	0.4437	0.4240	0.4382	0.4271	0.5251	0.5108	0.4549	0.5524	0.4588
	0.5435	0.4408	0.4133	0.4330	0.4129	0.4945	0.4982	0.4350	0.5292	0.4282
	0.5275	0.4384	0.4080	0.3905	0.4017	0.4860	0.4798	0.4346	0.5099	0.4279
	0.5205	0.4381	0.3913	0.3898	0.3990	0.4577	0.4730	0.4153	0.4852	0.4222
	0.5141	0.4059	0.3620	0.3889	0.3776	0.4548	0.4600	0.4090	0.4823	0.4172
	0.5012	0.3624	0.3605	0.3680	0.3561	0.4077	0.4315	0.3800	0.4457	0.3448
r'	**14**	**27**	**3**	**12**	**14**	**13**	**6**	**12**	**11**	**17**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}	S_{31}
All samples in the group a and c	0.7003	0.7520	0.7710	0.7549	0.8057	0.7439	0.7437	0.8271	0.6910	0.7271	0.7906
	0.6983	0.7303	0.6816	0.7352	0.7762	0.7116	0.7084	0.7790	0.6347	0.6651	0.7251
	0.6720	0.7176	0.6680	0.7249	0.7270	0.6947	0.6934	0.7656	0.6128	0.6511	0.7167
	0.6352	0.6310	0.6641	0.7219	0.7245	0.6844	0.6788	0.7601	0.5949	0.6495	0.6492
	0.6302	0.6296	0.6246	0.6984	0.7177	0.6574	0.6749	0.7382	0.5861	0.6395	0.6462
	0.6205	0.6200	0.6177	0.6870	0.7137	0.6544	0.6653	0.6986	0.5773	0.6386	0.6427
	0.5987	0.5877	0.6166	0.6677	0.7131	0.6542	0.6241	0.6966	0.5705	0.6226	0.6397
	0.5914	0.5856	0.6130	0.6502	0.7027	0.6537	0.6219	0.6455	0.5659	0.6070	0.6273
	0.5804	0.5817	0.6043	0.6342	0.7006	0.6424	0.5991	0.6333	0.5575	0.6018	0.6268
	0.5619	0.5737	0.6021	0.6330	0.6908	0.6419	0.5788	0.6314	0.5495	0.6007	0.6219
	0.5479	0.5547	0.5929	0.6304	0.6882	0.6411	0.5604	0.6309	0.5442	0.5963	0.6107
	0.5474	0.5480	0.5856	0.6240	0.6846	0.6317	0.5597	0.6303	0.5436	0.5956	0.6077
	0.5357	0.5398	0.5839	0.6174	0.6805	0.6128	0.5566	0.6299	0.5420	0.5924	0.6015
	0.5198	0.5330	0.5720	0.6161	0.6797	0.6123	0.5528	0.6269	0.5397	0.5915	0.5947
	0.5133	0.5122	0.5683	0.6079	0.6757	0.5936	0.5480	0.6140	0.5214	0.5863	0.5885
	0.5052	0.4930	0.5668	0.6069	0.6709	0.5788	0.5389	0.6082	0.5131	0.5773	0.5858
	0.5019	0.4913	0.5634	0.6027	0.6688	0.5776	0.5361	0.6014	0.4946	0.5723	0.5661
	0.4997	0.4886	0.5629	0.5964	0.6665	0.5774	0.5355	0.5994	0.4935	0.5712	0.5602
	0.4966	0.4822	0.5523	0.5903	0.6525	0.5717	0.5269	0.5978	0.4881	0.5679	0.5560
	0.4939	0.4693	0.5513	0.5781	0.6372	0.5657	0.5263	0.5909	0.4786	0.5608	0.5495
	0.4912	0.4548	0.5393	0.5767	0.6367	0.5545	0.5192	0.5843	0.4716	0.5535	0.5447
	0.4864	0.4514	0.5385	0.5764	0.6358	0.5377	0.5155	0.5821	0.4685	0.5507	0.5408
	0.4860	0.4506	0.5374	0.5625	0.6263	0.5357	0.5000	0.5821	0.4483	0.5493	0.5407
	0.4801	0.4496	0.5263	0.5599	0.6196	0.5307	0.4995	0.5790	0.4407	0.5294	0.5394
	0.4792	0.4387	0.5228	0.5597	0.6125	0.5209	0.4895	0.5654	0.4289	0.5215	0.5097
	0.4757	0.4347	0.5171	0.5517	0.6125	0.5164	0.4803	0.5468	0.4147	0.4988	0.5082
	0.4598	0.4326	0.5034	0.5311	0.5892	0.5046	0.4796	0.5452	0.4077	0.4967	0.5046
	0.4515	0.4323	0.4589	0.5252	0.5463	0.4928	0.4390	0.5356	0.4006	0.4583	0.4875
	0.4313	0.4161	0.4586	0.5198	0.5385	0.4857	0.4299	0.5287	0.3767	0.4564	0.4828
	0.4164	0.4155	0.4575	0.5171	0.5341	0.4606	0.4142	0.5271	0.3633	0.4402	0.4667
	0.4039	0.4054	0.4542	0.4697	0.5331	0.4506	0.3839	0.4964	0.3445	0.4181	0.4262
r'	**10**	**14**	**18**	**9**	**2**	**10**	**2**	**3**	**12**	**1**	**1**

Table A.19: Ranked *sim* values for $(c \vee \{a, b\})$ in GP dataset using FVR.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and b	0.7252	0.7342	0.6434	0.8641	0.6639	0.7881	0.6433	0.6023	0.7346	0.6612
	0.6681	0.7033	0.6209	0.8211	0.6512	0.6918	0.6210	0.5948	0.6983	0.6606
	0.6416	0.7013	0.6109	0.8050	0.6510	0.6852	0.5884	0.5790	0.6676	0.6151
	0.6275	0.6535	0.6062	0.7711	0.6133	0.6735	0.5562	0.5660	0.6515	0.6052
	0.6200	0.6480	0.5954	0.7640	0.6101	0.6369	0.5526	0.5556	0.6460	0.5912
	0.6200	0.6418	0.5769	0.7450	0.6055	0.6330	0.5405	0.5367	0.6437	0.5834
	0.6183	0.6227	0.5734	0.7435	0.6009	0.6329	0.5382	0.5264	0.6410	0.5808
	0.6177	0.6198	0.5532	0.7333	0.6000	0.6270	0.5243	0.5237	0.6350	0.5802
	0.6075	0.6046	0.5326	0.7178	0.5926	0.6181	0.5080	0.5208	0.6339	0.5643
	0.6044	0.5989	0.5289	0.7030	0.5917	0.6061	0.4963	0.5102	0.5992	0.5596
	0.6022	0.5966	0.5241	0.6998	0.5804	0.5983	0.4941	0.4991	0.5919	0.5471
	0.5886	0.5887	0.5089	0.6918	0.5663	0.5930	0.4914	0.4981	0.5899	0.5291
	0.5848	0.5846	0.5007	0.6894	0.5655	0.5873	0.4872	0.4974	0.5529	0.5283
	0.5754	0.5843	0.4944	0.6758	0.5596	0.5869	0.4866	0.4917	0.5451	0.5250
	0.5488	0.5813	0.4939	0.6704	0.5544	0.5808	0.4860	0.4862	0.5444	0.5188
	0.5486	0.5752	0.4937	0.6641	0.5534	0.5742	0.4859	0.4849	0.5428	0.5158
	0.5435	0.5707	0.4903	0.6599	0.5522	0.5655	0.4852	0.4769	0.5397	0.5107
	0.5415	0.5679	0.4807	0.6554	0.5519	0.5567	0.4759	0.4603	0.5385	0.5005
	0.5400	0.5615	0.4602	0.6489	0.5513	0.5434	0.4711	0.4578	0.5348	0.4973
	0.5351	0.5546	0.4550	0.6440	0.5489	0.5394	0.4680	0.4554	0.5341	0.4946
	0.5340	0.5525	0.4463	0.6376	0.5431	0.5387	0.4662	0.4536	0.5287	0.4932
	0.5308	0.5437	0.4393	0.6346	0.5380	0.5369	0.4657	0.4474	0.5280	0.4913
	0.5010	0.5398	0.4334	0.6258	0.5348	0.5259	0.4633	0.4373	0.5264	0.4854
	0.4773	0.5128	0.4321	0.6183	0.5327	0.5249	0.4493	0.4300	0.5187	0.4791
	0.4720	0.5008	0.4316	0.6112	0.5270	0.5206	0.4465	0.4257	0.4933	0.4726
	0.4715	0.4983	0.4296	0.6069	0.5038	0.4988	0.4464	0.4241	0.4917	0.4626
	0.4664	0.4886	0.4205	0.5787	0.4688	0.4932	0.4047	0.4225	0.4883	0.4590
	0.4630	0.4856	0.3790	0.5636	0.4637	0.4902	0.3923	0.4073	0.4743	0.4545
	0.4449	0.4616	0.3690	0.5482	0.4601	0.4841	0.3658	0.4022	0.4376	0.4268
	0.4412	0.4587	0.3583	0.5468	0.4063	0.4777	0.3617	0.3924	0.4238	0.4000
	0.4406	0.4483	0.3532	0.4975	0.3927	0.4309	0.3494	0.3815	0.4011	0.3990
r'	**2**	**8**	**6**	**10**	**17**	**4**	**14**	**4**	**7**	**9**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and b	0.7593	0.6675	0.8374	0.7150	0.7800	0.7318	0.7284	0.6880	0.7081	0.7430
	0.7593	0.6594	0.6891	0.6806	0.7736	0.7280	0.7160	0.6593	0.6714	0.7181
	0.7451	0.6444	0.6764	0.6606	0.7611	0.6583	0.6927	0.6507	0.6684	0.7143
	0.7277	0.6421	0.6751	0.6553	0.7414	0.6571	0.6669	0.6507	0.6392	0.7131
	0.7130	0.6367	0.6618	0.6506	0.7215	0.6471	0.6486	0.6369	0.6037	0.6635
	0.6955	0.6187	0.6584	0.6241	0.7207	0.6265	0.6419	0.6334	0.5798	0.6580
	0.6902	0.6022	0.6475	0.6240	0.6975	0.6255	0.6359	0.6246	0.5768	0.6555
	0.6784	0.5708	0.6396	0.5920	0.6817	0.6185	0.6320	0.6142	0.5732	0.6472
	0.6770	0.5682	0.6383	0.5833	0.6772	0.6163	0.6261	0.6129	0.5553	0.6405
	0.6745	0.5650	0.5964	0.5790	0.6719	0.6062	0.6257	0.6047	0.5547	0.6133
	0.6589	0.5525	0.5739	0.5746	0.6621	0.5938	0.6056	0.6024	0.5521	0.6069
	0.6585	0.5470	0.5728	0.5723	0.6619	0.5838	0.6056	0.5913	0.5488	0.5943
	0.6506	0.5462	0.5635	0.5629	0.6565	0.5786	0.6028	0.5696	0.5463	0.5940
	0.6400	0.5322	0.5508	0.5595	0.6534	0.5694	0.6017	0.5629	0.5455	0.5904
	0.6377	0.5224	0.5506	0.5504	0.6512	0.5657	0.5943	0.5577	0.5454	0.5860
	0.6367	0.5175	0.5494	0.5472	0.6384	0.5621	0.5940	0.5512	0.5265	0.5802
	0.6305	0.5121	0.5373	0.5432	0.6343	0.5620	0.5877	0.5502	0.5171	0.5759
	0.6234	0.5080	0.5348	0.5403	0.6267	0.5585	0.5861	0.5445	0.5112	0.5730
	0.6161	0.5076	0.5294	0.5308	0.6027	0.5570	0.5793	0.5434	0.4973	0.5723
	0.6154	0.5014	0.5232	0.5306	0.5993	0.5542	0.5791	0.5408	0.4924	0.5704
	0.6103	0.4987	0.5230	0.5289	0.5982	0.5471	0.5722	0.5297	0.4853	0.5520
	0.6061	0.4959	0.5153	0.5088	0.5881	0.5354	0.5705	0.5217	0.4830	0.5097
	0.5876	0.4923	0.5152	0.4933	0.5866	0.5189	0.5698	0.5202	0.4657	0.5023
	0.5848	0.4803	0.5112	0.4908	0.5857	0.5152	0.5471	0.5040	0.4586	0.4977
	0.5835	0.4795	0.4851	0.4848	0.5808	0.5091	0.5366	0.4986	0.4584	0.4944
	0.5679	0.4791	0.4808	0.4753	0.5647	0.5056	0.4922	0.4867	0.4430	0.4864
	0.5392	0.4608	0.4540	0.4749	0.5557	0.5050	0.4863	0.4780	0.4252	0.4849
	0.5311	0.4534	0.4490	0.4330	0.5535	0.5030	0.4853	0.4694	0.4197	0.4692
	0.5112	0.4406	0.4447	0.4269	0.5460	0.4718	0.4824	0.4459	0.3915	0.4685
	0.5057	0.4177	0.4438	0.4114	0.5070	0.4682	0.4503	0.4395	0.3839	0.4516
	0.4997	0.4115	0.4121	0.4016	0.4944	0.4462	0.4240	0.4189	0.3730	0.4406
r'	**10**	**6**	**2**	**11**	**12**	**11**	**7**	**13**	**14**	**1**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}	S_{31}
All samples in the group a and b	0.6911	0.7528	0.6659	0.6861	0.7032	0.7253	0.7195	0.7921	0.6377	0.7457	0.9135
	0.6780	0.6947	0.6450	0.6843	0.6660	0.7231	0.6904	0.7837	0.6248	0.6978	0.8408
	0.6707	0.6862	0.6406	0.6667	0.6616	0.7106	0.6669	0.7110	0.5992	0.6977	0.8252
	0.6612	0.6709	0.6196	0.6216	0.6598	0.7065	0.6534	0.7039	0.5886	0.5948	0.8057
	0.6486	0.6455	0.6043	0.6107	0.6447	0.6618	0.6459	0.6990	0.5885	0.5909	0.7983
	0.6056	0.6277	0.6037	0.6025	0.6229	0.6574	0.6400	0.6867	0.5858	0.5834	0.7542
	0.6052	0.6173	0.5932	0.6003	0.5796	0.6209	0.6258	0.6792	0.5842	0.5762	0.7519
	0.5974	0.6144	0.5876	0.5916	0.5742	0.6086	0.5908	0.6782	0.5832	0.5748	0.7509
	0.5969	0.6135	0.5860	0.5872	0.5726	0.6060	0.5884	0.6624	0.5790	0.5732	0.7284
	0.5936	0.5519	0.5734	0.5805	0.5695	0.5899	0.5820	0.6586	0.5700	0.5597	0.7231
	0.5856	0.5435	0.5719	0.5729	0.5673	0.5804	0.5768	0.6551	0.5604	0.5585	0.7186
	0.5825	0.5421	0.5674	0.5712	0.5668	0.5766	0.5725	0.6545	0.5585	0.5435	0.7178
	0.5733	0.5411	0.5518	0.5618	0.5630	0.5747	0.5684	0.6443	0.5481	0.5349	0.7146
	0.5661	0.5387	0.5518	0.5552	0.5522	0.5614	0.5646	0.6431	0.5380	0.5337	0.7037
	0.5554	0.5381	0.5320	0.5513	0.5379	0.5571	0.5616	0.6324	0.5375	0.5301	0.6896
	0.5548	0.5288	0.5271	0.5496	0.5366	0.5562	0.5591	0.6139	0.5353	0.4987	0.6893
	0.5480	0.5184	0.5229	0.5487	0.5326	0.5395	0.5562	0.6017	0.5234	0.4903	0.6859
	0.5409	0.5167	0.5130	0.5479	0.5278	0.5304	0.5540	0.5929	0.5229	0.4871	0.6821
	0.5272	0.5133	0.5101	0.5439	0.5209	0.5262	0.5535	0.5923	0.5187	0.4639	0.6815
	0.5133	0.5030	0.5098	0.5384	0.5206	0.5242	0.5438	0.5885	0.5166	0.4638	0.6815
	0.5117	0.4933	0.5078	0.5381	0.5168	0.5185	0.5276	0.5870	0.5156	0.4471	0.6596
	0.5111	0.4922	0.5047	0.5227	0.4932	0.5042	0.5252	0.5855	0.5112	0.4351	0.6474
	0.5049	0.4870	0.4974	0.5159	0.4915	0.5032	0.5125	0.5795	0.5087	0.4306	0.6430
	0.4987	0.4841	0.4930	0.5082	0.4912	0.4768	0.5000	0.5692	0.4971	0.4273	0.6424
	0.4933	0.4825	0.4929	0.5007	0.4880	0.4713	0.4968	0.5668	0.4940	0.4186	0.6354
	0.4910	0.4788	0.4820	0.4999	0.4812	0.4640	0.4935	0.5637	0.4770	0.4173	0.6284
	0.4689	0.4679	0.4701	0.4960	0.4625	0.4590	0.4908	0.5514	0.4702	0.4006	0.6275
	0.4477	0.4555	0.4483	0.4938	0.4596	0.4567	0.4817	0.5434	0.4464	0.3911	0.6112
	0.4357	0.4445	0.4198	0.4644	0.4410	0.4211	0.4764	0.5273	0.4431	0.3894	0.6056
	0.4208	0.4257	0.3844	0.4596	0.4249	0.4182	0.4608	0.5059	0.4229	0.3720	0.5956
	0.4048	0.3711	0.3815	0.4509	0.4214	0.4080	0.4035	0.5030	0.4183	0.3456	0.5427
r'	**11**	**12**	**18**	**6**	**3**	**10**	**1**	**2**	**9**	**3**	**3**

TABLE A.20: Ranked *sim* values for $(a \vee \{b, c\})$ in VHHS dataset using U-KTS $\mathcal{F}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group b and c	0.0104	0.0162	0.0131	0.0168	0.0355	0.0261	0.0312	0.0160	0.0107	0.0188
	0.0131	0.0167	0.0154	0.0173	0.0383	0.0277	0.0328	0.0164	0.0135	0.0203
	0.0147	0.0174	0.0155	0.0177	0.0388	0.0277	0.0332	0.0165	0.0144	0.0209
	0.0149	0.0198	0.0168	0.0185	0.0411	0.0277	0.0333	0.0175	0.0152	0.0210
	0.0163	0.0202	0.0171	0.0189	0.0419	0.0284	0.0340	0.0180	0.0153	0.0221
	0.0165	0.0203	0.0172	0.0201	0.0437	0.0293	0.0340	0.0187	0.0171	0.0224
	0.0165	0.0204	0.0172	0.0203	0.0437	0.0301	0.0343	0.0190	0.0191	0.0229
	0.0166	0.0208	0.0183	0.0206	0.0441	0.0304	0.0345	0.0194	0.0192	0.0230
	0.0180	0.0214	0.0184	0.0206	0.0442	0.0308	0.0356	0.0200	0.0196	0.0231
	0.0184	0.0216	0.0191	0.0206	0.0447	0.0313	0.0362	0.0202	0.0225	0.0238
	0.0201	0.0227	0.0206	0.0213	0.0466	0.0314	0.0363	0.0204	0.0229	0.0243
	0.0204	0.0227	0.0214	0.0218	0.0467	0.0317	0.0379	0.0209	0.0229	0.0245
	0.0205	0.0229	0.0217	0.0218	0.0472	0.0320	0.0395	0.0216	0.0231	0.0248
	0.0209	0.0238	0.0218	0.0219	0.0490	0.0324	0.0397	0.0227	0.0241	0.0258
	0.0210	0.0241	0.0228	0.0224	0.0496	0.0325	0.0398	0.0228	0.0241	0.0261
	0.0223	0.0242	0.0231	0.0227	0.0513	0.0325	0.0399	0.0234	0.0246	0.0263
	0.0235	0.0243	0.0235	0.0227	0.0515	0.0328	0.0399	0.0246	0.0258	0.0269
	0.0237	0.0244	0.0248	0.0228	0.0515	0.0333	0.0402	0.0252	0.0262	0.0269
	0.0242	0.0249	0.0251	0.0252	0.0527	0.0340	0.0409	0.0259	0.0282	0.0274
	0.0243	0.0250	0.0262	0.0255	0.0532	0.0346	0.0411	0.0263	0.0286	0.0277
	0.0252	0.0277	0.0263	0.0257	0.0534	0.0349	0.0413	0.0264	0.0289	0.0283
	0.0254	0.0278	0.0268	0.0260	0.0545	0.0362	0.0414	0.0301	0.0296	0.0284
	0.0275	0.0284	0.0270	0.0262	0.0547	0.0370	0.0416	0.0317	0.0312	0.0284
	0.0282	0.0285	0.0282	0.0263	0.0550	0.0378	0.0422	0.0333	0.0340	0.0298
	0.0315	0.0286	0.0307	0.0270	0.0556	0.0380	0.0430	0.0345	0.0342	0.0299
	0.0335	0.0290	0.0313	0.0307	0.0562	0.0411	0.0434	0.0354	0.0354	0.0305
	0.0341	0.0293	0.0329	0.0318	0.0563	0.0417	0.0435	0.0354	0.0385	0.0306
	0.0347	0.0312	0.0330	0.0339	0.0567	0.0433	0.0437	0.0368	0.0386	0.0319
	0.0356	0.0316	0.0356	0.0360	0.0582	0.0451	0.0439	0.0379	0.0407	0.0332
	0.0375	0.0320	0.0366	0.0360	0.0588	0.0468	0.0440	0.0379	0.0409	0.0340
	0.0382	0.0377	0.0390	0.0379	0.0589	0.0470	0.0445	0.0391	0.0444	0.0343
	0.0457	0.0381	0.0436	0.0424	0.0592	0.0495	0.0451	0.0509	0.0532	0.0369
	0.0507	0.0443	0.0487	0.0497	0.0593	0.0538	0.0460	0.0535	0.0562	0.0373
	0.0539	0.0444	0.0543	0.0524	0.0602	0.0618	0.0478	0.0545	0.0593	0.0388
	0.0588	0.0501	0.0573	0.0541	0.0616	0.0648	0.0511	0.0568	0.0654	0.0442
	0.0638	0.0579	0.0588	0.0588	0.0620	0.0692	0.0531	0.0570	0.0696	0.0497
	0.0655	0.0595	0.0610	0.0613	0.0642	0.0703	0.0589	0.0665	0.0710	0.0573
	0.0749	0.0657	0.0716	0.0705	0.0650	0.0811	0.0642	0.0749	0.0814	0.0587
	0.0816	0.0713	0.0813	0.0785	0.0654	0.0855	0.0748	0.0856	0.0884	0.0675
r'	**1**	**1**	**2**	**3**	**2**	**9**	**2**	**12**	**1**	**2**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group b and c	0.0554	0.0216	0.0337	0.0220	0.0166	0.0219	0.0298	0.0170	0.0309	0.0266
	0.0578	0.0232	0.0343	0.0250	0.0207	0.0220	0.0320	0.0192	0.0340	0.0327
	0.0590	0.0250	0.0350	0.0250	0.0226	0.0229	0.0327	0.0193	0.0343	0.0328
	0.0617	0.0254	0.0365	0.0256	0.0228	0.0253	0.0329	0.0198	0.0377	0.0336
	0.0636	0.0255	0.0366	0.0262	0.0233	0.0260	0.0339	0.0199	0.0389	0.0347
	0.0648	0.0263	0.0374	0.0268	0.0234	0.0269	0.0339	0.0202	0.0406	0.0348
	0.0655	0.0272	0.0381	0.0268	0.0238	0.0284	0.0350	0.0203	0.0407	0.0351
	0.0695	0.0275	0.0384	0.0272	0.0239	0.0288	0.0363	0.0208	0.0420	0.0354
	0.0704	0.0279	0.0392	0.0275	0.0245	0.0291	0.0369	0.0218	0.0421	0.0356
	0.0708	0.0280	0.0393	0.0281	0.0252	0.0295	0.0369	0.0224	0.0423	0.0358
	0.0737	0.0283	0.0398	0.0281	0.0253	0.0298	0.0371	0.0227	0.0424	0.0363
	0.0741	0.0284	0.0399	0.0283	0.0253	0.0298	0.0371	0.0229	0.0424	0.0363
	0.0771	0.0294	0.0405	0.0286	0.0263	0.0299	0.0380	0.0231	0.0425	0.0364
	0.0778	0.0296	0.0407	0.0289	0.0265	0.0301	0.0382	0.0231	0.0433	0.0367
	0.0786	0.0304	0.0407	0.0291	0.0272	0.0305	0.0388	0.0233	0.0439	0.0369
	0.0787	0.0308	0.0416	0.0293	0.0276	0.0305	0.0399	0.0240	0.0439	0.0378
	0.0792	0.0314	0.0419	0.0296	0.0280	0.0306	0.0402	0.0242	0.0443	0.0386
	0.0822	0.0315	0.0421	0.0296	0.0283	0.0309	0.0402	0.0258	0.0446	0.0392
	0.0827	0.0318	0.0422	0.0302	0.0285	0.0311	0.0404	0.0259	0.0447	0.0394
	0.0828	0.0323	0.0424	0.0304	0.0298	0.0312	0.0407	0.0282	0.0449	0.0401
	0.0831	0.0324	0.0424	0.0305	0.0310	0.0315	0.0416	0.0284	0.0452	0.0403
	0.0831	0.0324	0.0424	0.0313	0.0315	0.0319	0.0423	0.0298	0.0454	0.0409
	0.0832	0.0331	0.0429	0.0314	0.0319	0.0319	0.0427	0.0298	0.0455	0.0409
	0.0832	0.0332	0.0438	0.0315	0.0323	0.0327	0.0432	0.0311	0.0461	0.0422
	0.0845	0.0334	0.0448	0.0322	0.0325	0.0331	0.0432	0.0314	0.0462	0.0431
	0.0847	0.0339	0.0450	0.0323	0.0328	0.0333	0.0435	0.0329	0.0463	0.0437
	0.0849	0.0348	0.0458	0.0324	0.0331	0.0336	0.0438	0.0340	0.0463	0.0443
	0.0851	0.0358	0.0459	0.0341	0.0335	0.0347	0.0445	0.0354	0.0463	0.0456
	0.0855	0.0361	0.0463	0.0344	0.0336	0.0351	0.0449	0.0359	0.0466	0.0475
	0.0856	0.0364	0.0468	0.0348	0.0357	0.0355	0.0449	0.0371	0.0467	0.0492
	0.0865	0.0380	0.0478	0.0364	0.0360	0.0356	0.0454	0.0399	0.0473	0.0499
	0.0867	0.0393	0.0480	0.0367	0.0376	0.0361	0.0457	0.0455	0.0487	0.0507
	0.0894	0.0436	0.0486	0.0374	0.0386	0.0408	0.0459	0.0469	0.0495	0.0588
	0.0897	0.0477	0.0491	0.0430	0.0421	0.0419	0.0464	0.0477	0.0508	0.0623
	0.0909	0.0483	0.0500	0.0451	0.0481	0.0421	0.0466	0.0529	0.0521	0.0628
	0.0917	0.0525	0.0512	0.0537	0.0494	0.0465	0.0469	0.0582	0.0523	0.0652
	0.0943	0.0550	0.0514	0.0547	0.0587	0.0508	0.0474	0.0632	0.0531	0.0674
	0.0947	0.0644	0.0536	0.0591	0.0624	0.0553	0.0544	0.0720	0.0548	0.0681
	0.0961	0.0703	0.0593	0.0648	0.0680	0.0661	0.0656	0.0819	0.0596	0.0745
r'	**2**	**3**	**1**	**6**	**2**	**1**	**9**	**3**	**2**	**3**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group b and c	0.0207	0.0202	0.0363	0.0138	0.0143	0.0269	0.0176	0.0199	0.0146	0.0265
	0.0212	0.0203	0.0401	0.0143	0.0145	0.0302	0.0184	0.0207	0.0151	0.0273
	0.0235	0.0204	0.0406	0.0166	0.0155	0.0304	0.0191	0.0216	0.0160	0.0283
	0.0240	0.0210	0.0425	0.0171	0.0155	0.0306	0.0197	0.0223	0.0164	0.0287
	0.0243	0.0216	0.0431	0.0173	0.0164	0.0312	0.0206	0.0228	0.0175	0.0295
	0.0250	0.0216	0.0439	0.0189	0.0164	0.0320	0.0212	0.0236	0.0178	0.0300
	0.0253	0.0222	0.0444	0.0189	0.0165	0.0334	0.0213	0.0236	0.0192	0.0303
	0.0257	0.0227	0.0470	0.0191	0.0166	0.0336	0.0214	0.0238	0.0214	0.0305
	0.0261	0.0233	0.0472	0.0193	0.0171	0.0346	0.0215	0.0244	0.0216	0.0311
	0.0264	0.0234	0.0480	0.0196	0.0172	0.0347	0.0220	0.0251	0.0219	0.0313
	0.0265	0.0235	0.0489	0.0202	0.0178	0.0347	0.0228	0.0255	0.0221	0.0314
	0.0271	0.0244	0.0496	0.0207	0.0196	0.0349	0.0232	0.0257	0.0233	0.0317
	0.0277	0.0245	0.0496	0.0212	0.0197	0.0351	0.0237	0.0264	0.0239	0.0317
	0.0277	0.0248	0.0501	0.0228	0.0202	0.0354	0.0252	0.0268	0.0245	0.0317
	0.0279	0.0255	0.0501	0.0230	0.0205	0.0355	0.0258	0.0269	0.0247	0.0318
	0.0287	0.0256	0.0502	0.0260	0.0207	0.0358	0.0263	0.0272	0.0252	0.0328
	0.0287	0.0257	0.0503	0.0264	0.0207	0.0359	0.0264	0.0276	0.0255	0.0333
	0.0288	0.0267	0.0505	0.0265	0.0216	0.0363	0.0264	0.0278	0.0260	0.0335
	0.0293	0.0270	0.0511	0.0265	0.0224	0.0365	0.0274	0.0280	0.0261	0.0336
	0.0302	0.0273	0.0513	0.0267	0.0233	0.0366	0.0281	0.0281	0.0301	0.0339
	0.0302	0.0273	0.0515	0.0275	0.0260	0.0370	0.0286	0.0289	0.0306	0.0340
	0.0316	0.0275	0.0523	0.0279	0.0261	0.0376	0.0311	0.0290	0.0307	0.0340
	0.0319	0.0291	0.0532	0.0283	0.0262	0.0381	0.0311	0.0291	0.0319	0.0343
	0.0321	0.0294	0.0537	0.0292	0.0267	0.0382	0.0313	0.0295	0.0342	0.0347
	0.0324	0.0294	0.0538	0.0299	0.0281	0.0384	0.0344	0.0298	0.0347	0.0351
	0.0330	0.0307	0.0541	0.0312	0.0320	0.0387	0.0350	0.0302	0.0352	0.0363
	0.0337	0.0317	0.0541	0.0344	0.0329	0.0388	0.0367	0.0309	0.0361	0.0380
	0.0345	0.0341	0.0547	0.0365	0.0337	0.0391	0.0367	0.0332	0.0364	0.0383
	0.0390	0.0361	0.0548	0.0368	0.0346	0.0395	0.0370	0.0332	0.0374	0.0387
	0.0393	0.0370	0.0550	0.0377	0.0347	0.0402	0.0370	0.0339	0.0382	0.0416
	0.0410	0.0385	0.0552	0.0382	0.0355	0.0409	0.0371	0.0352	0.0408	0.0416
	0.0414	0.0405	0.0557	0.0444	0.0444	0.0413	0.0470	0.0370	0.0493	0.0422
	0.0436	0.0425	0.0573	0.0493	0.0485	0.0415	0.0476	0.0413	0.0523	0.0447
	0.0436	0.0453	0.0577	0.0537	0.0489	0.0417	0.0504	0.0416	0.0581	0.0458
	0.0451	0.0491	0.0581	0.0578	0.0545	0.0428	0.0549	0.0466	0.0614	0.0477
	0.0527	0.0538	0.0587	0.0610	0.0585	0.0483	0.0558	0.0475	0.0629	0.0488
	0.0588	0.0563	0.0588	0.0633	0.0622	0.0523	0.0651	0.0550	0.0654	0.0621
	0.0589	0.0668	0.0593	0.0731	0.0719	0.0574	0.0713	0.0597	0.0772	0.0628
	0.0729	0.0716	0.0596	0.0792	0.0797	0.0676	0.0804	0.0718	0.0830	0.0759
r'	**1**	**7**	**2**	**3**	**8**	**7**	**10**	**1**	**3**	**1**

	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}
All samples in the group b and c	0.0306	0.0395	0.0115	0.0222	0.0173	0.0314	0.0276	0.0412	0.0196
	0.0328	0.0438	0.0135	0.0224	0.0184	0.0346	0.0278	0.0460	0.0202
	0.0330	0.0440	0.0145	0.0234	0.0186	0.0355	0.0280	0.0478	0.0206
	0.0343	0.0468	0.0145	0.0254	0.0188	0.0360	0.0284	0.0512	0.0208
	0.0343	0.0477	0.0152	0.0254	0.0196	0.0364	0.0293	0.0524	0.0226
	0.0344	0.0481	0.0177	0.0259	0.0200	0.0375	0.0294	0.0531	0.0230
	0.0348	0.0483	0.0181	0.0263	0.0200	0.0383	0.0297	0.0547	0.0238
	0.0348	0.0490	0.0184	0.0272	0.0202	0.0396	0.0297	0.0552	0.0243
	0.0355	0.0497	0.0199	0.0282	0.0203	0.0397	0.0300	0.0552	0.0246
	0.0374	0.0498	0.0224	0.0284	0.0207	0.0406	0.0301	0.0581	0.0250
	0.0385	0.0505	0.0226	0.0285	0.0226	0.0413	0.0306	0.0582	0.0251
	0.0389	0.0526	0.0234	0.0287	0.0229	0.0418	0.0307	0.0582	0.0252
	0.0391	0.0528	0.0235	0.0287	0.0233	0.0430	0.0310	0.0591	0.0255
	0.0399	0.0537	0.0236	0.0288	0.0236	0.0431	0.0312	0.0601	0.0263
	0.0401	0.0539	0.0241	0.0292	0.0236	0.0432	0.0312	0.0602	0.0263
	0.0408	0.0542	0.0249	0.0299	0.0236	0.0433	0.0316	0.0612	0.0263
	0.0410	0.0546	0.0251	0.0299	0.0244	0.0434	0.0319	0.0620	0.0265
	0.0413	0.0548	0.0258	0.0302	0.0249	0.0435	0.0319	0.0624	0.0266
	0.0422	0.0550	0.0260	0.0302	0.0251	0.0439	0.0320	0.0649	0.0269
	0.0423	0.0551	0.0263	0.0305	0.0255	0.0439	0.0324	0.0652	0.0272
	0.0428	0.0556	0.0303	0.0305	0.0258	0.0441	0.0324	0.0654	0.0272
	0.0434	0.0560	0.0305	0.0307	0.0268	0.0449	0.0325	0.0658	0.0275
	0.0443	0.0571	0.0325	0.0308	0.0273	0.0452	0.0327	0.0658	0.0284
	0.0446	0.0580	0.0331	0.0311	0.0282	0.0453	0.0327	0.0661	0.0284
	0.0450	0.0585	0.0349	0.0313	0.0296	0.0455	0.0333	0.0673	0.0286
	0.0453	0.0586	0.0364	0.0316	0.0323	0.0459	0.0337	0.0678	0.0288
	0.0461	0.0587	0.0375	0.0320	0.0333	0.0459	0.0338	0.0684	0.0324
	0.0468	0.0592	0.0380	0.0325	0.0333	0.0460	0.0339	0.0687	0.0338
	0.0470	0.0597	0.0395	0.0335	0.0338	0.0462	0.0340	0.0688	0.0350
	0.0475	0.0599	0.0406	0.0366	0.0357	0.0464	0.0356	0.0689	0.0353
	0.0478	0.0600	0.0439	0.0368	0.0401	0.0474	0.0364	0.0692	0.0372
	0.0485	0.0608	0.0541	0.0420	0.0402	0.0479	0.0365	0.0692	0.0385
	0.0517	0.0608	0.0568	0.0440	0.0446	0.0480	0.0368	0.0705	0.0406
	0.0528	0.0611	0.0578	0.0493	0.0490	0.0481	0.0400	0.0716	0.0479
	0.0552	0.0641	0.0648	0.0498	0.0527	0.0523	0.0402	0.0716	0.0522
	0.0552	0.0643	0.0650	0.0508	0.0584	0.0528	0.0465	0.0721	0.0522
	0.0555	0.0656	0.0709	0.0555	0.0606	0.0532	0.0536	0.0731	0.0596
	0.0563	0.0660	0.0802	0.0622	0.0683	0.0558	0.0542	0.0755	0.0673
	0.0637	0.0680	0.0889	0.0673	0.0747	0.0664	0.0639	0.0758	0.0755
r'	**6**	**1**	**2**	**1**	**4**	**1**	**3**	**1**	**1**

TABLE A.21: Ranked *sim* values for $(b \vee \{a, c\})$ in VHHS dataset using U-KTS $\mathcal{F}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and c	0.0295	0.0347	0.0338	0.0246	0.0575	0.0451	0.0431	0.0371	0.0306	0.0315
	0.0295	0.0348	0.0356	0.0247	0.0611	0.0465	0.0468	0.0406	0.0322	0.0315
	0.0298	0.0357	0.0365	0.0251	0.0626	0.0484	0.0473	0.0407	0.0328	0.0345
	0.0299	0.0364	0.0367	0.0251	0.0630	0.0493	0.0485	0.0407	0.0329	0.0353
	0.0303	0.0371	0.0369	0.0257	0.0636	0.0495	0.0485	0.0409	0.0329	0.0355
	0.0306	0.0372	0.0385	0.0258	0.0653	0.0504	0.0487	0.0411	0.0331	0.0357
	0.0308	0.0378	0.0387	0.0258	0.0661	0.0511	0.0488	0.0416	0.0335	0.0361
	0.0309	0.0380	0.0389	0.0261	0.0665	0.0512	0.0492	0.0417	0.0339	0.0361
	0.0309	0.0381	0.0391	0.0262	0.0676	0.0519	0.0496	0.0419	0.0343	0.0367
	0.0310	0.0382	0.0399	0.0262	0.0680	0.0533	0.0502	0.0420	0.0345	0.0367
	0.0311	0.0383	0.0401	0.0264	0.0682	0.0534	0.0508	0.0421	0.0348	0.0368
	0.0312	0.0386	0.0402	0.0264	0.0693	0.0536	0.0513	0.0424	0.0355	0.0371
	0.0314	0.0388	0.0407	0.0265	0.0695	0.0539	0.0524	0.0427	0.0355	0.0371
	0.0315	0.0390	0.0407	0.0266	0.0701	0.0546	0.0537	0.0428	0.0358	0.0372
	0.0317	0.0392	0.0409	0.0266	0.0707	0.0547	0.0544	0.0428	0.0361	0.0372
	0.0318	0.0395	0.0409	0.0269	0.0715	0.0558	0.0545	0.0429	0.0367	0.0374
	0.0319	0.0396	0.0411	0.0271	0.0718	0.0563	0.0553	0.0429	0.0368	0.0374
	0.0320	0.0398	0.0414	0.0272	0.0728	0.0564	0.0554	0.0432	0.0368	0.0378
	0.0321	0.0399	0.0415	0.0276	0.0729	0.0565	0.0554	0.0437	0.0369	0.0380
	0.0325	0.0400	0.0423	0.0279	0.0734	0.0569	0.0555	0.0441	0.0373	0.0381
	0.0326	0.0409	0.0424	0.0281	0.0735	0.0575	0.0557	0.0441	0.0374	0.0384
	0.0329	0.0414	0.0428	0.0282	0.0740	0.0576	0.0559	0.0443	0.0380	0.0387
	0.0329	0.0416	0.0429	0.0282	0.0741	0.0577	0.0565	0.0443	0.0385	0.0389
	0.0333	0.0420	0.0429	0.0284	0.0744	0.0578	0.0567	0.0443	0.0385	0.0390
	0.0334	0.0420	0.0433	0.0289	0.0748	0.0580	0.0571	0.0445	0.0386	0.0390
	0.0336	0.0423	0.0435	0.0291	0.0749	0.0581	0.0576	0.0448	0.0390	0.0390
	0.0337	0.0425	0.0436	0.0296	0.0749	0.0582	0.0577	0.0449	0.0391	0.0391
	0.0337	0.0426	0.0437	0.0301	0.0754	0.0583	0.0580	0.0454	0.0391	0.0393
	0.0340	0.0429	0.0441	0.0301	0.0759	0.0583	0.0582	0.0454	0.0393	0.0395
	0.0341	0.0432	0.0444	0.0310	0.0764	0.0588	0.0589	0.0456	0.0393	0.0399
	0.0346	0.0434	0.0448	0.0319	0.0769	0.0604	0.0593	0.0459	0.0394	0.0400
	0.0348	0.0435	0.0451	0.0323	0.0772	0.0610	0.0593	0.0461	0.0394	0.0404
	0.0349	0.0435	0.0452	0.0329	0.0774	0.0614	0.0593	0.0463	0.0401	0.0411
	0.0355	0.0446	0.0453	0.0330	0.0779	0.0626	0.0597	0.0470	0.0402	0.0411
	0.0362	0.0446	0.0453	0.0365	0.0783	0.0627	0.0607	0.0470	0.0405	0.0419
	0.0404	0.0450	0.0458	0.0367	0.0790	0.0632	0.0613	0.0476	0.0406	0.0421
	0.0417	0.0456	0.0462	0.0370	0.0793	0.0634	0.0629	0.0477	0.0414	0.0442
	0.0431	0.0461	0.0473	0.0438	0.0806	0.0634	0.0629	0.0479	0.0433	0.0457
	0.0464	0.0468	0.0481	0.0493	0.0820	0.0647	0.0632	0.0531	0.0446	0.0477
r'	**2**	**1**	**2**	**2**	**3**	**8**	**3**	**10**	**1**	**1**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and c	0.0973	0.0420	0.0494	0.0422	0.0389	0.0330	0.0478	0.0324	0.0486	0.0296
	0.1072	0.0431	0.0616	0.0437	0.0408	0.0348	0.0511	0.0325	0.0526	0.0297
	0.1075	0.0442	0.0620	0.0452	0.0417	0.0357	0.0531	0.0336	0.0564	0.0315
	0.1101	0.0444	0.0624	0.0471	0.0432	0.0358	0.0539	0.0338	0.0588	0.0321
	0.1129	0.0446	0.0641	0.0473	0.0436	0.0358	0.0546	0.0338	0.0589	0.0328
	0.1129	0.0462	0.0644	0.0476	0.0445	0.0367	0.0567	0.0341	0.0600	0.0329
	0.1134	0.0470	0.0648	0.0476	0.0451	0.0369	0.0578	0.0344	0.0601	0.0332
	0.1143	0.0471	0.0655	0.0479	0.0458	0.0374	0.0584	0.0359	0.0603	0.0334
	0.1150	0.0475	0.0675	0.0480	0.0458	0.0374	0.0588	0.0360	0.0604	0.0334
	0.1150	0.0476	0.0676	0.0482	0.0463	0.0376	0.0594	0.0363	0.0607	0.0335
	0.1158	0.0483	0.0684	0.0495	0.0464	0.0376	0.0595	0.0364	0.0623	0.0340
	0.1172	0.0484	0.0688	0.0495	0.0471	0.0382	0.0602	0.0367	0.0647	0.0344
	0.1180	0.0487	0.0701	0.0498	0.0475	0.0383	0.0603	0.0368	0.0651	0.0344
	0.1198	0.0487	0.0708	0.0498	0.0488	0.0385	0.0610	0.0370	0.0655	0.0345
	0.1207	0.0488	0.0710	0.0499	0.0491	0.0386	0.0629	0.0373	0.0659	0.0348
	0.1216	0.0491	0.0711	0.0499	0.0491	0.0388	0.0630	0.0373	0.0671	0.0356
	0.1217	0.0492	0.0716	0.0511	0.0499	0.0389	0.0630	0.0374	0.0673	0.0359
	0.1220	0.0501	0.0720	0.0511	0.0501	0.0389	0.0630	0.0377	0.0675	0.0364
	0.1222	0.0503	0.0730	0.0513	0.0501	0.0392	0.0632	0.0378	0.0676	0.0368
	0.1228	0.0509	0.0734	0.0514	0.0504	0.0394	0.0637	0.0379	0.0678	0.0370
	0.1229	0.0512	0.0734	0.0524	0.0507	0.0396	0.0644	0.0380	0.0679	0.0370
	0.1231	0.0512	0.0736	0.0525	0.0520	0.0396	0.0644	0.0382	0.0682	0.0372
	0.1235	0.0516	0.0739	0.0527	0.0521	0.0397	0.0644	0.0392	0.0687	0.0376
	0.1239	0.0517	0.0740	0.0531	0.0522	0.0399	0.0648	0.0393	0.0688	0.0380
	0.1241	0.0517	0.0744	0.0532	0.0524	0.0403	0.0648	0.0394	0.0692	0.0383
	0.1244	0.0520	0.0744	0.0533	0.0524	0.0403	0.0651	0.0394	0.0698	0.0383
	0.1246	0.0520	0.0748	0.0534	0.0525	0.0406	0.0660	0.0395	0.0703	0.0388
	0.1252	0.0529	0.0751	0.0535	0.0529	0.0408	0.0662	0.0397	0.0705	0.0393
	0.1255	0.0530	0.0751	0.0536	0.0533	0.0410	0.0664	0.0400	0.0711	0.0393
	0.1263	0.0530	0.0753	0.0543	0.0535	0.0414	0.0665	0.0403	0.0711	0.0395
	0.1271	0.0530	0.0754	0.0546	0.0535	0.0420	0.0668	0.0405	0.0714	0.0397
	0.1274	0.0535	0.0758	0.0547	0.0540	0.0420	0.0669	0.0409	0.0715	0.0399
	0.1276	0.0536	0.0761	0.0551	0.0540	0.0422	0.0672	0.0410	0.0720	0.0401
	0.1279	0.0540	0.0764	0.0556	0.0544	0.0432	0.0674	0.0412	0.0724	0.0404
	0.1284	0.0545	0.0770	0.0557	0.0547	0.0436	0.0684	0.0416	0.0748	0.0405
	0.1292	0.0547	0.0774	0.0562	0.0552	0.0438	0.0686	0.0419	0.0750	0.0410
	0.1303	0.0556	0.0775	0.0564	0.0561	0.0460	0.0704	0.0432	0.0751	0.0422
	0.1315	0.0562	0.0783	0.0577	0.0572	0.0496	0.0706	0.0449	0.0751	0.0426
	0.1329	0.0594	0.0783	0.0588	0.0582	0.0511	0.0716	0.0449	0.0755	0.0454
r'	**3**	**3**	**2**	**5**	**1**	**1**	**7**	**4**	**1**	**2**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and c	0.0452	0.0311	0.0428	0.0279	0.0301	0.0832	0.0298	0.0401	0.0234	0.0442
	0.0463	0.0315	0.0475	0.0293	0.0305	0.0890	0.0311	0.0404	0.0254	0.0444
	0.0483	0.0317	0.0501	0.0296	0.0306	0.0906	0.0311	0.0407	0.0257	0.0457
	0.0487	0.0318	0.0511	0.0296	0.0309	0.0908	0.0313	0.0415	0.0265	0.0463
	0.0491	0.0319	0.0512	0.0298	0.0312	0.0932	0.0320	0.0418	0.0265	0.0484
	0.0492	0.0324	0.0522	0.0300	0.0312	0.0934	0.0320	0.0422	0.0268	0.0487
	0.0501	0.0325	0.0526	0.0302	0.0316	0.0950	0.0327	0.0426	0.0269	0.0493
	0.0506	0.0325	0.0527	0.0304	0.0319	0.0956	0.0330	0.0429	0.0270	0.0495
	0.0506	0.0327	0.0529	0.0305	0.0322	0.0969	0.0333	0.0429	0.0272	0.0498
	0.0510	0.0330	0.0535	0.0306	0.0323	0.0971	0.0334	0.0437	0.0276	0.0502
	0.0515	0.0332	0.0541	0.0309	0.0328	0.0978	0.0335	0.0438	0.0277	0.0509
	0.0519	0.0334	0.0544	0.0310	0.0328	0.0978	0.0336	0.0439	0.0278	0.0510
	0.0524	0.0334	0.0550	0.0311	0.0329	0.0986	0.0336	0.0440	0.0278	0.0511
	0.0532	0.0337	0.0553	0.0312	0.0332	0.1000	0.0336	0.0440	0.0278	0.0515
	0.0546	0.0337	0.0559	0.0313	0.0332	0.1001	0.0337	0.0445	0.0281	0.0517
	0.0547	0.0338	0.0560	0.0313	0.0336	0.1010	0.0338	0.0445	0.0281	0.0518
	0.0547	0.0340	0.0565	0.0313	0.0340	0.1011	0.0338	0.0446	0.0282	0.0523
	0.0548	0.0341	0.0566	0.0315	0.0347	0.1014	0.0338	0.0447	0.0282	0.0529
	0.0552	0.0349	0.0570	0.0315	0.0347	0.1018	0.0340	0.0448	0.0284	0.0530
	0.0553	0.0350	0.0570	0.0317	0.0348	0.1018	0.0342	0.0449	0.0285	0.0532
	0.0555	0.0351	0.0577	0.0322	0.0349	0.1019	0.0342	0.0451	0.0285	0.0536
	0.0555	0.0351	0.0578	0.0322	0.0353	0.1020	0.0343	0.0451	0.0286	0.0536
	0.0564	0.0352	0.0579	0.0325	0.0353	0.1023	0.0347	0.0453	0.0287	0.0543
	0.0570	0.0354	0.0587	0.0325	0.0353	0.1041	0.0347	0.0453	0.0291	0.0548
	0.0574	0.0354	0.0593	0.0326	0.0355	0.1043	0.0348	0.0455	0.0292	0.0550
	0.0578	0.0358	0.0597	0.0331	0.0359	0.1046	0.0350	0.0456	0.0299	0.0550
	0.0579	0.0360	0.0600	0.0332	0.0361	0.1054	0.0354	0.0459	0.0300	0.0550
	0.0584	0.0361	0.0601	0.0334	0.0362	0.1055	0.0354	0.0462	0.0301	0.0552
	0.0587	0.0362	0.0602	0.0338	0.0364	0.1059	0.0357	0.0464	0.0303	0.0553
	0.0588	0.0369	0.0608	0.0347	0.0364	0.1066	0.0357	0.0465	0.0307	0.0559
	0.0589	0.0370	0.0609	0.0358	0.0368	0.1069	0.0365	0.0467	0.0307	0.0563
	0.0590	0.0370	0.0616	0.0361	0.0369	0.1073	0.0370	0.0467	0.0320	0.0564
	0.0592	0.0371	0.0617	0.0374	0.0371	0.1073	0.0371	0.0468	0.0329	0.0567
	0.0595	0.0377	0.0617	0.0380	0.0379	0.1075	0.0374	0.0475	0.0330	0.0573
	0.0603	0.0378	0.0618	0.0381	0.0382	0.1089	0.0376	0.0484	0.0335	0.0579
	0.0604	0.0388	0.0635	0.0385	0.0396	0.1093	0.0377	0.0491	0.0349	0.0583
	0.0612	0.0400	0.0636	0.0388	0.0397	0.1100	0.0402	0.0491	0.0358	0.0591
	0.0619	0.0433	0.0644	0.0432	0.0399	0.1102	0.0402	0.0506	0.0427	0.0591
	0.0633	0.0452	0.0650	0.0469	0.0464	0.1121	0.0407	0.0516	0.0443	0.0637
r'	**1**	**6**	**2**	**3**	**6**	**8**	**7**	**4**	**3**	**1**

	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}
All samples in the group a and c	0.0352	0.0598	0.0217	0.0279	0.0310	0.0408	0.0452	0.0528	0.0376
	0.0368	0.0669	0.0219	0.0298	0.0316	0.0412	0.0501	0.0549	0.0382
	0.0372	0.0695	0.0220	0.0306	0.0321	0.0442	0.0541	0.0550	0.0400
	0.0373	0.0707	0.0221	0.0307	0.0340	0.0452	0.0547	0.0558	0.0408
	0.0381	0.0720	0.0222	0.0307	0.0340	0.0452	0.0561	0.0564	0.0410
	0.0386	0.0725	0.0223	0.0309	0.0341	0.0456	0.0567	0.0569	0.0420
	0.0386	0.0725	0.0223	0.0309	0.0347	0.0456	0.0582	0.0573	0.0427
	0.0386	0.0729	0.0234	0.0310	0.0347	0.0460	0.0591	0.0575	0.0430
	0.0389	0.0733	0.0238	0.0313	0.0350	0.0464	0.0592	0.0588	0.0432
	0.0389	0.0744	0.0239	0.0313	0.0356	0.0465	0.0610	0.0594	0.0436
	0.0394	0.0744	0.0241	0.0314	0.0357	0.0466	0.0613	0.0596	0.0436
	0.0394	0.0745	0.0243	0.0317	0.0360	0.0468	0.0617	0.0596	0.0438
	0.0396	0.0761	0.0245	0.0317	0.0362	0.0470	0.0620	0.0599	0.0441
	0.0398	0.0762	0.0249	0.0318	0.0363	0.0470	0.0620	0.0600	0.0445
	0.0403	0.0774	0.0250	0.0322	0.0363	0.0472	0.0621	0.0600	0.0449
	0.0403	0.0781	0.0251	0.0322	0.0363	0.0478	0.0627	0.0602	0.0449
	0.0405	0.0781	0.0251	0.0323	0.0363	0.0480	0.0629	0.0613	0.0450
	0.0407	0.0786	0.0256	0.0326	0.0364	0.0480	0.0629	0.0614	0.0452
	0.0407	0.0787	0.0261	0.0329	0.0365	0.0480	0.0630	0.0618	0.0456
	0.0407	0.0791	0.0261	0.0330	0.0373	0.0483	0.0630	0.0619	0.0463
	0.0408	0.0791	0.0262	0.0331	0.0373	0.0488	0.0634	0.0626	0.0468
	0.0410	0.0799	0.0269	0.0332	0.0378	0.0489	0.0636	0.0629	0.0472
	0.0410	0.0803	0.0274	0.0336	0.0378	0.0490	0.0636	0.0630	0.0473
	0.0411	0.0806	0.0277	0.0337	0.0379	0.0491	0.0637	0.0631	0.0474
	0.0412	0.0806	0.0280	0.0338	0.0383	0.0494	0.0654	0.0637	0.0478
	0.0415	0.0813	0.0282	0.0345	0.0384	0.0503	0.0654	0.0637	0.0481
	0.0418	0.0815	0.0284	0.0347	0.0387	0.0505	0.0657	0.0640	0.0484
	0.0423	0.0818	0.0287	0.0347	0.0388	0.0507	0.0662	0.0644	0.0486
	0.0424	0.0819	0.0305	0.0350	0.0389	0.0507	0.0662	0.0648	0.0488
	0.0426	0.0826	0.0306	0.0353	0.0390	0.0508	0.0666	0.0653	0.0488
	0.0429	0.0827	0.0308	0.0363	0.0391	0.0514	0.0669	0.0654	0.0490
	0.0430	0.0837	0.0341	0.0363	0.0400	0.0515	0.0669	0.0659	0.0490
	0.0435	0.0838	0.0349	0.0364	0.0405	0.0521	0.0673	0.0660	0.0491
	0.0442	0.0839	0.0351	0.0365	0.0411	0.0527	0.0678	0.0666	0.0497
	0.0448	0.0842	0.0379	0.0372	0.0414	0.0527	0.0688	0.0673	0.0500
	0.0449	0.0856	0.0386	0.0388	0.0420	0.0528	0.0690	0.0676	0.0503
	0.0449	0.0858	0.0398	0.0389	0.0430	0.0538	0.0698	0.0676	0.0504
	0.0457	0.0873	0.0453	0.0400	0.0431	0.0547	0.0701	0.0679	0.0510
	0.0479	0.0889	0.0521	0.0402	0.0433	0.0563	0.0709	0.0701	0.0535
r'	**4**	**2**	**2**	**1**	**3**	**1**	**2**	**1**	**1**

TABLE A.22: Ranked *sim* values for $(c \vee \{a, b\})$ in VHHS dataset using U-KTS $\mathcal{F}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and b	0.0285	0.0335	0.0347	0.0242	0.0579	0.0474	0.0425	0.0358	0.0317	0.0343
	0.0292	0.0352	0.0355	0.0243	0.0640	0.0481	0.0474	0.0384	0.0325	0.0344
	0.0293	0.0361	0.0363	0.0247	0.0644	0.0486	0.0481	0.0390	0.0326	0.0345
	0.0296	0.0362	0.0368	0.0250	0.0644	0.0486	0.0484	0.0398	0.0328	0.0349
	0.0299	0.0367	0.0373	0.0251	0.0646	0.0513	0.0499	0.0400	0.0332	0.0350
	0.0300	0.0371	0.0382	0.0252	0.0648	0.0515	0.0501	0.0402	0.0335	0.0353
	0.0302	0.0380	0.0382	0.0253	0.0651	0.0517	0.0505	0.0405	0.0341	0.0356
	0.0304	0.0381	0.0385	0.0254	0.0684	0.0520	0.0512	0.0407	0.0342	0.0356
	0.0305	0.0384	0.0385	0.0254	0.0691	0.0527	0.0513	0.0411	0.0342	0.0360
	0.0308	0.0390	0.0390	0.0256	0.0694	0.0536	0.0516	0.0421	0.0348	0.0363
	0.0308	0.0390	0.0390	0.0257	0.0694	0.0541	0.0519	0.0424	0.0352	0.0367
	0.0308	0.0390	0.0399	0.0261	0.0700	0.0549	0.0521	0.0425	0.0356	0.0368
	0.0310	0.0393	0.0407	0.0261	0.0702	0.0551	0.0523	0.0425	0.0357	0.0368
	0.0311	0.0400	0.0410	0.0262	0.0702	0.0555	0.0535	0.0429	0.0360	0.0369
	0.0314	0.0401	0.0412	0.0264	0.0718	0.0556	0.0546	0.0429	0.0365	0.0370
	0.0315	0.0403	0.0413	0.0264	0.0720	0.0567	0.0549	0.0431	0.0366	0.0371
	0.0315	0.0404	0.0420	0.0267	0.0726	0.0570	0.0550	0.0432	0.0366	0.0372
	0.0316	0.0404	0.0422	0.0270	0.0734	0.0572	0.0551	0.0432	0.0367	0.0373
	0.0316	0.0405	0.0422	0.0271	0.0736	0.0573	0.0562	0.0433	0.0368	0.0375
	0.0318	0.0405	0.0423	0.0274	0.0739	0.0574	0.0563	0.0433	0.0371	0.0377
	0.0320	0.0406	0.0430	0.0275	0.0740	0.0576	0.0563	0.0436	0.0374	0.0380
	0.0320	0.0407	0.0430	0.0276	0.0743	0.0577	0.0564	0.0437	0.0375	0.0381
	0.0322	0.0410	0.0430	0.0277	0.0743	0.0578	0.0566	0.0440	0.0376	0.0383
	0.0323	0.0413	0.0431	0.0280	0.0746	0.0579	0.0568	0.0443	0.0376	0.0383
	0.0323	0.0416	0.0431	0.0283	0.0747	0.0579	0.0570	0.0445	0.0377	0.0383
	0.0326	0.0418	0.0431	0.0291	0.0747	0.0581	0.0574	0.0445	0.0381	0.0384
	0.0329	0.0418	0.0432	0.0292	0.0747	0.0584	0.0575	0.0446	0.0385	0.0384
	0.0330	0.0424	0.0437	0.0298	0.0756	0.0587	0.0576	0.0447	0.0386	0.0385
	0.0331	0.0424	0.0438	0.0304	0.0757	0.0588	0.0579	0.0447	0.0386	0.0387
	0.0335	0.0427	0.0439	0.0309	0.0760	0.0591	0.0581	0.0450	0.0388	0.0390
	0.0342	0.0432	0.0439	0.0321	0.0763	0.0594	0.0585	0.0451	0.0389	0.0392
	0.0342	0.0435	0.0441	0.0322	0.0765	0.0597	0.0588	0.0454	0.0390	0.0394
	0.0348	0.0440	0.0442	0.0337	0.0767	0.0604	0.0591	0.0455	0.0390	0.0403
	0.0358	0.0443	0.0444	0.0347	0.0769	0.0611	0.0598	0.0459	0.0391	0.0403
	0.0362	0.0445	0.0452	0.0357	0.0775	0.0611	0.0600	0.0461	0.0394	0.0410
	0.0367	0.0445	0.0454	0.0361	0.0777	0.0623	0.0603	0.0462	0.0396	0.0417
	0.0385	0.0446	0.0455	0.0391	0.0788	0.0625	0.0605	0.0463	0.0399	0.0427
	0.0385	0.0449	0.0459	0.0402	0.0790	0.0629	0.0609	0.0471	0.0432	0.0443
	0.0478	0.0459	0.0468	0.0463	0.0800	0.0637	0.0614	0.0480	0.0445	0.0504
r'	**1**	**1**	**2**	**3**	**2**	**8**	**2**	**9**	**2**	**2**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and b	0.1045	0.0417	0.0595	0.0439	0.0372	0.0338	0.0477	0.0307	0.0500	0.0300
	0.1057	0.0443	0.0635	0.0442	0.0385	0.0346	0.0491	0.0317	0.0527	0.0304
	0.1082	0.0451	0.0637	0.0457	0.0408	0.0352	0.0496	0.0318	0.0549	0.0305
	0.1106	0.0452	0.0643	0.0464	0.0435	0.0360	0.0533	0.0320	0.0562	0.0309
	0.1116	0.0454	0.0654	0.0468	0.0442	0.0363	0.0537	0.0328	0.0576	0.0313
	0.1119	0.0454	0.0660	0.0471	0.0442	0.0365	0.0560	0.0333	0.0588	0.0331
	0.1141	0.0459	0.0661	0.0476	0.0444	0.0367	0.0581	0.0345	0.0596	0.0334
	0.1152	0.0460	0.0666	0.0482	0.0462	0.0368	0.0587	0.0353	0.0615	0.0336
	0.1165	0.0461	0.0667	0.0482	0.0467	0.0368	0.0595	0.0355	0.0616	0.0339
	0.1186	0.0474	0.0672	0.0485	0.0468	0.0369	0.0598	0.0363	0.0638	0.0341
	0.1187	0.0476	0.0690	0.0488	0.0472	0.0374	0.0598	0.0365	0.0642	0.0347
	0.1192	0.0487	0.0692	0.0492	0.0479	0.0374	0.0599	0.0366	0.0645	0.0352
	0.1194	0.0490	0.0693	0.0497	0.0482	0.0376	0.0600	0.0368	0.0657	0.0358
	0.1198	0.0491	0.0704	0.0503	0.0487	0.0378	0.0626	0.0372	0.0661	0.0358
	0.1210	0.0495	0.0705	0.0512	0.0503	0.0380	0.0630	0.0376	0.0668	0.0360
	0.1212	0.0504	0.0708	0.0513	0.0504	0.0381	0.0630	0.0378	0.0670	0.0362
	0.1221	0.0508	0.0710	0.0514	0.0508	0.0381	0.0638	0.0379	0.0674	0.0362
	0.1223	0.0509	0.0711	0.0518	0.0508	0.0383	0.0641	0.0381	0.0675	0.0365
	0.1228	0.0510	0.0712	0.0519	0.0509	0.0383	0.0642	0.0381	0.0678	0.0366
	0.1229	0.0511	0.0714	0.0520	0.0510	0.0383	0.0642	0.0383	0.0685	0.0367
	0.1229	0.0513	0.0716	0.0522	0.0514	0.0385	0.0646	0.0384	0.0686	0.0367
	0.1230	0.0514	0.0717	0.0523	0.0516	0.0386	0.0648	0.0386	0.0688	0.0369
	0.1238	0.0516	0.0719	0.0523	0.0517	0.0388	0.0651	0.0386	0.0690	0.0370
	0.1239	0.0518	0.0728	0.0525	0.0519	0.0391	0.0653	0.0387	0.0690	0.0372
	0.1239	0.0518	0.0731	0.0525	0.0521	0.0394	0.0656	0.0388	0.0690	0.0375
	0.1239	0.0523	0.0731	0.0525	0.0522	0.0397	0.0657	0.0390	0.0691	0.0376
	0.1241	0.0524	0.0734	0.0528	0.0524	0.0398	0.0658	0.0393	0.0696	0.0376
	0.1243	0.0526	0.0742	0.0531	0.0525	0.0398	0.0658	0.0398	0.0698	0.0378
	0.1243	0.0527	0.0742	0.0536	0.0526	0.0399	0.0661	0.0398	0.0701	0.0381
	0.1249	0.0528	0.0744	0.0536	0.0530	0.0402	0.0666	0.0399	0.0702	0.0382
	0.1264	0.0530	0.0744	0.0538	0.0531	0.0403	0.0668	0.0400	0.0710	0.0382
	0.1266	0.0534	0.0745	0.0541	0.0533	0.0408	0.0670	0.0401	0.0714	0.0386
	0.1267	0.0534	0.0746	0.0546	0.0536	0.0415	0.0672	0.0406	0.0714	0.0386
	0.1274	0.0534	0.0757	0.0546	0.0541	0.0416	0.0673	0.0409	0.0717	0.0388
	0.1284	0.0540	0.0757	0.0554	0.0547	0.0417	0.0673	0.0409	0.0721	0.0388
	0.1292	0.0542	0.0775	0.0555	0.0551	0.0420	0.0684	0.0413	0.0721	0.0392
	0.1301	0.0542	0.0780	0.0561	0.0552	0.0422	0.0685	0.0414	0.0733	0.0401
	0.1304	0.0543	0.0781	0.0567	0.0554	0.0428	0.0692	0.0415	0.0736	0.0404
	0.1306	0.0550	0.0803	0.0570	0.0556	0.0506	0.0697	0.0419	0.0749	0.0423
r'	**3**	**2**	**1**	**5**	**1**	**1**	**9**	**2**	**4**	**2**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and b	0.0419	0.0302	0.0477	0.0283	0.0287	0.0852	0.0293	0.0397	0.0240	0.0419
	0.0474	0.0308	0.0483	0.0290	0.0301	0.0873	0.0305	0.0397	0.0256	0.0462
	0.0475	0.0311	0.0496	0.0292	0.0302	0.0886	0.0308	0.0398	0.0257	0.0471
	0.0482	0.0312	0.0504	0.0292	0.0304	0.0904	0.0312	0.0401	0.0258	0.0486
	0.0495	0.0314	0.0517	0.0293	0.0305	0.0934	0.0312	0.0406	0.0259	0.0495
	0.0496	0.0317	0.0523	0.0295	0.0315	0.0939	0.0314	0.0414	0.0261	0.0502
	0.0516	0.0317	0.0528	0.0299	0.0319	0.0952	0.0321	0.0420	0.0261	0.0511
	0.0517	0.0322	0.0534	0.0301	0.0321	0.0963	0.0321	0.0423	0.0265	0.0514
	0.0520	0.0323	0.0544	0.0302	0.0327	0.0973	0.0322	0.0423	0.0266	0.0519
	0.0521	0.0325	0.0544	0.0302	0.0332	0.0976	0.0327	0.0433	0.0266	0.0521
	0.0529	0.0328	0.0548	0.0303	0.0332	0.0988	0.0330	0.0434	0.0268	0.0524
	0.0534	0.0330	0.0551	0.0304	0.0332	0.0989	0.0330	0.0434	0.0268	0.0525
	0.0534	0.0331	0.0551	0.0306	0.0333	0.1002	0.0332	0.0436	0.0269	0.0528
	0.0538	0.0331	0.0557	0.0309	0.0333	0.1002	0.0335	0.0436	0.0272	0.0528
	0.0550	0.0332	0.0571	0.0310	0.0336	0.1013	0.0336	0.0438	0.0275	0.0532
	0.0556	0.0333	0.0574	0.0311	0.0341	0.1013	0.0337	0.0440	0.0278	0.0532
	0.0559	0.0336	0.0574	0.0311	0.0342	0.1020	0.0337	0.0441	0.0280	0.0533
	0.0560	0.0337	0.0577	0.0314	0.0342	0.1024	0.0338	0.0443	0.0283	0.0533
	0.0560	0.0339	0.0587	0.0315	0.0344	0.1025	0.0339	0.0444	0.0283	0.0536
	0.0562	0.0345	0.0588	0.0316	0.0345	0.1027	0.0339	0.0445	0.0284	0.0539
	0.0565	0.0346	0.0588	0.0316	0.0346	0.1032	0.0342	0.0445	0.0285	0.0542
	0.0566	0.0346	0.0589	0.0320	0.0347	0.1037	0.0343	0.0446	0.0290	0.0546
	0.0567	0.0346	0.0589	0.0324	0.0349	0.1038	0.0345	0.0447	0.0290	0.0548
	0.0570	0.0351	0.0589	0.0325	0.0349	0.1039	0.0346	0.0448	0.0290	0.0548
	0.0572	0.0354	0.0591	0.0326	0.0350	0.1042	0.0348	0.0450	0.0291	0.0551
	0.0574	0.0354	0.0592	0.0331	0.0351	0.1046	0.0351	0.0453	0.0292	0.0556
	0.0577	0.0358	0.0593	0.0340	0.0352	0.1047	0.0351	0.0453	0.0292	0.0556
	0.0578	0.0358	0.0595	0.0341	0.0358	0.1050	0.0351	0.0459	0.0292	0.0556
	0.0583	0.0359	0.0595	0.0341	0.0359	0.1051	0.0352	0.0462	0.0300	0.0563
	0.0589	0.0362	0.0595	0.0342	0.0360	0.1051	0.0352	0.0462	0.0302	0.0570
	0.0591	0.0365	0.0603	0.0345	0.0364	0.1052	0.0362	0.0462	0.0312	0.0571
	0.0593	0.0366	0.0606	0.0353	0.0364	0.1058	0.0363	0.0463	0.0314	0.0571
	0.0595	0.0370	0.0606	0.0354	0.0368	0.1066	0.0365	0.0471	0.0327	0.0573
	0.0595	0.0373	0.0607	0.0356	0.0369	0.1068	0.0370	0.0473	0.0338	0.0575
	0.0596	0.0375	0.0615	0.0369	0.0371	0.1071	0.0378	0.0473	0.0340	0.0575
	0.0602	0.0386	0.0618	0.0378	0.0372	0.1075	0.0382	0.0477	0.0342	0.0577
	0.0607	0.0398	0.0624	0.0381	0.0374	0.1078	0.0386	0.0479	0.0369	0.0582
	0.0608	0.0400	0.0627	0.0419	0.0384	0.1091	0.0411	0.0483	0.0388	0.0582
	0.0609	0.0474	0.0640	0.0507	0.0410	0.1100	0.0449	0.0509	0.0428	0.0587
r'	**1**	**4**	**2**	**4**	**5**	**6**	**8**	**3**	**2**	**1**

	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}
All samples in the group a and b	0.0343	0.0604	0.0211	0.0279	0.0335	0.0422	0.0482	0.0541	0.0375
	0.0366	0.0658	0.0217	0.0284	0.0340	0.0423	0.0483	0.0542	0.0383
	0.0368	0.0674	0.0219	0.0296	0.0351	0.0431	0.0493	0.0562	0.0389
	0.0368	0.0684	0.0220	0.0303	0.0351	0.0441	0.0542	0.0562	0.0390
	0.0376	0.0686	0.0221	0.0304	0.0353	0.0446	0.0547	0.0564	0.0418
	0.0379	0.0706	0.0228	0.0312	0.0355	0.0446	0.0555	0.0574	0.0419
	0.0379	0.0720	0.0229	0.0313	0.0355	0.0454	0.0572	0.0578	0.0429
	0.0379	0.0727	0.0230	0.0315	0.0356	0.0454	0.0590	0.0582	0.0431
	0.0383	0.0741	0.0232	0.0317	0.0356	0.0455	0.0591	0.0588	0.0434
	0.0385	0.0745	0.0233	0.0317	0.0356	0.0458	0.0600	0.0593	0.0434
	0.0387	0.0757	0.0235	0.0321	0.0356	0.0460	0.0602	0.0598	0.0445
	0.0390	0.0766	0.0236	0.0324	0.0357	0.0462	0.0611	0.0599	0.0449
	0.0391	0.0773	0.0241	0.0326	0.0361	0.0463	0.0619	0.0607	0.0451
	0.0391	0.0773	0.0243	0.0327	0.0362	0.0466	0.0622	0.0609	0.0454
	0.0393	0.0773	0.0243	0.0328	0.0364	0.0467	0.0625	0.0613	0.0462
	0.0393	0.0781	0.0245	0.0330	0.0364	0.0474	0.0626	0.0613	0.0462
	0.0395	0.0788	0.0246	0.0330	0.0365	0.0477	0.0632	0.0614	0.0464
	0.0395	0.0791	0.0249	0.0330	0.0366	0.0479	0.0635	0.0617	0.0464
	0.0397	0.0797	0.0251	0.0331	0.0366	0.0482	0.0639	0.0618	0.0465
	0.0398	0.0797	0.0253	0.0331	0.0366	0.0483	0.0642	0.0621	0.0466
	0.0400	0.0802	0.0255	0.0333	0.0371	0.0487	0.0643	0.0621	0.0470
	0.0400	0.0802	0.0265	0.0335	0.0374	0.0487	0.0644	0.0624	0.0472
	0.0404	0.0803	0.0267	0.0337	0.0375	0.0492	0.0644	0.0625	0.0473
	0.0407	0.0805	0.0271	0.0341	0.0375	0.0494	0.0648	0.0626	0.0473
	0.0408	0.0806	0.0276	0.0342	0.0376	0.0494	0.0648	0.0626	0.0477
	0.0408	0.0809	0.0277	0.0342	0.0376	0.0495	0.0650	0.0628	0.0478
	0.0409	0.0809	0.0279	0.0343	0.0376	0.0495	0.0651	0.0629	0.0478
	0.0410	0.0810	0.0284	0.0343	0.0377	0.0501	0.0652	0.0634	0.0478
	0.0413	0.0813	0.0288	0.0344	0.0378	0.0502	0.0653	0.0636	0.0479
	0.0413	0.0815	0.0301	0.0344	0.0379	0.0503	0.0656	0.0637	0.0483
	0.0416	0.0823	0.0305	0.0350	0.0384	0.0504	0.0658	0.0642	0.0486
	0.0418	0.0830	0.0339	0.0351	0.0385	0.0507	0.0662	0.0648	0.0489
	0.0424	0.0832	0.0346	0.0352	0.0400	0.0507	0.0662	0.0655	0.0489
	0.0432	0.0832	0.0364	0.0354	0.0400	0.0508	0.0664	0.0658	0.0489
	0.0435	0.0836	0.0369	0.0359	0.0402	0.0514	0.0666	0.0667	0.0497
	0.0439	0.0836	0.0393	0.0367	0.0403	0.0518	0.0668	0.0670	0.0499
	0.0454	0.0858	0.0393	0.0370	0.0408	0.0524	0.0671	0.0676	0.0501
	0.0459	0.0859	0.0419	0.0376	0.0421	0.0526	0.0686	0.0682	0.0506
	0.0508	0.0860	0.0483	0.0405	0.0461	0.0541	0.0692	0.0689	0.0517
r'	**6**	**2**	**2**	**1**	**4**	**2**	**1**	**4**	**2**

TABLE A.23: Ranked *sim* values for $(a \vee \{b, c\})$ in VHHS dataset using U-KTS $\mathcal{KH}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group b and c	0.0146	0.0216	0.0157	0.0179	0.0482	0.0293	0.0404	0.0232	0.0124	0.0206
	0.0152	0.0235	0.0261	0.0232	0.0501	0.0313	0.0430	0.0237	0.0204	0.0228
	0.0213	0.0238	0.0267	0.0236	0.0569	0.0332	0.0432	0.0240	0.0209	0.0264
	0.0235	0.0249	0.0276	0.0241	0.0580	0.0333	0.0459	0.0247	0.0234	0.0300
	0.0254	0.0265	0.0277	0.0269	0.0589	0.0343	0.0465	0.0281	0.0241	0.0323
	0.0275	0.0275	0.0277	0.0273	0.0592	0.0349	0.0485	0.0287	0.0275	0.0325
	0.0280	0.0300	0.0284	0.0274	0.0595	0.0355	0.0489	0.0291	0.0316	0.0325
	0.0286	0.0311	0.0289	0.0287	0.0598	0.0368	0.0495	0.0300	0.0318	0.0326
	0.0292	0.0312	0.0297	0.0291	0.0606	0.0372	0.0503	0.0314	0.0321	0.0327
	0.0305	0.0316	0.0302	0.0297	0.0615	0.0376	0.0522	0.0318	0.0384	0.0335
	0.0315	0.0322	0.0303	0.0303	0.0628	0.0394	0.0524	0.0319	0.0386	0.0350
	0.0317	0.0324	0.0308	0.0313	0.0628	0.0394	0.0537	0.0324	0.0390	0.0358
	0.0322	0.0336	0.0326	0.0315	0.0636	0.0396	0.0541	0.0332	0.0405	0.0359
	0.0339	0.0341	0.0331	0.0330	0.0661	0.0398	0.0546	0.0339	0.0407	0.0366
	0.0345	0.0343	0.0331	0.0332	0.0674	0.0398	0.0566	0.0342	0.0407	0.0374
	0.0349	0.0350	0.0360	0.0353	0.0680	0.0401	0.0581	0.0346	0.0414	0.0383
	0.0355	0.0360	0.0375	0.0354	0.0692	0.0410	0.0581	0.0357	0.0423	0.0384
	0.0368	0.0368	0.0379	0.0358	0.0695	0.0430	0.0597	0.0373	0.0441	0.0385
	0.0382	0.0380	0.0387	0.0361	0.0698	0.0443	0.0629	0.0385	0.0447	0.0386
	0.0387	0.0382	0.0394	0.0364	0.0705	0.0449	0.0636	0.0412	0.0448	0.0394
	0.0389	0.0390	0.0399	0.0384	0.0712	0.0452	0.0656	0.0427	0.0502	0.0396
	0.0410	0.0392	0.0420	0.0384	0.0733	0.0453	0.0662	0.0428	0.0507	0.0401
	0.0427	0.0392	0.0421	0.0385	0.0738	0.0453	0.0686	0.0438	0.0531	0.0408
	0.0436	0.0394	0.0422	0.0395	0.0748	0.0460	0.0693	0.0464	0.0536	0.0435
	0.0470	0.0425	0.0422	0.0451	0.0756	0.0460	0.0703	0.0464	0.0542	0.0442
	0.0496	0.0430	0.0444	0.0461	0.0757	0.0480	0.0707	0.0474	0.0548	0.0448
	0.0498	0.0432	0.0476	0.0475	0.0769	0.0492	0.0723	0.0490	0.0556	0.0467
	0.0520	0.0446	0.0486	0.0496	0.0785	0.0514	0.0738	0.0507	0.0582	0.0468
	0.0528	0.0466	0.0551	0.0507	0.0798	0.0539	0.0771	0.0550	0.0661	0.0469
	0.0584	0.0480	0.0568	0.0513	0.0799	0.0540	0.0771	0.0569	0.0711	0.0500
	0.0661	0.0481	0.0618	0.0653	0.0801	0.0644	0.0774	0.0598	0.0721	0.0506
	0.0795	0.0501	0.0710	0.0698	0.0840	0.0740	0.0782	0.0808	0.0956	0.0528
	0.0808	0.0616	0.0756	0.0765	0.0854	0.0752	0.0789	0.0819	0.0974	0.0532
	0.0907	0.0620	0.0838	0.0785	0.0872	0.0764	0.0799	0.0840	0.1028	0.0533
	0.0958	0.0688	0.0904	0.0814	0.0880	0.0839	0.0821	0.0847	0.1078	0.0635
	0.0971	0.0714	0.0926	0.0946	0.0893	0.0868	0.0831	0.0946	0.1102	0.0661
	0.1129	0.0857	0.0975	0.0999	0.0898	0.0883	0.0836	0.1018	0.1274	0.0823
	0.1129	0.0888	0.0999	0.1124	0.0923	0.1048	0.0907	0.1043	0.1294	0.0862
	0.1293	0.0892	0.1233	0.1150	0.1104	0.1184	0.0968	0.1258	0.1411	0.0873
r'	**2**	**2**	**3**	**1**	**1**	**11**	**3**	**3**	**5**	**8**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group b and c	0.0564	0.0351	0.0507	0.0269	0.0272	0.0244	0.0375	0.0262	0.0417	0.0354
	0.0601	0.0361	0.0571	0.0293	0.0273	0.0269	0.0395	0.0271	0.0445	0.0444
	0.0726	0.0366	0.0578	0.0324	0.0290	0.0330	0.0403	0.0281	0.0457	0.0461
	0.0749	0.0391	0.0591	0.0339	0.0290	0.0330	0.0414	0.0294	0.0500	0.0464
	0.0773	0.0394	0.0600	0.0368	0.0305	0.0345	0.0415	0.0294	0.0526	0.0465
	0.0808	0.0399	0.0607	0.0373	0.0318	0.0384	0.0443	0.0299	0.0539	0.0490
	0.0912	0.0412	0.0623	0.0413	0.0320	0.0388	0.0473	0.0326	0.0549	0.0498
	0.1022	0.0416	0.0626	0.0419	0.0321	0.0391	0.0473	0.0341	0.0575	0.0514
	0.1027	0.0426	0.0644	0.0420	0.0328	0.0394	0.0474	0.0345	0.0577	0.0523
	0.1035	0.0428	0.0655	0.0425	0.0335	0.0397	0.0476	0.0351	0.0578	0.0531
	0.1090	0.0454	0.0658	0.0426	0.0342	0.0399	0.0488	0.0351	0.0584	0.0534
	0.1137	0.0460	0.0658	0.0426	0.0361	0.0406	0.0520	0.0361	0.0586	0.0540
	0.1167	0.0467	0.0663	0.0431	0.0368	0.0419	0.0526	0.0366	0.0593	0.0545
	0.1171	0.0470	0.0694	0.0437	0.0387	0.0421	0.0530	0.0388	0.0598	0.0545
	0.1174	0.0478	0.0700	0.0437	0.0391	0.0421	0.0531	0.0391	0.0599	0.0547
	0.1180	0.0480	0.0701	0.0439	0.0399	0.0421	0.0535	0.0392	0.0603	0.0556
	0.1197	0.0499	0.0706	0.0441	0.0404	0.0432	0.0537	0.0397	0.0606	0.0556
	0.1200	0.0515	0.0717	0.0449	0.0404	0.0434	0.0546	0.0399	0.0606	0.0556
	0.1208	0.0515	0.0718	0.0466	0.0408	0.0445	0.0555	0.0402	0.0606	0.0572
	0.1219	0.0523	0.0718	0.0476	0.0425	0.0450	0.0560	0.0414	0.0612	0.0574
	0.1244	0.0533	0.0726	0.0483	0.0440	0.0458	0.0565	0.0433	0.0618	0.0575
	0.1256	0.0537	0.0727	0.0489	0.0447	0.0464	0.0578	0.0437	0.0620	0.0584
	0.1284	0.0544	0.0728	0.0492	0.0460	0.0468	0.0580	0.0444	0.0635	0.0600
	0.1300	0.0548	0.0730	0.0494	0.0476	0.0474	0.0581	0.0459	0.0643	0.0601
	0.1302	0.0564	0.0742	0.0497	0.0479	0.0480	0.0588	0.0460	0.0650	0.0603
	0.1325	0.0568	0.0751	0.0498	0.0481	0.0482	0.0591	0.0489	0.0657	0.0615
	0.1327	0.0587	0.0757	0.0501	0.0498	0.0484	0.0591	0.0503	0.0659	0.0618
	0.1329	0.0591	0.0762	0.0507	0.0525	0.0486	0.0610	0.0514	0.0660	0.0618
	0.1341	0.0615	0.0770	0.0509	0.0550	0.0505	0.0622	0.0552	0.0664	0.0658
	0.1360	0.0624	0.0771	0.0533	0.0560	0.0532	0.0623	0.0557	0.0673	0.0663
	0.1365	0.0648	0.0775	0.0544	0.0614	0.0558	0.0636	0.0589	0.0675	0.0685
	0.1459	0.0725	0.0807	0.0566	0.0616	0.0566	0.0649	0.0711	0.0703	0.0753
	0.1493	0.0736	0.0808	0.0584	0.0658	0.0567	0.0654	0.0736	0.0708	0.0756
	0.1496	0.0762	0.0829	0.0611	0.0692	0.0616	0.0657	0.0771	0.0726	0.0785
	0.1580	0.0784	0.0835	0.0614	0.0703	0.0634	0.0675	0.0859	0.0734	0.0797
	0.1591	0.0834	0.0866	0.0633	0.0753	0.0643	0.0676	0.0864	0.0774	0.0837
	0.1615	0.0853	0.0876	0.0650	0.1015	0.0834	0.0712	0.0900	0.0777	0.0918
	0.1638	0.0888	0.0891	0.0798	0.1033	0.0842	0.0746	0.1029	0.0809	0.0961
	0.1760	0.0912	0.0946	0.0829	0.1034	0.0855	0.0795	0.1149	0.0810	0.0977
r'	**3**	**8**	**3**	**6**	**4**	**1**	**12**	**1**	**15**	**1**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group b and c	0.0273	0.0288	0.0531	0.0221	0.0174	0.0401	0.0225	0.0296	0.0189	0.0412
	0.0304	0.0295	0.0572	0.0249	0.0194	0.0423	0.0234	0.0306	0.0211	0.0438
	0.0307	0.0330	0.0591	0.0252	0.0217	0.0429	0.0268	0.0326	0.0216	0.0446
	0.0308	0.0334	0.0606	0.0263	0.0227	0.0434	0.0269	0.0327	0.0216	0.0452
	0.0308	0.0339	0.0608	0.0264	0.0227	0.0456	0.0307	0.0328	0.0274	0.0453
	0.0315	0.0339	0.0650	0.0275	0.0236	0.0468	0.0313	0.0339	0.0304	0.0458
	0.0322	0.0342	0.0683	0.0287	0.0239	0.0472	0.0327	0.0357	0.0306	0.0463
	0.0350	0.0353	0.0691	0.0287	0.0243	0.0473	0.0332	0.0361	0.0308	0.0470
	0.0354	0.0363	0.0701	0.0288	0.0251	0.0474	0.0344	0.0386	0.0316	0.0471
	0.0362	0.0365	0.0705	0.0292	0.0265	0.0484	0.0349	0.0386	0.0343	0.0476
	0.0365	0.0366	0.0714	0.0312	0.0284	0.0485	0.0355	0.0388	0.0356	0.0487
	0.0370	0.0371	0.0715	0.0314	0.0286	0.0498	0.0357	0.0389	0.0365	0.0494
	0.0372	0.0372	0.0717	0.0346	0.0293	0.0510	0.0364	0.0397	0.0374	0.0506
	0.0381	0.0373	0.0717	0.0357	0.0317	0.0511	0.0367	0.0401	0.0380	0.0508
	0.0407	0.0376	0.0723	0.0362	0.0329	0.0522	0.0378	0.0404	0.0384	0.0514
	0.0412	0.0376	0.0734	0.0374	0.0335	0.0523	0.0396	0.0409	0.0385	0.0516
	0.0424	0.0386	0.0743	0.0384	0.0347	0.0530	0.0404	0.0409	0.0396	0.0523
	0.0433	0.0388	0.0745	0.0391	0.0348	0.0544	0.0411	0.0415	0.0406	0.0539
	0.0438	0.0402	0.0757	0.0397	0.0355	0.0551	0.0437	0.0422	0.0424	0.0544
	0.0440	0.0410	0.0778	0.0412	0.0367	0.0556	0.0443	0.0435	0.0430	0.0555
	0.0447	0.0420	0.0781	0.0425	0.0393	0.0557	0.0452	0.0439	0.0469	0.0576
	0.0452	0.0426	0.0781	0.0426	0.0400	0.0558	0.0483	0.0452	0.0473	0.0578
	0.0474	0.0443	0.0784	0.0447	0.0401	0.0561	0.0485	0.0456	0.0475	0.0588
	0.0477	0.0443	0.0784	0.0451	0.0401	0.0595	0.0491	0.0461	0.0499	0.0591
	0.0498	0.0447	0.0793	0.0463	0.0432	0.0597	0.0510	0.0461	0.0505	0.0598
	0.0498	0.0451	0.0794	0.0468	0.0468	0.0608	0.0510	0.0484	0.0509	0.0599
	0.0510	0.0479	0.0807	0.0485	0.0471	0.0611	0.0532	0.0488	0.0515	0.0603
	0.0524	0.0485	0.0809	0.0486	0.0477	0.0615	0.0534	0.0490	0.0555	0.0604
	0.0546	0.0518	0.0814	0.0551	0.0501	0.0616	0.0536	0.0491	0.0564	0.0605
	0.0593	0.0518	0.0820	0.0563	0.0560	0.0617	0.0536	0.0532	0.0595	0.0606
	0.0606	0.0548	0.0837	0.0567	0.0567	0.0623	0.0590	0.0533	0.0662	0.0671
	0.0627	0.0574	0.0846	0.0693	0.0758	0.0630	0.0673	0.0537	0.0749	0.0683
	0.0681	0.0684	0.0848	0.0767	0.0787	0.0641	0.0766	0.0584	0.0866	0.0698
	0.0721	0.0707	0.0865	0.0793	0.0806	0.0666	0.0772	0.0597	0.0892	0.0774
	0.0756	0.0745	0.0866	0.0840	0.0844	0.0686	0.0794	0.0602	0.0912	0.0777
	0.0803	0.0772	0.0896	0.0850	0.0880	0.0763	0.0814	0.0753	0.1011	0.0787
	0.0903	0.0817	0.0901	0.0967	0.0994	0.0765	0.0948	0.0805	0.1023	0.0818
	0.0948	0.0862	0.0977	0.1080	0.1039	0.0869	0.0961	0.0808	0.1111	0.0819
	0.1011	0.0965	0.1003	0.1094	0.1155	0.0895	0.1181	0.0868	0.1135	0.1018
r'	**5**	**1**	**3**	**1**	**2**	**3**	**2**	**4**	**5**	**3**

	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}
All samples in the group b and c	0.0433	0.0565	0.0157	0.0336	0.0228	0.0438	0.0330	0.0414	0.0274
	0.0445	0.0573	0.0182	0.0338	0.0265	0.0496	0.0336	0.0469	0.0303
	0.0481	0.0591	0.0197	0.0376	0.0268	0.0505	0.0344	0.0475	0.0306
	0.0490	0.0605	0.0201	0.0383	0.0286	0.0506	0.0359	0.0483	0.0313
	0.0508	0.0616	0.0234	0.0385	0.0301	0.0507	0.0378	0.0506	0.0319
	0.0561	0.0624	0.0239	0.0394	0.0314	0.0512	0.0398	0.0509	0.0331
	0.0564	0.0643	0.0314	0.0405	0.0318	0.0514	0.0398	0.0515	0.0332
	0.0564	0.0656	0.0317	0.0410	0.0320	0.0530	0.0400	0.0532	0.0350
	0.0576	0.0656	0.0359	0.0415	0.0329	0.0537	0.0406	0.0571	0.0360
	0.0579	0.0659	0.0377	0.0423	0.0338	0.0551	0.0408	0.0579	0.0368
	0.0598	0.0659	0.0381	0.0437	0.0350	0.0552	0.0415	0.0587	0.0368
	0.0600	0.0670	0.0382	0.0440	0.0351	0.0571	0.0416	0.0602	0.0369
	0.0604	0.0678	0.0390	0.0443	0.0358	0.0575	0.0423	0.0609	0.0370
	0.0618	0.0679	0.0409	0.0451	0.0359	0.0576	0.0427	0.0616	0.0393
	0.0624	0.0685	0.0410	0.0470	0.0364	0.0582	0.0432	0.0618	0.0397
	0.0629	0.0685	0.0411	0.0471	0.0368	0.0620	0.0434	0.0643	0.0413
	0.0636	0.0694	0.0423	0.0473	0.0369	0.0620	0.0447	0.0645	0.0421
	0.0638	0.0710	0.0434	0.0483	0.0381	0.0626	0.0449	0.0653	0.0432
	0.0653	0.0711	0.0440	0.0488	0.0387	0.0636	0.0452	0.0685	0.0432
	0.0658	0.0721	0.0455	0.0491	0.0393	0.0636	0.0466	0.0686	0.0435
	0.0661	0.0725	0.0466	0.0495	0.0395	0.0646	0.0469	0.0754	0.0436
	0.0682	0.0737	0.0486	0.0511	0.0395	0.0648	0.0473	0.0766	0.0440
	0.0686	0.0739	0.0490	0.0526	0.0398	0.0649	0.0482	0.0777	0.0449
	0.0689	0.0752	0.0536	0.0527	0.0405	0.0651	0.0487	0.0780	0.0463
	0.0695	0.0765	0.0546	0.0547	0.0410	0.0664	0.0491	0.0785	0.0470
	0.0699	0.0788	0.0552	0.0550	0.0429	0.0668	0.0541	0.0788	0.0473
	0.0702	0.0795	0.0558	0.0554	0.0433	0.0678	0.0543	0.0793	0.0480
	0.0730	0.0819	0.0566	0.0555	0.0474	0.0682	0.0544	0.0793	0.0487
	0.0769	0.0826	0.0608	0.0561	0.0533	0.0690	0.0546	0.0811	0.0494
	0.0772	0.0852	0.0687	0.0564	0.0557	0.0698	0.0564	0.0830	0.0514
	0.0787	0.0855	0.0825	0.0576	0.0571	0.0705	0.0566	0.0847	0.0526
	0.0789	0.0866	0.0892	0.0580	0.0677	0.0711	0.0599	0.0859	0.0580
	0.0792	0.0868	0.0915	0.0583	0.0686	0.0731	0.0617	0.0869	0.0700
	0.0794	0.0887	0.0985	0.0668	0.0688	0.0740	0.0639	0.0943	0.0704
	0.0817	0.0894	0.1058	0.0675	0.0753	0.0776	0.0639	0.0978	0.0745
	0.0895	0.0936	0.1066	0.0718	0.0895	0.0824	0.0649	0.1010	0.0753
	0.0936	0.0975	0.1101	0.0721	0.0897	0.0837	0.0789	0.1013	0.0853
	0.1071	0.0985	0.1240	0.0780	0.0925	0.0880	0.0876	0.1039	0.0876
	0.1173	0.1082	0.1266	0.0889	0.1063	0.0897	0.0944	0.1183	0.1035
r'	**2**	**6**	**8**	**10**	**5**	**1**	**2**	**1**	**4**

TABLE A.24: Ranked *sim* values for $(b \vee \{a, c\})$ in VHHS dataset using U-KTS $\mathcal{KH}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and c	0.0515	0.0624	0.0550	0.0386	0.0910	0.0882	0.0714	0.0722	0.0511	0.0622
	0.0521	0.0627	0.0636	0.0416	0.0912	0.0892	0.0723	0.0730	0.0554	0.0624
	0.0523	0.0633	0.0644	0.0419	0.0947	0.0930	0.0734	0.0730	0.0591	0.0626
	0.0525	0.0641	0.0650	0.0426	0.0967	0.0954	0.0746	0.0754	0.0593	0.0627
	0.0527	0.0660	0.0660	0.0427	0.0973	0.0964	0.0747	0.0759	0.0594	0.0647
	0.0528	0.0664	0.0665	0.0429	0.0984	0.0972	0.0760	0.0762	0.0618	0.0647
	0.0531	0.0674	0.0671	0.0436	0.1026	0.0978	0.0762	0.0769	0.0619	0.0650
	0.0531	0.0685	0.0672	0.0440	0.1060	0.0979	0.0800	0.0776	0.0620	0.0651
	0.0542	0.0696	0.0675	0.0443	0.1062	0.1009	0.0807	0.0782	0.0624	0.0654
	0.0545	0.0700	0.0680	0.0443	0.1096	0.1023	0.0818	0.0782	0.0628	0.0660
	0.0546	0.0702	0.0682	0.0450	0.1099	0.1024	0.0822	0.0783	0.0631	0.0662
	0.0546	0.0707	0.0683	0.0450	0.1113	0.1025	0.0837	0.0788	0.0654	0.0664
	0.0550	0.0711	0.0687	0.0450	0.1113	0.1044	0.0852	0.0796	0.0664	0.0666
	0.0562	0.0712	0.0693	0.0454	0.1116	0.1048	0.0857	0.0798	0.0669	0.0670
	0.0563	0.0714	0.0703	0.0455	0.1132	0.1058	0.0863	0.0799	0.0674	0.0672
	0.0569	0.0718	0.0706	0.0461	0.1143	0.1059	0.0868	0.0802	0.0676	0.0672
	0.0570	0.0732	0.0708	0.0462	0.1144	0.1071	0.0868	0.0808	0.0679	0.0674
	0.0575	0.0735	0.0721	0.0463	0.1145	0.1072	0.0880	0.0810	0.0681	0.0679
	0.0577	0.0738	0.0722	0.0469	0.1145	0.1079	0.0881	0.0819	0.0684	0.0680
	0.0580	0.0742	0.0728	0.0470	0.1147	0.1081	0.0884	0.0820	0.0686	0.0681
	0.0586	0.0744	0.0738	0.0471	0.1158	0.1083	0.0885	0.0820	0.0688	0.0702
	0.0590	0.0745	0.0742	0.0478	0.1160	0.1086	0.0886	0.0825	0.0690	0.0703
	0.0598	0.0747	0.0748	0.0486	0.1190	0.1105	0.0888	0.0826	0.0691	0.0704
	0.0609	0.0752	0.0751	0.0493	0.1204	0.1109	0.0891	0.0832	0.0693	0.0704
	0.0609	0.0755	0.0753	0.0500	0.1221	0.1116	0.0908	0.0832	0.0693	0.0706
	0.0618	0.0760	0.0753	0.0507	0.1226	0.1127	0.0911	0.0832	0.0713	0.0707
	0.0620	0.0769	0.0755	0.0520	0.1233	0.1129	0.0916	0.0834	0.0722	0.0710
	0.0621	0.0770	0.0755	0.0521	0.1239	0.1131	0.0922	0.0837	0.0728	0.0716
	0.0622	0.0773	0.0764	0.0528	0.1253	0.1147	0.0935	0.0840	0.0739	0.0717
	0.0624	0.0776	0.0770	0.0533	0.1260	0.1159	0.0947	0.0844	0.0740	0.0723
	0.0628	0.0777	0.0773	0.0536	0.1272	0.1160	0.0964	0.0848	0.0741	0.0733
	0.0628	0.0777	0.0781	0.0549	0.1273	0.1164	0.0971	0.0860	0.0743	0.0734
	0.0644	0.0814	0.0791	0.0556	0.1275	0.1178	0.0973	0.0870	0.0746	0.0751
	0.0660	0.0814	0.0792	0.0574	0.1280	0.1182	0.0974	0.0884	0.0747	0.0754
	0.0665	0.0826	0.0797	0.0577	0.1310	0.1183	0.0977	0.0897	0.0749	0.0756
	0.0676	0.0834	0.0800	0.0626	0.1333	0.1184	0.0985	0.0898	0.0756	0.0778
	0.0712	0.0858	0.0801	0.0663	0.1336	0.1209	0.1022	0.0900	0.0777	0.0784
	0.0720	0.0892	0.0840	0.0720	0.1429	0.1238	0.1061	0.0922	0.0811	0.0798
	0.0837	0.0894	0.0871	0.0738	0.1438	0.1315	0.1108	0.0939	0.0870	0.0901
r'	**3**	**1**	**3**	**2**	**2**	**9**	**5**	**4**	**2**	**9**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and c	0.1774	0.0756	0.1118	0.0796	0.0664	0.0578	0.0709	0.0544	0.0902	0.0557
	0.1780	0.0776	0.1137	0.0796	0.0750	0.0588	0.0716	0.0558	0.0927	0.0578
	0.1859	0.0786	0.1163	0.0804	0.0766	0.0588	0.0719	0.0584	0.0964	0.0588
	0.1862	0.0804	0.1164	0.0818	0.0773	0.0602	0.0736	0.0596	0.0982	0.0596
	0.1882	0.0816	0.1202	0.0824	0.0778	0.0602	0.0744	0.0621	0.1017	0.0600
	0.1886	0.0817	0.1203	0.0828	0.0784	0.0606	0.0748	0.0625	0.1053	0.0601
	0.1910	0.0817	0.1210	0.0833	0.0794	0.0607	0.0751	0.0633	0.1078	0.0602
	0.1925	0.0822	0.1215	0.0846	0.0808	0.0612	0.0796	0.0634	0.1082	0.0604
	0.2052	0.0826	0.1248	0.0848	0.0811	0.0613	0.0814	0.0635	0.1107	0.0613
	0.2056	0.0831	0.1262	0.0852	0.0817	0.0616	0.0840	0.0640	0.1131	0.0614
	0.2079	0.0833	0.1264	0.0857	0.0820	0.0617	0.0865	0.0648	0.1132	0.0614
	0.2095	0.0834	0.1278	0.0858	0.0839	0.0638	0.0871	0.0650	0.1137	0.0624
	0.2117	0.0842	0.1278	0.0864	0.0858	0.0639	0.0898	0.0651	0.1139	0.0626
	0.2123	0.0860	0.1282	0.0875	0.0866	0.0639	0.0903	0.0652	0.1139	0.0628
	0.2124	0.0863	0.1284	0.0876	0.0867	0.0641	0.0911	0.0657	0.1156	0.0629
	0.2133	0.0863	0.1288	0.0880	0.0868	0.0642	0.0939	0.0666	0.1158	0.0630
	0.2146	0.0874	0.1302	0.0887	0.0879	0.0642	0.0955	0.0672	0.1160	0.0635
	0.2148	0.0880	0.1312	0.0894	0.0882	0.0642	0.0972	0.0674	0.1161	0.0643
	0.2150	0.0881	0.1318	0.0894	0.0884	0.0644	0.0986	0.0676	0.1161	0.0650
	0.2156	0.0890	0.1318	0.0915	0.0892	0.0645	0.1000	0.0678	0.1163	0.0669
	0.2156	0.0891	0.1320	0.0926	0.0900	0.0655	0.1008	0.0684	0.1195	0.0678
	0.2198	0.0896	0.1320	0.0931	0.0912	0.0655	0.1032	0.0684	0.1202	0.0684
	0.2216	0.0899	0.1322	0.0932	0.0922	0.0656	0.1036	0.0698	0.1215	0.0684
	0.2221	0.0912	0.1326	0.0938	0.0928	0.0658	0.1043	0.0712	0.1240	0.0685
	0.2242	0.0913	0.1340	0.0940	0.0937	0.0659	0.1044	0.0714	0.1271	0.0686
	0.2245	0.0915	0.1358	0.0942	0.0958	0.0666	0.1058	0.0725	0.1278	0.0694
	0.2247	0.0925	0.1362	0.0954	0.0959	0.0667	0.1082	0.0726	0.1279	0.0699
	0.2265	0.0933	0.1372	0.0958	0.0966	0.0676	0.1085	0.0728	0.1281	0.0702
	0.2275	0.0937	0.1393	0.0966	0.0967	0.0689	0.1087	0.0731	0.1282	0.0702
	0.2295	0.0937	0.1394	0.0968	0.0970	0.0696	0.1104	0.0734	0.1294	0.0718
	0.2325	0.0963	0.1395	0.0979	0.0972	0.0702	0.1105	0.0744	0.1300	0.0720
	0.2332	0.0965	0.1403	0.0986	0.0974	0.0731	0.1123	0.0749	0.1305	0.0720
	0.2338	0.0970	0.1452	0.0988	0.0983	0.0735	0.1125	0.0756	0.1306	0.0730
	0.2345	0.0970	0.1467	0.0992	0.0990	0.0739	0.1134	0.0760	0.1311	0.0749
	0.2390	0.0974	0.1485	0.1000	0.0994	0.0744	0.1148	0.0760	0.1355	0.0755
	0.2414	0.0991	0.1510	0.1002	0.1000	0.0752	0.1179	0.0774	0.1382	0.0764
	0.2439	0.1008	0.1511	0.1004	0.1031	0.0760	0.1196	0.0784	0.1390	0.0771
	0.2541	0.1018	0.1534	0.1051	0.1079	0.0774	0.1282	0.0806	0.1475	0.0804
	0.2552	0.1038	0.1563	0.1095	0.1097	0.0861	0.1300	0.0846	0.1484	0.0844
r'	**5**	**6**	**2**	**4**	**2**	**2**	**10**	**2**	**13**	**3**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and c	0.0856	0.0460	0.0985	0.0467	0.0466	0.1668	0.0517	0.0652	0.0366	0.0819
	0.0870	0.0470	0.0990	0.0468	0.0501	0.1668	0.0539	0.0678	0.0389	0.0843
	0.0891	0.0509	0.0996	0.0484	0.0501	0.1686	0.0540	0.0690	0.0422	0.0880
	0.0894	0.0512	0.1014	0.0491	0.0502	0.1717	0.0542	0.0711	0.0424	0.0880
	0.0899	0.0522	0.1032	0.0491	0.0510	0.1724	0.0551	0.0723	0.0461	0.0881
	0.0912	0.0529	0.1038	0.0494	0.0513	0.1750	0.0552	0.0728	0.0475	0.0886
	0.0918	0.0539	0.1039	0.0495	0.0535	0.1775	0.0553	0.0742	0.0476	0.0890
	0.0928	0.0541	0.1050	0.0497	0.0536	0.1796	0.0554	0.0750	0.0481	0.0900
	0.0929	0.0541	0.1057	0.0498	0.0537	0.1820	0.0555	0.0757	0.0487	0.0902
	0.0934	0.0544	0.1072	0.0500	0.0539	0.1821	0.0559	0.0761	0.0489	0.0905
	0.0942	0.0555	0.1074	0.0505	0.0549	0.1844	0.0565	0.0764	0.0490	0.0913
	0.0954	0.0556	0.1076	0.0513	0.0550	0.1849	0.0568	0.0766	0.0496	0.0921
	0.0957	0.0556	0.1084	0.0517	0.0551	0.1852	0.0570	0.0776	0.0500	0.0943
	0.0961	0.0557	0.1086	0.0518	0.0557	0.1869	0.0574	0.0779	0.0503	0.0959
	0.0962	0.0559	0.1088	0.0520	0.0557	0.1876	0.0581	0.0783	0.0509	0.0962
	0.0970	0.0564	0.1099	0.0523	0.0559	0.1890	0.0582	0.0790	0.0510	0.0972
	0.0973	0.0565	0.1109	0.0524	0.0560	0.1904	0.0582	0.0791	0.0512	0.0972
	0.0985	0.0570	0.1111	0.0524	0.0565	0.1913	0.0582	0.0793	0.0516	0.0976
	0.1002	0.0582	0.1111	0.0526	0.0566	0.1915	0.0587	0.0794	0.0521	0.0977
	0.1006	0.0588	0.1112	0.0530	0.0573	0.1918	0.0594	0.0800	0.0525	0.0979
	0.1008	0.0590	0.1130	0.0534	0.0583	0.1930	0.0598	0.0802	0.0532	0.0986
	0.1010	0.0590	0.1130	0.0536	0.0585	0.1931	0.0612	0.0804	0.0544	0.1022
	0.1020	0.0590	0.1130	0.0537	0.0592	0.1945	0.0616	0.0808	0.0550	0.1022
	0.1024	0.0592	0.1132	0.0537	0.0602	0.1951	0.0623	0.0811	0.0556	0.1025
	0.1033	0.0595	0.1142	0.0539	0.0606	0.1991	0.0623	0.0816	0.0558	0.1026
	0.1035	0.0600	0.1157	0.0540	0.0611	0.1994	0.0624	0.0817	0.0563	0.1035
	0.1052	0.0600	0.1158	0.0547	0.0612	0.1996	0.0625	0.0817	0.0570	0.1052
	0.1054	0.0601	0.1171	0.0551	0.0619	0.2012	0.0626	0.0819	0.0571	0.1058
	0.1058	0.0606	0.1175	0.0569	0.0630	0.2013	0.0627	0.0819	0.0587	0.1075
	0.1063	0.0609	0.1185	0.0573	0.0636	0.2020	0.0630	0.0820	0.0588	0.1083
	0.1063	0.0611	0.1189	0.0579	0.0640	0.2021	0.0632	0.0826	0.0600	0.1090
	0.1083	0.0613	0.1194	0.0589	0.0640	0.2028	0.0632	0.0831	0.0611	0.1106
	0.1085	0.0620	0.1194	0.0602	0.0649	0.2038	0.0639	0.0833	0.0625	0.1132
	0.1086	0.0628	0.1212	0.0618	0.0650	0.2043	0.0665	0.0848	0.0635	0.1139
	0.1106	0.0636	0.1219	0.0640	0.0652	0.2044	0.0671	0.0853	0.0663	0.1147
	0.1119	0.0638	0.1224	0.0645	0.0675	0.2063	0.0704	0.0896	0.0668	0.1164
	0.1124	0.0664	0.1228	0.0664	0.0701	0.2100	0.0707	0.0916	0.0720	0.1168
	0.1162	0.0702	0.1268	0.0690	0.0722	0.2164	0.0712	0.0931	0.0744	0.1188
	0.1176	0.0706	0.1295	0.0726	0.0826	0.2193	0.0722	0.0954	0.0768	0.1220
r'	**5**	**2**	**2**	**1**	**1**	**1**	**2**	**4**	**4**	**3**

	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}
All samples in the group a and c	0.0611	0.1091	0.0324	0.0527	0.0568	0.0761	0.0656	0.0867	0.0616
	0.0625	0.1154	0.0343	0.0545	0.0577	0.0766	0.0672	0.0873	0.0658
	0.0634	0.1156	0.0357	0.0552	0.0586	0.0769	0.0768	0.0904	0.0674
	0.0642	0.1174	0.0363	0.0560	0.0594	0.0774	0.0778	0.0907	0.0689
	0.0643	0.1184	0.0366	0.0560	0.0597	0.0775	0.0790	0.0909	0.0696
	0.0644	0.1186	0.0372	0.0561	0.0602	0.0790	0.0829	0.0917	0.0704
	0.0658	0.1188	0.0372	0.0567	0.0603	0.0794	0.0868	0.0925	0.0706
	0.0664	0.1216	0.0373	0.0571	0.0603	0.0794	0.0875	0.0926	0.0713
	0.0666	0.1240	0.0381	0.0599	0.0607	0.0813	0.0881	0.0978	0.0728
	0.0668	0.1268	0.0386	0.0602	0.0621	0.0824	0.0892	0.0997	0.0729
	0.0676	0.1274	0.0390	0.0602	0.0624	0.0836	0.0908	0.1000	0.0739
	0.0677	0.1274	0.0395	0.0604	0.0627	0.0836	0.0916	0.1002	0.0744
	0.0683	0.1282	0.0397	0.0605	0.0627	0.0838	0.0938	0.1003	0.0757
	0.0697	0.1286	0.0398	0.0610	0.0631	0.0839	0.0946	0.1005	0.0769
	0.0698	0.1296	0.0411	0.0610	0.0635	0.0842	0.0948	0.1006	0.0771
	0.0701	0.1304	0.0413	0.0614	0.0645	0.0844	0.0965	0.1006	0.0772
	0.0706	0.1314	0.0414	0.0614	0.0649	0.0846	0.0970	0.1008	0.0773
	0.0711	0.1316	0.0414	0.0616	0.0651	0.0850	0.0982	0.1009	0.0778
	0.0714	0.1326	0.0422	0.0618	0.0660	0.0860	0.0988	0.1015	0.0781
	0.0718	0.1331	0.0424	0.0622	0.0663	0.0862	0.1008	0.1032	0.0791
	0.0719	0.1340	0.0426	0.0622	0.0665	0.0866	0.1026	0.1038	0.0795
	0.0720	0.1350	0.0429	0.0644	0.0668	0.0876	0.1028	0.1054	0.0798
	0.0725	0.1357	0.0432	0.0646	0.0669	0.0888	0.1052	0.1061	0.0804
	0.0727	0.1380	0.0439	0.0656	0.0676	0.0891	0.1059	0.1063	0.0819
	0.0730	0.1395	0.0463	0.0662	0.0676	0.0903	0.1090	0.1068	0.0819
	0.0734	0.1396	0.0466	0.0664	0.0682	0.0912	0.1096	0.1076	0.0820
	0.0736	0.1397	0.0486	0.0664	0.0697	0.0921	0.1097	0.1085	0.0828
	0.0737	0.1400	0.0497	0.0669	0.0699	0.0921	0.1099	0.1086	0.0832
	0.0740	0.1412	0.0513	0.0670	0.0699	0.0930	0.1124	0.1088	0.0835
	0.0743	0.1413	0.0528	0.0677	0.0706	0.0932	0.1127	0.1100	0.0856
	0.0744	0.1427	0.0546	0.0678	0.0717	0.0933	0.1139	0.1113	0.0868
	0.0768	0.1439	0.0550	0.0686	0.0738	0.0955	0.1142	0.1120	0.0870
	0.0768	0.1478	0.0588	0.0692	0.0744	0.0957	0.1166	0.1132	0.0874
	0.0772	0.1486	0.0592	0.0694	0.0745	0.0964	0.1182	0.1148	0.0892
	0.0778	0.1488	0.0685	0.0702	0.0745	0.0968	0.1195	0.1154	0.0894
	0.0794	0.1489	0.0685	0.0714	0.0758	0.0970	0.1215	0.1159	0.0896
	0.0840	0.1524	0.0698	0.0726	0.0758	0.1016	0.1257	0.1208	0.0902
	0.0844	0.1618	0.0718	0.0739	0.0795	0.1018	0.1312	0.1245	0.0933
	0.0873	0.1683	0.0872	0.0754	0.0829	0.1032	0.1354	0.1291	0.0947
r'	**1**	**7**	**9**	**12**	**6**	**1**	**4**	**1**	**2**

TABLE A.25: Ranked *sim* values for $(c \vee \{a, b\})$ in VHHS dataset using U-KTS $\mathcal{KH}^t$.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and b	0.0493	0.0602	0.0509	0.0370	0.0869	0.0872	0.0682	0.0677	0.0541	0.0588
	0.0504	0.0623	0.0615	0.0404	0.0876	0.0895	0.0697	0.0719	0.0581	0.0606
	0.0506	0.0636	0.0624	0.0419	0.0923	0.0919	0.0702	0.0724	0.0593	0.0609
	0.0510	0.0658	0.0641	0.0420	0.0998	0.0930	0.0742	0.0725	0.0598	0.0614
	0.0510	0.0672	0.0665	0.0424	0.0999	0.0948	0.0756	0.0740	0.0601	0.0617
	0.0512	0.0680	0.0669	0.0424	0.1007	0.0986	0.0771	0.0746	0.0620	0.0627
	0.0514	0.0681	0.0675	0.0426	0.1018	0.0996	0.0790	0.0755	0.0626	0.0627
	0.0516	0.0688	0.0678	0.0426	0.1042	0.1011	0.0811	0.0763	0.0635	0.0629
	0.0521	0.0698	0.0678	0.0428	0.1065	0.1012	0.0816	0.0765	0.0636	0.0636
	0.0528	0.0704	0.0690	0.0430	0.1076	0.1026	0.0818	0.0766	0.0640	0.0637
	0.0535	0.0705	0.0692	0.0432	0.1084	0.1029	0.0836	0.0776	0.0644	0.0642
	0.0540	0.0705	0.0696	0.0432	0.1090	0.1040	0.0841	0.0779	0.0648	0.0644
	0.0542	0.0721	0.0698	0.0434	0.1102	0.1055	0.0847	0.0781	0.0656	0.0645
	0.0545	0.0723	0.0702	0.0438	0.1124	0.1057	0.0856	0.0783	0.0661	0.0645
	0.0545	0.0726	0.0705	0.0441	0.1127	0.1062	0.0863	0.0786	0.0664	0.0647
	0.0546	0.0732	0.0706	0.0447	0.1133	0.1070	0.0866	0.0789	0.0671	0.0652
	0.0547	0.0741	0.0708	0.0448	0.1152	0.1075	0.0871	0.0792	0.0672	0.0660
	0.0553	0.0744	0.0709	0.0452	0.1154	0.1077	0.0878	0.0795	0.0677	0.0669
	0.0554	0.0744	0.0712	0.0454	0.1160	0.1081	0.0880	0.0795	0.0679	0.0672
	0.0555	0.0744	0.0714	0.0457	0.1170	0.1081	0.0881	0.0798	0.0680	0.0677
	0.0558	0.0751	0.0717	0.0464	0.1172	0.1082	0.0882	0.0800	0.0682	0.0682
	0.0560	0.0752	0.0719	0.0465	0.1175	0.1082	0.0886	0.0802	0.0682	0.0682
	0.0565	0.0753	0.0720	0.0468	0.1178	0.1085	0.0886	0.0803	0.0687	0.0685
	0.0568	0.0753	0.0720	0.0478	0.1184	0.1086	0.0889	0.0809	0.0688	0.0687
	0.0575	0.0754	0.0722	0.0480	0.1189	0.1090	0.0890	0.0811	0.0688	0.0695
	0.0575	0.0757	0.0722	0.0481	0.1189	0.1096	0.0890	0.0814	0.0690	0.0699
	0.0575	0.0758	0.0726	0.0483	0.1192	0.1099	0.0902	0.0816	0.0694	0.0700
	0.0594	0.0761	0.0735	0.0494	0.1202	0.1101	0.0909	0.0817	0.0698	0.0701
	0.0594	0.0762	0.0746	0.0495	0.1209	0.1103	0.0912	0.0817	0.0699	0.0703
	0.0598	0.0763	0.0754	0.0496	0.1215	0.1117	0.0916	0.0820	0.0715	0.0722
	0.0598	0.0764	0.0762	0.0511	0.1232	0.1138	0.0932	0.0822	0.0716	0.0732
	0.0601	0.0770	0.0763	0.0512	0.1233	0.1150	0.0939	0.0824	0.0719	0.0736
	0.0612	0.0786	0.0764	0.0546	0.1257	0.1152	0.0944	0.0825	0.0724	0.0738
	0.0642	0.0788	0.0775	0.0556	0.1261	0.1154	0.0961	0.0826	0.0726	0.0746
	0.0662	0.0798	0.0777	0.0601	0.1280	0.1160	0.0963	0.0842	0.0732	0.0746
	0.0681	0.0799	0.0794	0.0627	0.1296	0.1165	0.0965	0.0846	0.0735	0.0748
	0.0686	0.0806	0.0796	0.0634	0.1302	0.1185	0.0977	0.0852	0.0762	0.0755
	0.0733	0.0810	0.0814	0.0670	0.1354	0.1238	0.1040	0.0856	0.0772	0.0772
	0.0736	0.0827	0.0819	0.0708	0.1412	0.1255	0.1071	0.0890	0.0789	0.0796
r'	**1**	**1**	**5**	**3**	**2**	**9**	**4**	**6**	**2**	**7**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and b	0.1739	0.0739	0.1085	0.0724	0.0690	0.0563	0.0621	0.0549	0.0943	0.0482
	0.1798	0.0748	0.1112	0.0739	0.0726	0.0576	0.0724	0.0588	0.0950	0.0484
	0.1799	0.0778	0.1143	0.0788	0.0728	0.0584	0.0744	0.0601	0.0959	0.0530
	0.1810	0.0780	0.1156	0.0796	0.0776	0.0587	0.0777	0.0602	0.0983	0.0573
	0.1914	0.0810	0.1177	0.0799	0.0786	0.0593	0.0784	0.0608	0.0994	0.0586
	0.1929	0.0826	0.1204	0.0821	0.0841	0.0594	0.0788	0.0611	0.1011	0.0601
	0.1950	0.0849	0.1211	0.0823	0.0843	0.0599	0.0837	0.0624	0.1045	0.0602
	0.2011	0.0855	0.1227	0.0824	0.0849	0.0601	0.0844	0.0634	0.1075	0.0604
	0.2065	0.0856	0.1228	0.0833	0.0852	0.0606	0.0864	0.0637	0.1085	0.0607
	0.2083	0.0857	0.1238	0.0834	0.0860	0.0606	0.0876	0.0644	0.1086	0.0612
	0.2084	0.0860	0.1249	0.0859	0.0861	0.0608	0.0879	0.0650	0.1088	0.0612
	0.2100	0.0861	0.1259	0.0860	0.0864	0.0610	0.0886	0.0652	0.1090	0.0617
	0.2118	0.0862	0.1264	0.0866	0.0866	0.0613	0.0896	0.0654	0.1132	0.0618
	0.2137	0.0863	0.1270	0.0867	0.0875	0.0614	0.0899	0.0656	0.1138	0.0619
	0.2166	0.0866	0.1277	0.0892	0.0878	0.0614	0.0908	0.0656	0.1145	0.0626
	0.2167	0.0872	0.1280	0.0898	0.0880	0.0616	0.0928	0.0659	0.1152	0.0637
	0.2171	0.0876	0.1287	0.0901	0.0882	0.0616	0.0932	0.0662	0.1168	0.0637
	0.2174	0.0877	0.1306	0.0907	0.0888	0.0616	0.0935	0.0664	0.1169	0.0658
	0.2174	0.0878	0.1308	0.0909	0.0889	0.0618	0.0936	0.0666	0.1182	0.0660
	0.2176	0.0882	0.1311	0.0915	0.0891	0.0630	0.0956	0.0667	0.1182	0.0667
	0.2179	0.0882	0.1317	0.0919	0.0900	0.0636	0.0992	0.0668	0.1191	0.0670
	0.2183	0.0887	0.1323	0.0919	0.0901	0.0643	0.0995	0.0671	0.1196	0.0678
	0.2183	0.0887	0.1325	0.0920	0.0902	0.0645	0.0997	0.0678	0.1205	0.0679
	0.2208	0.0898	0.1327	0.0921	0.0904	0.0648	0.0999	0.0681	0.1212	0.0681
	0.2209	0.0899	0.1331	0.0925	0.0910	0.0654	0.1010	0.0684	0.1215	0.0682
	0.2224	0.0901	0.1336	0.0932	0.0911	0.0661	0.1034	0.0686	0.1220	0.0682
	0.2245	0.0902	0.1336	0.0934	0.0915	0.0663	0.1039	0.0688	0.1225	0.0685
	0.2254	0.0906	0.1340	0.0934	0.0920	0.0663	0.1040	0.0701	0.1234	0.0690
	0.2259	0.0908	0.1346	0.0936	0.0921	0.0664	0.1041	0.0704	0.1239	0.0694
	0.2262	0.0911	0.1347	0.0946	0.0924	0.0673	0.1092	0.0708	0.1240	0.0697
	0.2280	0.0913	0.1356	0.0952	0.0927	0.0682	0.1102	0.0712	0.1246	0.0697
	0.2290	0.0917	0.1364	0.0952	0.0938	0.0687	0.1107	0.0715	0.1282	0.0698
	0.2330	0.0922	0.1373	0.0957	0.0946	0.0690	0.1109	0.0717	0.1302	0.0698
	0.2338	0.0926	0.1432	0.0958	0.0962	0.0696	0.1156	0.0726	0.1325	0.0703
	0.2378	0.0930	0.1448	0.0976	0.0980	0.0703	0.1158	0.0733	0.1339	0.0704
	0.2388	0.0955	0.1450	0.0978	0.0980	0.0713	0.1169	0.0733	0.1363	0.0723
	0.2393	0.0971	0.1457	0.0990	0.1010	0.0713	0.1197	0.0736	0.1374	0.0732
	0.2465	0.0996	0.1507	0.1012	0.1014	0.0738	0.1209	0.0781	0.1401	0.0736
	0.2517	0.0996	0.1572	0.1056	0.1040	0.0780	0.1235	0.0805	0.1434	0.0748
r'	**5**	**6**	**1**	**8**	**3**	**1**	**11**	**2**	**13**	**2**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and b	0.0781	0.0459	0.0936	0.0438	0.0506	0.1585	0.0496	0.0640	0.0414	0.0745
	0.0844	0.0476	0.0954	0.0466	0.0510	0.1643	0.0498	0.0675	0.0438	0.0817
	0.0848	0.0476	0.0996	0.0480	0.0510	0.1681	0.0521	0.0699	0.0442	0.0853
	0.0873	0.0484	0.1024	0.0480	0.0515	0.1693	0.0534	0.0719	0.0453	0.0879
	0.0902	0.0496	0.1058	0.0483	0.0518	0.1697	0.0543	0.0720	0.0460	0.0882
	0.0916	0.0499	0.1065	0.0486	0.0522	0.1772	0.0553	0.0734	0.0464	0.0886
	0.0954	0.0502	0.1069	0.0486	0.0532	0.1792	0.0554	0.0743	0.0469	0.0911
	0.0960	0.0505	0.1070	0.0492	0.0534	0.1813	0.0554	0.0744	0.0474	0.0920
	0.0962	0.0508	0.1072	0.0492	0.0534	0.1846	0.0564	0.0745	0.0474	0.0924
	0.0965	0.0513	0.1076	0.0495	0.0538	0.1849	0.0567	0.0746	0.0475	0.0924
	0.0966	0.0516	0.1076	0.0495	0.0544	0.1852	0.0571	0.0747	0.0487	0.0955
	0.0969	0.0524	0.1098	0.0496	0.0545	0.1853	0.0572	0.0748	0.0487	0.0965
	0.0978	0.0527	0.1098	0.0496	0.0545	0.1856	0.0573	0.0750	0.0488	0.0976
	0.0982	0.0528	0.1100	0.0498	0.0547	0.1859	0.0574	0.0754	0.0491	0.0977
	0.0986	0.0534	0.1101	0.0502	0.0548	0.1860	0.0578	0.0764	0.0494	0.0991
	0.0989	0.0537	0.1102	0.0503	0.0549	0.1860	0.0580	0.0766	0.0495	0.0992
	0.0990	0.0539	0.1105	0.0507	0.0560	0.1872	0.0586	0.0770	0.0497	0.0992
	0.0992	0.0542	0.1108	0.0512	0.0564	0.1879	0.0588	0.0772	0.0498	0.0995
	0.0997	0.0547	0.1109	0.0518	0.0566	0.1884	0.0588	0.0776	0.0500	0.0998
	0.1000	0.0556	0.1117	0.0519	0.0566	0.1890	0.0590	0.0777	0.0506	0.0999
	0.1000	0.0563	0.1121	0.0524	0.0568	0.1900	0.0598	0.0787	0.0508	0.1018
	0.1007	0.0563	0.1123	0.0530	0.0572	0.1927	0.0601	0.0789	0.0510	0.1021
	0.1008	0.0564	0.1125	0.0533	0.0580	0.1935	0.0604	0.0789	0.0512	0.1028
	0.1011	0.0565	0.1129	0.0533	0.0580	0.1936	0.0608	0.0789	0.0516	0.1034
	0.1014	0.0576	0.1130	0.0535	0.0587	0.1936	0.0608	0.0790	0.0517	0.1038
	0.1018	0.0579	0.1132	0.0537	0.0587	0.1939	0.0610	0.0791	0.0524	0.1043
	0.1024	0.0580	0.1132	0.0537	0.0595	0.1940	0.0613	0.0792	0.0545	0.1045
	0.1028	0.0581	0.1142	0.0538	0.0596	0.1942	0.0616	0.0796	0.0546	0.1048
	0.1035	0.0590	0.1148	0.0545	0.0602	0.1956	0.0616	0.0801	0.0553	0.1048
	0.1035	0.0592	0.1150	0.0546	0.0602	0.1977	0.0621	0.0803	0.0557	0.1049
	0.1036	0.0595	0.1151	0.0548	0.0603	0.1985	0.0623	0.0808	0.0559	0.1054
	0.1036	0.0599	0.1152	0.0566	0.0610	0.1992	0.0625	0.0819	0.0568	0.1058
	0.1037	0.0609	0.1155	0.0574	0.0613	0.2012	0.0628	0.0831	0.0581	0.1062
	0.1050	0.0618	0.1163	0.0602	0.0616	0.2020	0.0638	0.0832	0.0591	0.1066
	0.1063	0.0627	0.1182	0.0620	0.0625	0.2024	0.0639	0.0842	0.0671	0.1087
	0.1072	0.0631	0.1206	0.0627	0.0634	0.2032	0.0646	0.0845	0.0672	0.1088
	0.1074	0.0635	0.1206	0.0632	0.0640	0.2039	0.0676	0.0850	0.0706	0.1092
	0.1108	0.0655	0.1223	0.0636	0.0642	0.2043	0.0680	0.0862	0.0708	0.1124
	0.1115	0.0682	0.1250	0.0651	0.0676	0.2160	0.0738	0.0865	0.0727	0.1173
r'	**2**	**3**	**5**	**3**	**1**	**3**	**1**	**6**	**5**	**3**

	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}
All samples in the group a and b	0.0613	0.0998	0.0304	0.0520	0.0567	0.0674	0.0684	0.0794	0.0613
	0.0618	0.1067	0.0332	0.0530	0.0576	0.0740	0.0685	0.0841	0.0621
	0.0619	0.1092	0.0335	0.0534	0.0606	0.0764	0.0803	0.0892	0.0651
	0.0625	0.1179	0.0342	0.0569	0.0614	0.0779	0.0852	0.0916	0.0707
	0.0626	0.1183	0.0345	0.0572	0.0614	0.0780	0.0853	0.0929	0.0717
	0.0626	0.1185	0.0350	0.0588	0.0616	0.0786	0.0884	0.0931	0.0717
	0.0640	0.1190	0.0353	0.0590	0.0617	0.0790	0.0884	0.0938	0.0720
	0.0649	0.1198	0.0353	0.0590	0.0619	0.0791	0.0892	0.0954	0.0729
	0.0657	0.1212	0.0354	0.0592	0.0620	0.0792	0.0901	0.0954	0.0732
	0.0663	0.1246	0.0356	0.0599	0.0624	0.0806	0.0930	0.0968	0.0732
	0.0666	0.1251	0.0366	0.0602	0.0624	0.0818	0.0941	0.0976	0.0735
	0.0675	0.1276	0.0367	0.0605	0.0627	0.0823	0.0941	0.0979	0.0737
	0.0676	0.1284	0.0369	0.0607	0.0631	0.0826	0.0950	0.0989	0.0750
	0.0676	0.1295	0.0382	0.0608	0.0633	0.0830	0.0950	0.0989	0.0769
	0.0676	0.1298	0.0383	0.0611	0.0636	0.0837	0.0954	0.1002	0.0770
	0.0680	0.1310	0.0393	0.0612	0.0637	0.0838	0.0957	0.1004	0.0782
	0.0682	0.1314	0.0405	0.0614	0.0638	0.0843	0.0970	0.1008	0.0784
	0.0686	0.1316	0.0406	0.0614	0.0639	0.0852	0.1008	0.1010	0.0788
	0.0687	0.1328	0.0415	0.0614	0.0640	0.0852	0.1020	0.1031	0.0794
	0.0688	0.1329	0.0425	0.0614	0.0646	0.0856	0.1034	0.1031	0.0797
	0.0693	0.1343	0.0426	0.0617	0.0647	0.0860	0.1038	0.1034	0.0800
	0.0697	0.1349	0.0428	0.0620	0.0653	0.0860	0.1041	0.1038	0.0803
	0.0698	0.1351	0.0430	0.0624	0.0660	0.0862	0.1042	0.1039	0.0805
	0.0701	0.1351	0.0437	0.0629	0.0662	0.0869	0.1047	0.1040	0.0808
	0.0708	0.1352	0.0439	0.0629	0.0668	0.0869	0.1050	0.1042	0.0808
	0.0710	0.1364	0.0446	0.0640	0.0668	0.0870	0.1056	0.1053	0.0808
	0.0714	0.1372	0.0452	0.0642	0.0672	0.0871	0.1057	0.1054	0.0812
	0.0715	0.1384	0.0453	0.0642	0.0672	0.0878	0.1064	0.1054	0.0820
	0.0720	0.1386	0.0466	0.0644	0.0678	0.0884	0.1082	0.1060	0.0821
	0.0738	0.1396	0.0476	0.0648	0.0684	0.0900	0.1091	0.1060	0.0832
	0.0744	0.1411	0.0484	0.0649	0.0708	0.0901	0.1100	0.1072	0.0850
	0.0745	0.1415	0.0492	0.0652	0.0711	0.0903	0.1108	0.1074	0.0856
	0.0748	0.1432	0.0501	0.0662	0.0712	0.0906	0.1112	0.1078	0.0858
	0.0761	0.1438	0.0638	0.0666	0.0713	0.0908	0.1157	0.1112	0.0862
	0.0768	0.1440	0.0645	0.0670	0.0715	0.0926	0.1164	0.1126	0.0864
	0.0773	0.1474	0.0662	0.0674	0.0721	0.0926	0.1194	0.1142	0.0868
	0.0788	0.1475	0.0681	0.0692	0.0734	0.0932	0.1206	0.1171	0.0888
	0.0794	0.1560	0.0694	0.0698	0.0740	0.0992	0.1214	0.1225	0.0910
	0.0804	0.1587	0.0719	0.0714	0.0779	0.1002	0.1308	0.1233	0.0924
r'	**3**	**4**	**9**	**11**	**3**	**2**	**1**	**4**	**1**

TABLE A.26: Ranked *sim* values for $(a \vee \{b, c\})$ in VHHS dataset using M-KTS.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group b and c	0.0288	0.0487	0.0775	0.0399	0.0948	0.0893	0.1029	0.0460	0.0379	0.0617
	0.0297	0.0517	0.0802	0.0438	0.0980	0.0993	0.1163	0.0475	0.0383	0.0657
	0.0342	0.0571	0.0837	0.0529	0.1020	0.1022	0.1221	0.0508	0.0405	0.0675
	0.0423	0.0599	0.0849	0.0550	0.1037	0.1027	0.1237	0.0533	0.0420	0.0683
	0.0425	0.0603	0.0856	0.0555	0.1038	0.1073	0.1262	0.0572	0.0466	0.0691
	0.0448	0.0627	0.0868	0.0557	0.1084	0.1097	0.1324	0.0576	0.0480	0.0723
	0.0487	0.0651	0.0873	0.0559	0.1084	0.1104	0.1325	0.0600	0.0516	0.0745
	0.0493	0.0724	0.0887	0.0565	0.1086	0.1116	0.1352	0.0639	0.0519	0.0755
	0.0512	0.0726	0.0888	0.0578	0.1103	0.1126	0.1377	0.0646	0.0610	0.0758
	0.0556	0.0729	0.0908	0.0610	0.1130	0.1127	0.1387	0.0652	0.0648	0.0763
	0.0590	0.0734	0.0926	0.0620	0.1142	0.1154	0.1406	0.0663	0.0666	0.0799
	0.0632	0.0742	0.0939	0.0649	0.1143	0.1156	0.1462	0.0677	0.0677	0.0799
	0.0647	0.0750	0.0949	0.0660	0.1150	0.1157	0.1464	0.0686	0.0685	0.0819
	0.0690	0.0781	0.0951	0.0673	0.1165	0.1173	0.1483	0.0692	0.0687	0.0842
	0.0706	0.0789	0.0966	0.0696	0.1171	0.1214	0.1488	0.0692	0.0695	0.0843
	0.0738	0.0792	0.0972	0.0711	0.1180	0.1221	0.1503	0.0722	0.0756	0.0843
	0.0750	0.0808	0.0974	0.0734	0.1184	0.1234	0.1533	0.0737	0.0781	0.0880
	0.0757	0.0821	0.0982	0.0735	0.1189	0.1251	0.1534	0.0743	0.0798	0.0904
	0.0777	0.0830	0.0982	0.0746	0.1210	0.1260	0.1549	0.0756	0.0813	0.0920
	0.0780	0.0852	0.0984	0.0753	0.1244	0.1309	0.1575	0.0772	0.0828	0.0924
	0.0829	0.0873	0.0996	0.0756	0.1249	0.1314	0.1580	0.0779	0.0831	0.0935
	0.0856	0.0879	0.0997	0.0825	0.1258	0.1321	0.1621	0.0789	0.0871	0.0942
	0.0934	0.0880	0.1013	0.0830	0.1264	0.1336	0.1657	0.0826	0.0888	0.0947
	0.0985	0.0882	0.1052	0.0832	0.1274	0.1342	0.1677	0.0838	0.0942	0.0955
	0.1001	0.0924	0.1055	0.0839	0.1276	0.1346	0.1749	0.0850	0.0961	0.0957
	0.1044	0.0941	0.1060	0.0841	0.1290	0.1410	0.1760	0.0924	0.1020	0.0963
	0.1047	0.0947	0.1114	0.0895	0.1307	0.1418	0.1773	0.0930	0.1036	0.0995
	0.1049	0.0961	0.1123	0.0986	0.1314	0.1421	0.1782	0.0965	0.1043	0.0997
	0.1235	0.0987	0.1123	0.1015	0.1328	0.1440	0.1785	0.0998	0.1204	0.0999
	0.1319	0.0991	0.1130	0.1027	0.1329	0.1447	0.1796	0.1008	0.1219	0.1016
	0.1321	0.1119	0.1144	0.1186	0.1344	0.1467	0.1818	0.1027	0.1288	0.1089
	0.1387	0.1159	0.1199	0.1195	0.1346	0.1473	0.1830	0.1069	0.1387	0.1138
	0.1440	0.1177	0.1205	0.1218	0.1346	0.1473	0.1857	0.1199	0.1397	0.1198
	0.1527	0.1209	0.1254	0.1235	0.1359	0.1580	0.1922	0.1235	0.1462	0.1236
	0.1663	0.1266	0.1258	0.1291	0.1430	0.1593	0.1936	0.1391	0.1645	0.1247
	0.1797	0.1549	0.1341	0.1358	0.1461	0.1605	0.1938	0.1413	0.1734	0.1285
	0.1837	0.1549	0.1353	0.1367	0.1500	0.1737	0.1986	0.1512	0.1735	0.1337
	0.1998	0.1667	0.1481	0.1369	0.1538	0.1771	0.2121	0.1591	0.1920	0.1380
	0.2759	0.2260	0.2056	0.2204	0.1616	0.1811	0.2176	0.2253	0.2688	0.2077
r'	**1**	**2**	**1**	**2**	**1**	**6**	**1**	**6**	**1**	**2**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group b and c	0.1259	0.0498	0.0906	0.0582	0.0499	0.0474	0.1092	0.0483	0.0548	0.0431
	0.1273	0.0528	0.0928	0.0615	0.0505	0.0510	0.1139	0.0507	0.0575	0.0505
	0.1316	0.0595	0.0932	0.0665	0.0639	0.0529	0.1155	0.0511	0.0578	0.0507
	0.1380	0.0609	0.0937	0.0672	0.0706	0.0529	0.1166	0.0525	0.0599	0.0510
	0.1395	0.0627	0.0986	0.0688	0.0730	0.0569	0.1176	0.0540	0.0609	0.0530
	0.1408	0.0643	0.1039	0.0735	0.0746	0.0599	0.1186	0.0545	0.0633	0.0541
	0.1517	0.0661	0.1046	0.0754	0.0755	0.0624	0.1197	0.0551	0.0634	0.0549
	0.1565	0.0661	0.1060	0.0785	0.0757	0.0637	0.1197	0.0586	0.0678	0.0567
	0.1596	0.0673	0.1069	0.0828	0.0793	0.0653	0.1223	0.0590	0.0695	0.0567
	0.1621	0.0697	0.1106	0.0841	0.0803	0.0674	0.1229	0.0605	0.0705	0.0569
	0.1652	0.0698	0.1114	0.0857	0.0819	0.0674	0.1277	0.0627	0.0717	0.0577
	0.1690	0.0708	0.1124	0.0859	0.0838	0.0684	0.1339	0.0640	0.0765	0.0598
	0.1696	0.0712	0.1146	0.0881	0.0844	0.0726	0.1356	0.0660	0.0768	0.0642
	0.1716	0.0712	0.1169	0.0889	0.0864	0.0737	0.1371	0.0687	0.0769	0.0650
	0.1728	0.0756	0.1185	0.0895	0.0866	0.0741	0.1378	0.0689	0.0785	0.0663
	0.1736	0.0764	0.1195	0.0897	0.0874	0.0747	0.1389	0.0699	0.0810	0.0688
	0.1778	0.0790	0.1196	0.0898	0.0894	0.0751	0.1463	0.0742	0.0811	0.0697
	0.1804	0.0791	0.1212	0.0899	0.0907	0.0752	0.1465	0.0765	0.0818	0.0738
	0.1805	0.0800	0.1242	0.0902	0.0908	0.0759	0.1515	0.0783	0.0826	0.0740
	0.1806	0.0820	0.1247	0.0902	0.0953	0.0766	0.1520	0.0784	0.0840	0.0745
	0.1830	0.0861	0.1265	0.0979	0.0954	0.0774	0.1536	0.0803	0.0842	0.0752
	0.1852	0.0866	0.1268	0.0985	0.0989	0.0778	0.1543	0.0814	0.0850	0.0758
	0.1890	0.0896	0.1286	0.0992	0.1033	0.0802	0.1546	0.0819	0.0851	0.0796
	0.1913	0.0900	0.1316	0.1013	0.1037	0.0821	0.1549	0.0855	0.0858	0.0806
	0.1915	0.0901	0.1328	0.1017	0.1038	0.0834	0.1555	0.0873	0.0869	0.0818
	0.1984	0.0951	0.1347	0.1033	0.1059	0.0841	0.1570	0.0914	0.0878	0.0834
	0.2004	0.0965	0.1350	0.1109	0.1067	0.0865	0.1577	0.0928	0.0921	0.0852
	0.2061	0.1001	0.1351	0.1120	0.1070	0.0934	0.1580	0.0981	0.0926	0.0892
	0.2062	0.1047	0.1366	0.1145	0.1086	0.0944	0.1583	0.1003	0.0934	0.1042
	0.2064	0.1060	0.1391	0.1151	0.1160	0.0953	0.1591	0.1024	0.0946	0.1055
	0.2087	0.1075	0.1409	0.1250	0.1172	0.1129	0.1601	0.1027	0.0953	0.1063
	0.2090	0.1087	0.1451	0.1255	0.1244	0.1196	0.1658	0.1097	0.1037	0.1130
	0.2116	0.1092	0.1474	0.1261	0.1270	0.1227	0.1676	0.1167	0.1056	0.1138
	0.2117	0.1126	0.1475	0.1274	0.1291	0.1236	0.1681	0.1197	0.1060	0.1413
	0.2145	0.1266	0.1536	0.1291	0.1298	0.1356	0.1697	0.1344	0.1256	0.1423
	0.2181	0.1382	0.1690	0.1370	0.1307	0.1371	0.1717	0.1401	0.1286	0.1506
	0.2239	0.1445	0.1693	0.1421	0.1377	0.1393	0.1754	0.1463	0.1310	0.1548
	0.2257	0.1521	0.1707	0.1525	0.1586	0.1512	0.1793	0.1668	0.1333	0.1559
	0.2396	0.2076	0.1794	0.2213	0.2189	0.2328	0.1875	0.2449	0.2026	0.2333
r'	**3**	**2**	**2**	**3**	**1**	**2**	**6**	**1**	**3**	**2**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group b and c	0.0512	0.0332	0.0674	0.0373	0.0451	0.1224	0.0528	0.0606	0.0387	0.0486
	0.0575	0.0429	0.0679	0.0381	0.0458	0.1332	0.0551	0.0647	0.0392	0.0550
	0.0578	0.0432	0.0701	0.0410	0.0499	0.1366	0.0570	0.0717	0.0426	0.0615
	0.0580	0.0469	0.0713	0.0424	0.0540	0.1388	0.0587	0.0793	0.0456	0.0650
	0.0582	0.0488	0.0725	0.0451	0.0545	0.1389	0.0597	0.0795	0.0457	0.0669
	0.0597	0.0489	0.0744	0.0461	0.0562	0.1407	0.0611	0.0844	0.0485	0.0672
	0.0599	0.0490	0.0773	0.0492	0.0596	0.1412	0.0626	0.0848	0.0491	0.0677
	0.0616	0.0500	0.0816	0.0505	0.0602	0.1441	0.0640	0.0859	0.0494	0.0682
	0.0647	0.0554	0.0817	0.0544	0.0604	0.1445	0.0647	0.0862	0.0551	0.0683
	0.0683	0.0559	0.0835	0.0593	0.0612	0.1458	0.0647	0.0880	0.0559	0.0699
	0.0689	0.0560	0.0842	0.0605	0.0615	0.1476	0.0665	0.0898	0.0592	0.0700
	0.0704	0.0573	0.0847	0.0611	0.0615	0.1482	0.0665	0.0901	0.0592	0.0710
	0.0717	0.0597	0.0855	0.0656	0.0616	0.1487	0.0668	0.0907	0.0597	0.0721
	0.0740	0.0629	0.0857	0.0658	0.0650	0.1495	0.0682	0.0910	0.0629	0.0724
	0.0746	0.0654	0.0873	0.0661	0.0675	0.1498	0.0689	0.0916	0.0635	0.0748
	0.0750	0.0660	0.0929	0.0685	0.0677	0.1500	0.0707	0.0916	0.0640	0.0750
	0.0765	0.0671	0.0930	0.0693	0.0689	0.1518	0.0727	0.0929	0.0643	0.0750
	0.0768	0.0704	0.0931	0.0719	0.0757	0.1523	0.0732	0.0938	0.0679	0.0755
	0.0777	0.0710	0.0936	0.0724	0.0781	0.1579	0.0732	0.0976	0.0683	0.0757
	0.0780	0.0743	0.0937	0.0741	0.0792	0.1592	0.0732	0.0981	0.0703	0.0806
	0.0780	0.0815	0.0952	0.0750	0.0793	0.1592	0.0749	0.0983	0.0760	0.0815
	0.0782	0.0876	0.0979	0.0819	0.0802	0.1597	0.0750	0.0986	0.0766	0.0842
	0.0825	0.0912	0.0992	0.0853	0.0812	0.1616	0.0802	0.1004	0.0770	0.0849
	0.0879	0.0938	0.1005	0.0877	0.0834	0.1617	0.0858	0.1016	0.0799	0.0861
	0.0885	0.0939	0.1006	0.0940	0.0836	0.1633	0.0933	0.1021	0.0838	0.0901
	0.0903	0.0993	0.1016	0.0971	0.0920	0.1635	0.0953	0.1023	0.0841	0.0908
	0.0939	0.0998	0.1026	0.1011	0.0958	0.1637	0.1018	0.1053	0.0867	0.0917
	0.0991	0.1063	0.1031	0.1032	0.1022	0.1663	0.1044	0.1059	0.0970	0.0954
	0.1009	0.1071	0.1057	0.1165	0.1120	0.1682	0.1055	0.1069	0.1017	0.0974
	0.1036	0.1085	0.1079	0.1229	0.1136	0.1702	0.1083	0.1108	0.1089	0.0975
	0.1072	0.1132	0.1141	0.1249	0.1159	0.1732	0.1154	0.1120	0.1168	0.1002
	0.1101	0.1225	0.1151	0.1271	0.1198	0.1737	0.1157	0.1129	0.1210	0.1020
	0.1111	0.1290	0.1154	0.1347	0.1259	0.1746	0.1209	0.1130	0.1269	0.1041
	0.1120	0.1311	0.1308	0.1351	0.1290	0.1747	0.1327	0.1141	0.1310	0.1071
	0.1335	0.1539	0.1311	0.1583	0.1480	0.1748	0.1368	0.1165	0.1455	0.1081
	0.1358	0.1543	0.1316	0.1658	0.1493	0.1795	0.1383	0.1180	0.1556	0.1114
	0.1401	0.1547	0.1376	0.1684	0.1595	0.1810	0.1491	0.1245	0.1587	0.1272
	0.1522	0.1819	0.1414	0.1824	0.1756	0.1824	0.1622	0.1314	0.1598	0.1280
	0.2121	0.2563	0.2137	0.2592	0.2482	0.2053	0.2220	0.1543	0.2439	0.2096
r'	**1**	**5**	**1**	**3**	**5**	**6**	**3**	**2**	**2**	**3**

	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}
All samples in the group b and c	0.0940	0.0518	0.0366	0.0518	0.0337	0.0683	0.0555	0.0597	0.0466
	0.1021	0.0591	0.0376	0.0546	0.0451	0.0804	0.0610	0.0599	0.0490
	0.1022	0.0603	0.0378	0.0546	0.0471	0.0806	0.0636	0.0653	0.0528
	0.1032	0.0603	0.0432	0.0571	0.0474	0.0808	0.0639	0.0658	0.0548
	0.1042	0.0620	0.0440	0.0578	0.0487	0.0813	0.0643	0.0665	0.0561
	0.1073	0.0642	0.0451	0.0636	0.0488	0.0819	0.0680	0.0680	0.0568
	0.1115	0.0644	0.0465	0.0638	0.0561	0.0825	0.0722	0.0685	0.0575
	0.1126	0.0648	0.0491	0.0638	0.0576	0.0828	0.0731	0.0706	0.0626
	0.1148	0.0667	0.0551	0.0643	0.0588	0.0858	0.0732	0.0711	0.0636
	0.1153	0.0671	0.0566	0.0661	0.0591	0.0874	0.0760	0.0719	0.0645
	0.1168	0.0688	0.0613	0.0662	0.0628	0.0905	0.0762	0.0721	0.0656
	0.1220	0.0691	0.0647	0.0670	0.0664	0.0918	0.0782	0.0722	0.0674
	0.1236	0.0709	0.0655	0.0672	0.0666	0.0921	0.0784	0.0752	0.0687
	0.1247	0.0714	0.0674	0.0673	0.0686	0.0964	0.0784	0.0816	0.0695
	0.1263	0.0741	0.0682	0.0701	0.0708	0.0971	0.0787	0.0832	0.0695
	0.1267	0.0749	0.0711	0.0714	0.0712	0.0978	0.0820	0.0842	0.0701
	0.1280	0.0749	0.0720	0.0718	0.0734	0.0995	0.0828	0.0850	0.0710
	0.1284	0.0762	0.0727	0.0721	0.0748	0.1001	0.0835	0.0852	0.0738
	0.1311	0.0793	0.0755	0.0735	0.0772	0.1012	0.0840	0.0861	0.0739
	0.1313	0.0824	0.0778	0.0746	0.0776	0.1037	0.0867	0.0868	0.0751
	0.1330	0.0836	0.0823	0.0748	0.0785	0.1039	0.0873	0.0872	0.0753
	0.1335	0.0846	0.0825	0.0750	0.0821	0.1049	0.0897	0.0876	0.0759
	0.1337	0.0860	0.0835	0.0757	0.0849	0.1060	0.0903	0.0880	0.0800
	0.1370	0.0874	0.0862	0.0774	0.0861	0.1060	0.0913	0.0919	0.0824
	0.1371	0.0895	0.0880	0.0860	0.0861	0.1067	0.0930	0.0927	0.0825
	0.1391	0.0897	0.0959	0.0915	0.0877	0.1081	0.0950	0.0955	0.0842
	0.1392	0.0916	0.0961	0.0936	0.0898	0.1097	0.0971	0.0959	0.0906
	0.1395	0.0950	0.1089	0.0971	0.0920	0.1117	0.0980	0.0969	0.0932
	0.1424	0.0978	0.1184	0.0986	0.0945	0.1177	0.0986	0.0991	0.1034
	0.1433	0.1041	0.1233	0.1093	0.0977	0.1183	0.1062	0.0998	0.1058
	0.1482	0.1092	0.1261	0.1115	0.1059	0.1188	0.1065	0.0998	0.1068
	0.1484	0.1111	0.1265	0.1122	0.1063	0.1206	0.1122	0.1012	0.1159
	0.1489	0.1143	0.1362	0.1198	0.1173	0.1213	0.1186	0.1021	0.1163
	0.1532	0.1176	0.1441	0.1256	0.1244	0.1253	0.1189	0.1037	0.1215
	0.1549	0.1326	0.1582	0.1383	0.1403	0.1269	0.1233	0.1178	0.1222
	0.1592	0.1329	0.1600	0.1431	0.1479	0.1284	0.1276	0.1203	0.1235
	0.1596	0.1385	0.1697	0.1431	0.1514	0.1343	0.1347	0.1294	0.1284
	0.1789	0.1546	0.1851	0.1456	0.1532	0.1452	0.1434	0.1316	0.1434
	0.1835	0.2225	0.2675	0.2248	0.2258	0.1824	0.2079	0.2082	0.2143
r'	**4**	**2**	**2**	**2**	**3**	**1**	**1**	**1**	**2**

TABLE A.27: Ranked *sim* values for $(b \vee \{a, c\})$ in VHHS dataset using M-KTS.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and c	0.0832	0.1004	0.1027	0.0681	0.1523	0.1395	0.1202	0.1157	0.0906	0.0982
	0.0847	0.1042	0.1031	0.0690	0.1559	0.1455	0.1214	0.1165	0.0938	0.0993
	0.0850	0.1056	0.1035	0.0700	0.1582	0.1472	0.1221	0.1188	0.0940	0.1015
	0.0851	0.1056	0.1050	0.0700	0.1588	0.1502	0.1221	0.1188	0.0944	0.1022
	0.0871	0.1080	0.1053	0.0711	0.1605	0.1518	0.1246	0.1191	0.0954	0.1034
	0.0872	0.1087	0.1059	0.0724	0.1621	0.1530	0.1297	0.1214	0.0956	0.1037
	0.0877	0.1105	0.1065	0.0724	0.1640	0.1542	0.1299	0.1215	0.0962	0.1038
	0.0887	0.1107	0.1073	0.0725	0.1680	0.1542	0.1311	0.1215	0.0983	0.1041
	0.0887	0.1108	0.1076	0.0727	0.1751	0.1554	0.1315	0.1230	0.0997	0.1046
	0.0888	0.1114	0.1084	0.0731	0.1772	0.1562	0.1318	0.1234	0.1013	0.1059
	0.0891	0.1117	0.1085	0.0735	0.1779	0.1564	0.1326	0.1237	0.1027	0.1060
	0.0900	0.1120	0.1088	0.0746	0.1781	0.1566	0.1335	0.1247	0.1028	0.1069
	0.0902	0.1126	0.1090	0.0747	0.1783	0.1587	0.1339	0.1252	0.1031	0.1070
	0.0904	0.1130	0.1112	0.0751	0.1786	0.1598	0.1347	0.1253	0.1031	0.1070
	0.0917	0.1149	0.1115	0.0759	0.1807	0.1614	0.1348	0.1254	0.1034	0.1073
	0.0925	0.1149	0.1118	0.0765	0.1809	0.1635	0.1349	0.1264	0.1035	0.1077
	0.0929	0.1153	0.1121	0.0774	0.1813	0.1635	0.1375	0.1265	0.1040	0.1080
	0.0930	0.1154	0.1122	0.0775	0.1818	0.1639	0.1388	0.1274	0.1040	0.1081
	0.0933	0.1156	0.1128	0.0782	0.1831	0.1639	0.1391	0.1283	0.1044	0.1082
	0.0937	0.1157	0.1130	0.0789	0.1845	0.1643	0.1411	0.1288	0.1056	0.1099
	0.0940	0.1165	0.1150	0.0791	0.1849	0.1659	0.1413	0.1291	0.1058	0.1104
	0.0944	0.1179	0.1155	0.0791	0.1850	0.1663	0.1415	0.1297	0.1064	0.1104
	0.0955	0.1179	0.1169	0.0793	0.1870	0.1675	0.1420	0.1299	0.1079	0.1107
	0.0961	0.1187	0.1172	0.0817	0.1935	0.1707	0.1425	0.1306	0.1083	0.1107
	0.0970	0.1191	0.1176	0.0820	0.1947	0.1709	0.1427	0.1325	0.1093	0.1110
	0.0975	0.1199	0.1178	0.0826	0.1958	0.1724	0.1435	0.1330	0.1104	0.1115
	0.0985	0.1209	0.1179	0.0837	0.1959	0.1724	0.1461	0.1334	0.1110	0.1121
	0.0988	0.1211	0.1179	0.0857	0.1963	0.1725	0.1468	0.1338	0.1122	0.1129
	0.0988	0.1213	0.1179	0.0873	0.1966	0.1726	0.1468	0.1345	0.1130	0.1130
	0.0990	0.1215	0.1180	0.0875	0.1975	0.1746	0.1482	0.1358	0.1132	0.1133
	0.1006	0.1233	0.1189	0.0880	0.1975	0.1753	0.1491	0.1362	0.1147	0.1154
	0.1022	0.1242	0.1197	0.0932	0.1979	0.1757	0.1503	0.1371	0.1157	0.1172
	0.1055	0.1252	0.1212	0.0934	0.2009	0.1766	0.1512	0.1372	0.1161	0.1176
	0.1058	0.1260	0.1218	0.0950	0.2024	0.1774	0.1535	0.1381	0.1169	0.1187
	0.1061	0.1260	0.1234	0.0952	0.2044	0.1788	0.1536	0.1391	0.1172	0.1190
	0.1071	0.1284	0.1245	0.1021	0.2065	0.1806	0.1545	0.1403	0.1179	0.1219
	0.1151	0.1313	0.1249	0.1071	0.2097	0.1828	0.1600	0.1425	0.1182	0.1226
	0.1165	0.1343	0.1299	0.1124	0.2181	0.1842	0.1638	0.1455	0.1273	0.1260
	0.1327	0.1363	0.1313	0.1233	0.2203	0.1934	0.1708	0.1456	0.1274	0.1391
r'	**1**	**1**	**2**	**1**	**1**	**4**	**1**	**5**	**2**	**1**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and c	0.2617	0.1247	0.1696	0.1284	0.1191	0.0953	0.1188	0.0882	0.1486	0.0910
	0.2734	0.1250	0.1772	0.1293	0.1197	0.0961	0.1239	0.0929	0.1506	0.0921
	0.2781	0.1260	0.1840	0.1325	0.1198	0.0963	0.1247	0.0956	0.1513	0.0936
	0.2814	0.1291	0.1846	0.1325	0.1233	0.0984	0.1248	0.0974	0.1551	0.0942
	0.2830	0.1292	0.1850	0.1330	0.1238	0.1002	0.1267	0.0974	0.1600	0.0942
	0.2864	0.1311	0.1853	0.1343	0.1253	0.1028	0.1272	0.0978	0.1622	0.0955
	0.2911	0.1338	0.1861	0.1346	0.1271	0.1031	0.1299	0.1002	0.1657	0.0969
	0.2915	0.1344	0.1863	0.1353	0.1275	0.1032	0.1335	0.1012	0.1712	0.0976
	0.3029	0.1359	0.1914	0.1361	0.1277	0.1032	0.1374	0.1023	0.1713	0.0977
	0.3037	0.1360	0.1915	0.1371	0.1277	0.1034	0.1389	0.1023	0.1726	0.0978
	0.3039	0.1362	0.1931	0.1379	0.1285	0.1035	0.1404	0.1031	0.1732	0.0982
	0.3043	0.1368	0.1936	0.1390	0.1290	0.1045	0.1408	0.1034	0.1734	0.0988
	0.3090	0.1370	0.1938	0.1390	0.1293	0.1050	0.1427	0.1038	0.1751	0.0996
	0.3099	0.1375	0.1953	0.1391	0.1303	0.1052	0.1429	0.1042	0.1754	0.1013
	0.3114	0.1382	0.1954	0.1394	0.1320	0.1059	0.1446	0.1051	0.1757	0.1026
	0.3122	0.1392	0.1957	0.1396	0.1327	0.1062	0.1453	0.1052	0.1757	0.1032
	0.3122	0.1398	0.1957	0.1397	0.1329	0.1063	0.1485	0.1061	0.1795	0.1032
	0.3125	0.1408	0.1969	0.1420	0.1329	0.1065	0.1510	0.1077	0.1799	0.1034
	0.3157	0.1412	0.1999	0.1421	0.1354	0.1066	0.1525	0.1083	0.1813	0.1055
	0.3164	0.1412	0.2005	0.1431	0.1363	0.1070	0.1536	0.1085	0.1818	0.1059
	0.3183	0.1413	0.2008	0.1452	0.1372	0.1076	0.1538	0.1094	0.1823	0.1064
	0.3216	0.1417	0.2010	0.1466	0.1381	0.1077	0.1548	0.1096	0.1830	0.1065
	0.3222	0.1427	0.2040	0.1475	0.1396	0.1087	0.1556	0.1099	0.1852	0.1077
	0.3228	0.1441	0.2055	0.1480	0.1410	0.1093	0.1563	0.1113	0.1859	0.1081
	0.3243	0.1444	0.2058	0.1483	0.1415	0.1095	0.1596	0.1113	0.1870	0.1085
	0.3265	0.1454	0.2072	0.1486	0.1417	0.1098	0.1600	0.1115	0.1875	0.1094
	0.3266	0.1459	0.2074	0.1492	0.1439	0.1107	0.1602	0.1121	0.1894	0.1097
	0.3271	0.1466	0.2081	0.1497	0.1440	0.1110	0.1618	0.1121	0.1915	0.1101
	0.3283	0.1467	0.2097	0.1509	0.1455	0.1133	0.1625	0.1133	0.1919	0.1101
	0.3285	0.1487	0.2102	0.1512	0.1460	0.1144	0.1639	0.1135	0.1943	0.1106
	0.3340	0.1493	0.2122	0.1514	0.1463	0.1170	0.1652	0.1153	0.1956	0.1108
	0.3362	0.1493	0.2131	0.1523	0.1465	0.1171	0.1653	0.1154	0.1960	0.1108
	0.3369	0.1500	0.2168	0.1548	0.1466	0.1193	0.1674	0.1154	0.1971	0.1111
	0.3371	0.1506	0.2188	0.1561	0.1469	0.1214	0.1682	0.1173	0.1975	0.1137
	0.3426	0.1506	0.2224	0.1570	0.1486	0.1225	0.1704	0.1174	0.1979	0.1142
	0.3474	0.1515	0.2237	0.1574	0.1490	0.1234	0.1716	0.1189	0.2025	0.1146
	0.3510	0.1552	0.2251	0.1576	0.1549	0.1246	0.1773	0.1204	0.2083	0.1159
	0.3589	0.1565	0.2285	0.1602	0.1596	0.1324	0.1855	0.1230	0.2137	0.1253
	0.3631	0.1603	0.2296	0.1649	0.1611	0.1385	0.1856	0.1264	0.2166	0.1362
r'	**1**	**4**	**3**	**1**	**2**	**3**	**5**	**1**	**2**	**2**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and c	0.1359	0.0816	0.1519	0.0765	0.0832	0.2377	0.0851	0.1108	0.0662	0.1302
	0.1360	0.0847	0.1555	0.0795	0.0843	0.2486	0.0877	0.1112	0.0663	0.1342
	0.1362	0.0847	0.1560	0.0797	0.0849	0.2499	0.0888	0.1151	0.0687	0.1374
	0.1378	0.0861	0.1566	0.0819	0.0867	0.2524	0.0892	0.1160	0.0744	0.1381
	0.1402	0.0891	0.1572	0.0824	0.0876	0.2527	0.0906	0.1161	0.0748	0.1386
	0.1409	0.0893	0.1581	0.0827	0.0877	0.2592	0.0906	0.1163	0.0755	0.1397
	0.1413	0.0896	0.1593	0.0839	0.0909	0.2594	0.0908	0.1186	0.0757	0.1401
	0.1424	0.0904	0.1595	0.0845	0.0920	0.2595	0.0910	0.1200	0.0767	0.1404
	0.1440	0.0907	0.1614	0.0847	0.0920	0.2596	0.0911	0.1209	0.0770	0.1413
	0.1446	0.0910	0.1614	0.0848	0.0921	0.2598	0.0913	0.1210	0.0771	0.1424
	0.1452	0.0915	0.1618	0.0851	0.0923	0.2622	0.0917	0.1216	0.0772	0.1424
	0.1464	0.0918	0.1627	0.0851	0.0925	0.2622	0.0925	0.1220	0.0776	0.1464
	0.1469	0.0919	0.1634	0.0855	0.0927	0.2681	0.0934	0.1225	0.0785	0.1467
	0.1477	0.0923	0.1638	0.0864	0.0930	0.2691	0.0939	0.1232	0.0786	0.1473
	0.1488	0.0923	0.1663	0.0871	0.0935	0.2692	0.0941	0.1237	0.0806	0.1477
	0.1488	0.0936	0.1664	0.0872	0.0942	0.2714	0.0949	0.1250	0.0808	0.1495
	0.1489	0.0948	0.1671	0.0873	0.0947	0.2724	0.0953	0.1251	0.0814	0.1496
	0.1493	0.0954	0.1681	0.0875	0.0961	0.2735	0.0962	0.1251	0.0818	0.1499
	0.1505	0.0955	0.1686	0.0876	0.0963	0.2739	0.0962	0.1259	0.0831	0.1501
	0.1512	0.0956	0.1691	0.0890	0.0981	0.2749	0.0964	0.1276	0.0832	0.1505
	0.1514	0.0959	0.1693	0.0890	0.0987	0.2753	0.0978	0.1278	0.0848	0.1505
	0.1527	0.0972	0.1710	0.0912	0.0990	0.2769	0.0978	0.1278	0.0852	0.1517
	0.1533	0.0976	0.1711	0.0912	0.0999	0.2770	0.0979	0.1281	0.0864	0.1541
	0.1540	0.0989	0.1713	0.0916	0.1011	0.2785	0.0980	0.1288	0.0872	0.1556
	0.1546	0.0993	0.1719	0.0925	0.1011	0.2803	0.0991	0.1288	0.0877	0.1563
	0.1557	0.0993	0.1729	0.0926	0.1025	0.2807	0.0991	0.1290	0.0879	0.1574
	0.1558	0.0998	0.1737	0.0928	0.1031	0.2824	0.1004	0.1302	0.0896	0.1596
	0.1584	0.1006	0.1739	0.0929	0.1038	0.2826	0.1009	0.1304	0.0896	0.1602
	0.1584	0.1010	0.1765	0.0936	0.1039	0.2833	0.1022	0.1304	0.0897	0.1608
	0.1586	0.1011	0.1768	0.0942	0.1054	0.2847	0.1027	0.1308	0.0907	0.1617
	0.1588	0.1011	0.1770	0.0951	0.1058	0.2859	0.1034	0.1308	0.0951	0.1622
	0.1614	0.1033	0.1775	0.0959	0.1059	0.2873	0.1055	0.1316	0.0994	0.1652
	0.1614	0.1052	0.1784	0.1000	0.1071	0.2888	0.1060	0.1324	0.1003	0.1660
	0.1625	0.1058	0.1804	0.1012	0.1072	0.2890	0.1072	0.1327	0.1046	0.1679
	0.1627	0.1059	0.1810	0.1036	0.1073	0.2909	0.1088	0.1374	0.1073	0.1712
	0.1630	0.1063	0.1812	0.1058	0.1096	0.2911	0.1106	0.1377	0.1082	0.1718
	0.1676	0.1103	0.1833	0.1076	0.1098	0.2920	0.1106	0.1399	0.1105	0.1745
	0.1699	0.1108	0.1871	0.1153	0.1156	0.3044	0.1110	0.1440	0.1197	0.1759
	0.1743	0.1181	0.1915	0.1235	0.1340	0.3047	0.1221	0.1499	0.1214	0.1763
r'	**4**	**2**	**1**	**1**	**4**	**7**	**1**	**3**	**2**	**4**

	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}
All samples in the group a and c	0.1009	0.1798	0.0560	0.0871	0.0922	0.1235	0.1206	0.1438	0.1088
	0.1032	0.1824	0.0591	0.0875	0.0925	0.1238	0.1225	0.1453	0.1106
	0.1048	0.1849	0.0594	0.0896	0.0950	0.1241	0.1239	0.1513	0.1112
	0.1056	0.1894	0.0609	0.0897	0.0951	0.1243	0.1324	0.1514	0.1140
	0.1066	0.1901	0.0619	0.0902	0.0953	0.1248	0.1391	0.1516	0.1144
	0.1072	0.1902	0.0620	0.0905	0.0969	0.1272	0.1407	0.1524	0.1151
	0.1073	0.1924	0.0625	0.0911	0.0987	0.1295	0.1431	0.1536	0.1159
	0.1081	0.1939	0.0638	0.0924	0.0992	0.1299	0.1471	0.1559	0.1160
	0.1086	0.1986	0.0662	0.0927	0.0992	0.1302	0.1471	0.1577	0.1169
	0.1089	0.1990	0.0668	0.0937	0.0994	0.1309	0.1477	0.1577	0.1172
	0.1096	0.1995	0.0674	0.0937	0.0996	0.1309	0.1503	0.1580	0.1174
	0.1107	0.2012	0.0677	0.0944	0.1011	0.1313	0.1517	0.1594	0.1190
	0.1108	0.2023	0.0684	0.0947	0.1014	0.1322	0.1523	0.1595	0.1190
	0.1131	0.2030	0.0688	0.0947	0.1015	0.1326	0.1536	0.1607	0.1200
	0.1134	0.2044	0.0692	0.0967	0.1036	0.1340	0.1541	0.1609	0.1204
	0.1143	0.2058	0.0696	0.0978	0.1036	0.1341	0.1554	0.1616	0.1219
	0.1150	0.2068	0.0697	0.0982	0.1045	0.1346	0.1560	0.1622	0.1222
	0.1156	0.2077	0.0698	0.0986	0.1048	0.1359	0.1568	0.1631	0.1224
	0.1158	0.2084	0.0698	0.0991	0.1050	0.1363	0.1584	0.1643	0.1231
	0.1160	0.2109	0.0703	0.1000	0.1060	0.1363	0.1591	0.1649	0.1233
	0.1167	0.2119	0.0708	0.1002	0.1076	0.1385	0.1608	0.1656	0.1234
	0.1170	0.2128	0.0739	0.1017	0.1082	0.1405	0.1613	0.1685	0.1243
	0.1173	0.2135	0.0765	0.1020	0.1095	0.1406	0.1631	0.1690	0.1246
	0.1174	0.2159	0.0772	0.1021	0.1098	0.1418	0.1657	0.1698	0.1258
	0.1177	0.2174	0.0788	0.1021	0.1102	0.1419	0.1705	0.1706	0.1260
	0.1179	0.2197	0.0802	0.1023	0.1103	0.1421	0.1708	0.1712	0.1260
	0.1188	0.2199	0.0805	0.1028	0.1104	0.1423	0.1710	0.1718	0.1273
	0.1190	0.2211	0.0823	0.1036	0.1115	0.1438	0.1710	0.1723	0.1285
	0.1191	0.2216	0.0869	0.1037	0.1117	0.1454	0.1721	0.1726	0.1298
	0.1193	0.2220	0.0886	0.1046	0.1125	0.1465	0.1721	0.1735	0.1301
	0.1194	0.2226	0.0893	0.1069	0.1125	0.1470	0.1746	0.1754	0.1319
	0.1213	0.2242	0.0898	0.1075	0.1130	0.1474	0.1778	0.1780	0.1334
	0.1215	0.2269	0.0931	0.1092	0.1166	0.1503	0.1779	0.1790	0.1340
	0.1234	0.2276	0.1024	0.1096	0.1182	0.1515	0.1800	0.1793	0.1343
	0.1256	0.2277	0.1081	0.1109	0.1188	0.1534	0.1805	0.1803	0.1355
	0.1292	0.2318	0.1102	0.1122	0.1220	0.1578	0.1850	0.1839	0.1373
	0.1307	0.2385	0.1129	0.1124	0.1230	0.1582	0.1878	0.1884	0.1387
	0.1320	0.2460	0.1153	0.1125	0.1236	0.1591	0.1932	0.1932	0.1393
	0.1341	0.2532	0.1398	0.1156	0.1271	0.1594	0.1992	0.1998	0.1415
r'	**2**	**3**	**5**	**2**	**1**	**1**	**2**	**2**	**3**

TABLE A.28: Ranked *sim* values for $(c \vee \{a, b\})$ in VHHS dataset using M-KTS.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and b	0.0820	0.1004	0.0968	0.0666	0.1506	0.1420	0.1209	0.1131	0.0937	0.0978
	0.0832	0.1038	0.0976	0.0674	0.1522	0.1453	0.1218	0.1140	0.0943	0.0979
	0.0840	0.1047	0.1064	0.0694	0.1604	0.1459	0.1224	0.1142	0.0950	0.1000
	0.0843	0.1090	0.1067	0.0706	0.1632	0.1487	0.1233	0.1148	0.0960	0.1005
	0.0845	0.1091	0.1077	0.0709	0.1637	0.1508	0.1263	0.1190	0.0966	0.1006
	0.0846	0.1092	0.1078	0.0710	0.1646	0.1532	0.1272	0.1193	0.0975	0.1017
	0.0860	0.1097	0.1078	0.0714	0.1685	0.1534	0.1286	0.1214	0.0986	0.1024
	0.0873	0.1100	0.1099	0.0715	0.1701	0.1551	0.1325	0.1218	0.0987	0.1024
	0.0875	0.1107	0.1100	0.0718	0.1704	0.1553	0.1326	0.1219	0.1002	0.1030
	0.0880	0.1110	0.1104	0.0720	0.1755	0.1579	0.1338	0.1223	0.1014	0.1037
	0.0881	0.1134	0.1109	0.0721	0.1758	0.1615	0.1343	0.1243	0.1024	0.1041
	0.0882	0.1140	0.1112	0.0723	0.1758	0.1616	0.1351	0.1250	0.1026	0.1048
	0.0888	0.1141	0.1112	0.0724	0.1826	0.1618	0.1358	0.1251	0.1032	0.1049
	0.0896	0.1142	0.1114	0.0724	0.1830	0.1631	0.1375	0.1252	0.1037	0.1055
	0.0901	0.1142	0.1114	0.0727	0.1834	0.1637	0.1382	0.1253	0.1038	0.1059
	0.0903	0.1147	0.1114	0.0738	0.1835	0.1641	0.1385	0.1255	0.1041	0.1062
	0.0906	0.1162	0.1116	0.0742	0.1845	0.1641	0.1392	0.1258	0.1041	0.1070
	0.0908	0.1169	0.1120	0.0748	0.1851	0.1645	0.1395	0.1259	0.1042	0.1085
	0.0910	0.1173	0.1121	0.0752	0.1865	0.1652	0.1399	0.1260	0.1045	0.1095
	0.0911	0.1176	0.1127	0.0761	0.1870	0.1652	0.1414	0.1261	0.1049	0.1095
	0.0918	0.1177	0.1129	0.0766	0.1872	0.1652	0.1422	0.1266	0.1064	0.1098
	0.0921	0.1178	0.1138	0.0766	0.1876	0.1653	0.1423	0.1267	0.1075	0.1099
	0.0922	0.1179	0.1146	0.0767	0.1884	0.1657	0.1424	0.1287	0.1076	0.1104
	0.0928	0.1179	0.1146	0.0776	0.1885	0.1661	0.1425	0.1287	0.1081	0.1104
	0.0932	0.1184	0.1151	0.0796	0.1895	0.1661	0.1430	0.1287	0.1087	0.1107
	0.0933	0.1186	0.1154	0.0797	0.1899	0.1672	0.1437	0.1291	0.1087	0.1107
	0.0941	0.1198	0.1156	0.0797	0.1907	0.1672	0.1438	0.1296	0.1088	0.1110
	0.0943	0.1199	0.1157	0.0808	0.1913	0.1674	0.1448	0.1300	0.1089	0.1112
	0.0946	0.1204	0.1160	0.0825	0.1918	0.1674	0.1448	0.1307	0.1092	0.1115
	0.0959	0.1212	0.1170	0.0854	0.1929	0.1677	0.1457	0.1320	0.1093	0.1121
	0.0962	0.1212	0.1183	0.0857	0.1931	0.1711	0.1482	0.1320	0.1096	0.1127
	0.0979	0.1212	0.1187	0.0871	0.1933	0.1729	0.1487	0.1324	0.1111	0.1144
	0.1000	0.1214	0.1191	0.0927	0.1970	0.1741	0.1498	0.1333	0.1117	0.1157
	0.1017	0.1224	0.1208	0.0982	0.1983	0.1769	0.1501	0.1333	0.1136	0.1161
	0.1031	0.1227	0.1214	0.0984	0.2002	0.1772	0.1512	0.1339	0.1151	0.1194
	0.1137	0.1243	0.1223	0.1036	0.2057	0.1787	0.1516	0.1358	0.1151	0.1205
	0.1138	0.1249	0.1231	0.1053	0.2064	0.1788	0.1545	0.1370	0.1169	0.1219
	0.1149	0.1282	0.1239	0.1129	0.2126	0.1825	0.1596	0.1375	0.1205	0.1251
	0.1175	0.1283	0.1260	0.1152	0.2142	0.1889	0.1642	0.1376	0.1263	0.1258
r'	**2**	**1**	**1**	**3**	**1**	**4**	**2**	**5**	**2**	**1**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and b	0.2664	0.1196	0.1762	0.1268	0.1162	0.0950	0.1155	0.0958	0.1509	0.0902
	0.2763	0.1221	0.1787	0.1272	0.1167	0.0969	0.1267	0.0961	0.1530	0.0913
	0.2775	0.1312	0.1796	0.1287	0.1188	0.0970	0.1281	0.0986	0.1556	0.0918
	0.2816	0.1317	0.1798	0.1301	0.1241	0.0975	0.1285	0.0988	0.1561	0.0920
	0.2867	0.1326	0.1817	0.1315	0.1242	0.0986	0.1298	0.0989	0.1567	0.0947
	0.2880	0.1336	0.1852	0.1323	0.1256	0.0995	0.1301	0.1001	0.1583	0.0959
	0.2898	0.1336	0.1855	0.1362	0.1267	0.1006	0.1382	0.1004	0.1611	0.0959
	0.2910	0.1359	0.1860	0.1363	0.1279	0.1008	0.1393	0.1009	0.1655	0.0960
	0.3022	0.1366	0.1924	0.1363	0.1288	0.1009	0.1404	0.1021	0.1696	0.0966
	0.3073	0.1366	0.1932	0.1364	0.1297	0.1017	0.1416	0.1032	0.1697	0.0976
	0.3085	0.1367	0.1939	0.1369	0.1324	0.1018	0.1422	0.1038	0.1726	0.0982
	0.3097	0.1369	0.1952	0.1382	0.1335	0.1023	0.1433	0.1040	0.1745	0.0990
	0.3101	0.1371	0.1959	0.1406	0.1339	0.1024	0.1434	0.1045	0.1755	0.0998
	0.3123	0.1372	0.1965	0.1408	0.1340	0.1033	0.1439	0.1047	0.1763	0.1001
	0.3140	0.1377	0.1969	0.1412	0.1347	0.1035	0.1444	0.1048	0.1767	0.1003
	0.3151	0.1381	0.1971	0.1412	0.1351	0.1044	0.1452	0.1050	0.1786	0.1012
	0.3163	0.1388	0.1982	0.1417	0.1352	0.1046	0.1457	0.1056	0.1789	0.1013
	0.3167	0.1395	0.1986	0.1420	0.1356	0.1048	0.1481	0.1061	0.1789	0.1019
	0.3170	0.1398	0.1992	0.1422	0.1359	0.1057	0.1495	0.1064	0.1794	0.1028
	0.3170	0.1398	0.2007	0.1443	0.1360	0.1057	0.1502	0.1065	0.1794	0.1033
	0.3184	0.1403	0.2009	0.1448	0.1364	0.1060	0.1508	0.1069	0.1795	0.1036
	0.3188	0.1403	0.2022	0.1457	0.1374	0.1063	0.1510	0.1070	0.1800	0.1047
	0.3191	0.1407	0.2025	0.1458	0.1376	0.1070	0.1530	0.1076	0.1823	0.1051
	0.3209	0.1416	0.2027	0.1464	0.1392	0.1070	0.1533	0.1077	0.1832	0.1053
	0.3216	0.1417	0.2029	0.1465	0.1395	0.1072	0.1535	0.1081	0.1839	0.1053
	0.3220	0.1420	0.2031	0.1477	0.1398	0.1091	0.1556	0.1098	0.1841	0.1054
	0.3235	0.1421	0.2032	0.1478	0.1398	0.1098	0.1561	0.1100	0.1851	0.1070
	0.3236	0.1421	0.2048	0.1479	0.1400	0.1108	0.1570	0.1103	0.1854	0.1073
	0.3259	0.1423	0.2049	0.1493	0.1405	0.1108	0.1587	0.1106	0.1874	0.1073
	0.3266	0.1430	0.2063	0.1496	0.1407	0.1109	0.1618	0.1109	0.1888	0.1075
	0.3272	0.1448	0.2072	0.1497	0.1418	0.1120	0.1639	0.1119	0.1914	0.1082
	0.3285	0.1453	0.2083	0.1507	0.1423	0.1121	0.1640	0.1121	0.1931	0.1098
	0.3353	0.1459	0.2084	0.1514	0.1425	0.1127	0.1670	0.1126	0.1940	0.1099
	0.3375	0.1466	0.2137	0.1527	0.1432	0.1169	0.1695	0.1136	0.1966	0.1101
	0.3412	0.1472	0.2151	0.1537	0.1489	0.1170	0.1706	0.1143	0.1992	0.1106
	0.3423	0.1472	0.2210	0.1541	0.1492	0.1185	0.1709	0.1145	0.1993	0.1130
	0.3461	0.1506	0.2229	0.1544	0.1493	0.1198	0.1739	0.1155	0.2043	0.1148
	0.3535	0.1510	0.2243	0.1549	0.1508	0.1202	0.1753	0.1184	0.2059	0.1155
	0.3564	0.1526	0.2310	0.1624	0.1535	0.1227	0.1772	0.1213	0.2084	0.1188
r'	**2**	**2**	**3**	**4**	**1**	**1**	**5**	**2**	**4**	**1**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and b	0.1305	0.0816	0.1495	0.0783	0.0857	0.2419	0.0820	0.1079	0.0706	0.1225
	0.1342	0.0830	0.1508	0.0801	0.0877	0.2429	0.0841	0.1123	0.0709	0.1338
	0.1355	0.0831	0.1534	0.0810	0.0880	0.2485	0.0884	0.1162	0.0712	0.1348
	0.1367	0.0838	0.1604	0.0818	0.0887	0.2507	0.0885	0.1163	0.0721	0.1384
	0.1404	0.0838	0.1613	0.0819	0.0889	0.2513	0.0892	0.1169	0.0739	0.1410
	0.1407	0.0839	0.1625	0.0828	0.0894	0.2568	0.0896	0.1190	0.0744	0.1439
	0.1439	0.0839	0.1628	0.0834	0.0896	0.2593	0.0905	0.1192	0.0749	0.1442
	0.1453	0.0848	0.1643	0.0836	0.0908	0.2603	0.0911	0.1199	0.0750	0.1451
	0.1458	0.0861	0.1644	0.0839	0.0911	0.2606	0.0913	0.1211	0.0753	0.1463
	0.1466	0.0865	0.1647	0.0840	0.0914	0.2617	0.0919	0.1211	0.0766	0.1464
	0.1479	0.0883	0.1649	0.0844	0.0923	0.2644	0.0929	0.1212	0.0771	0.1474
	0.1479	0.0885	0.1653	0.0844	0.0928	0.2658	0.0931	0.1216	0.0773	0.1483
	0.1481	0.0892	0.1653	0.0851	0.0929	0.2667	0.0934	0.1219	0.0774	0.1496
	0.1490	0.0894	0.1659	0.0855	0.0931	0.2678	0.0936	0.1222	0.0782	0.1509
	0.1491	0.0904	0.1659	0.0858	0.0936	0.2681	0.0942	0.1226	0.0782	0.1513
	0.1492	0.0904	0.1661	0.0858	0.0941	0.2685	0.0942	0.1226	0.0788	0.1515
	0.1494	0.0907	0.1673	0.0874	0.0941	0.2701	0.0943	0.1229	0.0804	0.1518
	0.1498	0.0908	0.1673	0.0877	0.0944	0.2704	0.0944	0.1236	0.0814	0.1525
	0.1499	0.0924	0.1674	0.0879	0.0950	0.2718	0.0944	0.1237	0.0815	0.1531
	0.1501	0.0943	0.1674	0.0886	0.0954	0.2735	0.0962	0.1237	0.0815	0.1531
	0.1502	0.0948	0.1681	0.0888	0.0955	0.2737	0.0963	0.1238	0.0821	0.1534
	0.1507	0.0952	0.1683	0.0888	0.0957	0.2738	0.0966	0.1250	0.0829	0.1540
	0.1508	0.0954	0.1683	0.0889	0.0960	0.2748	0.0968	0.1253	0.0832	0.1548
	0.1523	0.0960	0.1694	0.0895	0.0963	0.2749	0.0970	0.1254	0.0839	0.1550
	0.1530	0.0964	0.1704	0.0896	0.0966	0.2751	0.0971	0.1256	0.0845	0.1555
	0.1530	0.0966	0.1706	0.0907	0.0972	0.2773	0.0974	0.1259	0.0859	0.1565
	0.1535	0.0967	0.1716	0.0907	0.0976	0.2778	0.0980	0.1271	0.0872	0.1566
	0.1538	0.0976	0.1718	0.0914	0.0979	0.2784	0.0988	0.1276	0.0875	0.1569
	0.1551	0.0978	0.1730	0.0916	0.0987	0.2796	0.0996	0.1280	0.0876	0.1573
	0.1552	0.0983	0.1730	0.0916	0.0992	0.2810	0.1004	0.1280	0.0880	0.1603
	0.1553	0.0985	0.1731	0.0931	0.1006	0.2812	0.1013	0.1286	0.0905	0.1605
	0.1554	0.0994	0.1737	0.0934	0.1008	0.2812	0.1017	0.1299	0.0913	0.1619
	0.1557	0.0999	0.1739	0.0988	0.1011	0.2865	0.1035	0.1306	0.0983	0.1619
	0.1584	0.1008	0.1739	0.1005	0.1017	0.2887	0.1039	0.1307	0.1001	0.1621
	0.1584	0.1008	0.1769	0.1010	0.1026	0.2888	0.1052	0.1324	0.1002	0.1622
	0.1592	0.1044	0.1786	0.1037	0.1052	0.2896	0.1083	0.1332	0.1027	0.1623
	0.1615	0.1092	0.1811	0.1057	0.1059	0.2897	0.1087	0.1337	0.1054	0.1634
	0.1633	0.1114	0.1818	0.1092	0.1064	0.2900	0.1093	0.1339	0.1104	0.1636
	0.1645	0.1157	0.1827	0.1192	0.1076	0.3001	0.1180	0.1353	0.1117	0.1711
r'	**2**	**3**	**2**	**2**	**4**	**4**	**1**	**3**	**2**	**3**

	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}
All samples in the group a and b	0.0997	0.1673	0.0554	0.0821	0.0950	0.1170	0.1215	0.1449	0.1075
	0.1015	0.1757	0.0561	0.0848	0.0957	0.1211	0.1239	0.1453	0.1117
	0.1031	0.1827	0.0568	0.0909	0.0980	0.1262	0.1323	0.1497	0.1129
	0.1034	0.1855	0.0581	0.0918	0.0986	0.1263	0.1370	0.1505	0.1164
	0.1037	0.1873	0.0584	0.0937	0.0996	0.1284	0.1385	0.1507	0.1187
	0.1037	0.1909	0.0589	0.0938	0.0997	0.1287	0.1389	0.1517	0.1192
	0.1048	0.1912	0.0590	0.0944	0.1000	0.1288	0.1407	0.1542	0.1192
	0.1056	0.1931	0.0603	0.0945	0.1001	0.1301	0.1480	0.1553	0.1204
	0.1063	0.1932	0.0609	0.0948	0.1009	0.1306	0.1495	0.1559	0.1205
	0.1067	0.2000	0.0618	0.0950	0.1011	0.1311	0.1500	0.1591	0.1214
	0.1075	0.2016	0.0624	0.0950	0.1017	0.1331	0.1500	0.1592	0.1215
	0.1079	0.2028	0.0635	0.0952	0.1023	0.1331	0.1506	0.1594	0.1215
	0.1093	0.2037	0.0663	0.0954	0.1024	0.1334	0.1509	0.1616	0.1216
	0.1096	0.2057	0.0667	0.0956	0.1024	0.1341	0.1524	0.1619	0.1221
	0.1100	0.2070	0.0681	0.0962	0.1028	0.1341	0.1536	0.1628	0.1228
	0.1108	0.2090	0.0690	0.0965	0.1031	0.1342	0.1536	0.1635	0.1229
	0.1108	0.2095	0.0691	0.0965	0.1039	0.1353	0.1549	0.1643	0.1231
	0.1116	0.2106	0.0692	0.0971	0.1039	0.1360	0.1596	0.1645	0.1231
	0.1116	0.2109	0.0703	0.0978	0.1039	0.1361	0.1613	0.1645	0.1231
	0.1121	0.2109	0.0722	0.0980	0.1039	0.1365	0.1629	0.1650	0.1233
	0.1129	0.2120	0.0728	0.0981	0.1050	0.1366	0.1630	0.1651	0.1237
	0.1138	0.2121	0.0732	0.0983	0.1053	0.1371	0.1632	0.1651	0.1242
	0.1143	0.2122	0.0734	0.0986	0.1060	0.1373	0.1634	0.1657	0.1246
	0.1145	0.2125	0.0740	0.0989	0.1061	0.1383	0.1642	0.1666	0.1247
	0.1148	0.2131	0.0746	0.0994	0.1066	0.1386	0.1645	0.1686	0.1248
	0.1157	0.2141	0.0749	0.0997	0.1072	0.1388	0.1647	0.1689	0.1251
	0.1164	0.2150	0.0763	0.1009	0.1075	0.1389	0.1655	0.1692	0.1260
	0.1167	0.2157	0.0785	0.1028	0.1088	0.1396	0.1671	0.1710	0.1262
	0.1176	0.2173	0.0789	0.1029	0.1093	0.1405	0.1682	0.1710	0.1285
	0.1185	0.2186	0.0806	0.1031	0.1101	0.1406	0.1687	0.1713	0.1289
	0.1189	0.2189	0.0815	0.1032	0.1108	0.1413	0.1690	0.1718	0.1295
	0.1210	0.2206	0.0883	0.1038	0.1132	0.1421	0.1705	0.1721	0.1297
	0.1213	0.2206	0.0913	0.1044	0.1141	0.1422	0.1731	0.1746	0.1298
	0.1217	0.2261	0.1042	0.1052	0.1146	0.1437	0.1732	0.1748	0.1310
	0.1218	0.2264	0.1045	0.1060	0.1148	0.1440	0.1776	0.1800	0.1327
	0.1222	0.2296	0.1086	0.1072	0.1152	0.1477	0.1781	0.1813	0.1330
	0.1265	0.2303	0.1098	0.1079	0.1156	0.1491	0.1823	0.1834	0.1343
	0.1304	0.2396	0.1157	0.1087	0.1176	0.1517	0.1836	0.1902	0.1358
	0.1315	0.2434	0.1181	0.1097	0.1199	0.1541	0.1905	0.1910	0.1371
r'	**2**	**2**	**2**	**1**	**3**	**2**	**2**	**1**	**2**

TABLE A.29: Ranked *sim* values for $(a \vee \{b, c\})$ in VHHS dataset using U-KTS FVR.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group b and c	0.7097	0.6481	0.8791	0.4589	0.7008	0.6752	0.4874	0.5155	0.7263	0.5649
	0.6344	0.6389	0.7985	0.4038	0.6948	0.6547	0.4452	0.5152	0.6782	0.5428
	0.6246	0.6308	0.7739	0.3785	0.6447	0.6198	0.3967	0.4393	0.6726	0.5221
	0.6113	0.5927	0.6253	0.3628	0.6214	0.6020	0.3945	0.4292	0.6618	0.4240
	0.6041	0.5146	0.5819	0.3401	0.5954	0.5772	0.3837	0.4276	0.6497	0.4198
	0.5900	0.4714	0.5230	0.3385	0.5748	0.4911	0.3770	0.4046	0.6477	0.4184
	0.5819	0.4609	0.5075	0.3326	0.5641	0.4752	0.3604	0.3998	0.6188	0.4120
	0.5804	0.4363	0.4560	0.3317	0.5421	0.4498	0.3568	0.3838	0.6078	0.4105
	0.5627	0.4075	0.4517	0.3211	0.5412	0.4444	0.3557	0.3798	0.5975	0.3912
	0.5586	0.3820	0.4491	0.3206	0.5350	0.4224	0.3516	0.3778	0.5952	0.3790
	0.5315	0.3763	0.4334	0.3204	0.5186	0.4059	0.3505	0.3724	0.5916	0.3693
	0.5291	0.3408	0.4299	0.3093	0.5128	0.3991	0.3505	0.3698	0.5825	0.3561
	0.5263	0.3333	0.4290	0.3091	0.5001	0.3951	0.3259	0.3583	0.5759	0.3410
	0.5156	0.3317	0.4282	0.3081	0.4983	0.3208	0.3245	0.3575	0.5551	0.3288
	0.4997	0.3222	0.4226	0.2998	0.4962	0.3178	0.3157	0.3565	0.5543	0.3253
	0.4886	0.3177	0.4198	0.2988	0.4950	0.3135	0.3106	0.3551	0.5518	0.3239
	0.4825	0.3061	0.4168	0.2979	0.4908	0.3110	0.3070	0.3541	0.5063	0.3198
	0.4730	0.2995	0.4049	0.2873	0.4715	0.3102	0.3023	0.3503	0.4951	0.3190
	0.4554	0.2816	0.3943	0.2796	0.4690	0.3065	0.3015	0.3474	0.4924	0.3156
	0.4549	0.2737	0.3886	0.2616	0.4597	0.3037	0.2921	0.3464	0.4907	0.3137
	0.4439	0.2737	0.3856	0.2481	0.4587	0.2963	0.2829	0.3397	0.4841	0.3070
	0.4433	0.2707	0.3762	0.2466	0.4440	0.2958	0.2818	0.3377	0.4825	0.2990
	0.4395	0.2605	0.3628	0.2377	0.4172	0.2839	0.2808	0.3292	0.4799	0.2972
	0.4383	0.2565	0.3601	0.2370	0.4170	0.2772	0.2739	0.3249	0.4785	0.2925
	0.4369	0.2532	0.3555	0.2332	0.4053	0.2689	0.2662	0.3237	0.4768	0.2835
	0.4310	0.2473	0.3543	0.2309	0.3938	0.2647	0.2588	0.3154	0.4731	0.2773
	0.4230	0.2426	0.3523	0.2280	0.3928	0.2640	0.2410	0.3126	0.4635	0.2755
	0.4186	0.2335	0.3506	0.2267	0.3897	0.2593	0.2398	0.3004	0.4590	0.2728
	0.4143	0.2331	0.3443	0.2234	0.3763	0.2590	0.2378	0.2941	0.4583	0.2644
	0.4113	0.2314	0.3363	0.2210	0.3756	0.2485	0.2350	0.2938	0.4562	0.2616
	0.4032	0.2303	0.3315	0.2198	0.3691	0.2477	0.2281	0.2880	0.4452	0.2611
	0.3994	0.2289	0.3266	0.2192	0.3604	0.2438	0.2241	0.2751	0.4418	0.2411
	0.3974	0.2219	0.3180	0.2177	0.3582	0.2275	0.2183	0.2690	0.4393	0.2351
	0.3916	0.2218	0.3152	0.2167	0.3565	0.2243	0.2181	0.2657	0.4294	0.2305
	0.3848	0.2156	0.3092	0.2140	0.3470	0.2208	0.2116	0.2645	0.4076	0.2217
	0.3524	0.1976	0.3035	0.2137	0.3401	0.2154	0.2110	0.2398	0.3989	0.2087
	0.3475	0.1900	0.3021	0.2123	0.3295	0.2140	0.2046	0.2380	0.3799	0.2082
	0.3391	0.1807	0.3020	0.2062	0.3197	0.2082	0.2004	0.2297	0.3756	0.2062
	0.2850	0.1607	0.2852	0.1710	0.2799	0.1982	0.1572	0.2258	0.3675	0.2044
r'	**3**	**6**	**1**	**7**	**10**	**9**	**20**	**12**	**21**	**1**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group b and c	0.8914	0.7232	0.7591	0.6213	0.4463	0.7730	0.7485	0.7569	0.8367	0.7970
	0.8076	0.5401	0.7395	0.6112	0.4035	0.7263	0.6474	0.6975	0.6822	0.7046
	0.5852	0.5371	0.7108	0.5811	0.3958	0.6579	0.6102	0.6599	0.6470	0.6386
	0.5639	0.5334	0.6931	0.5800	0.3386	0.6230	0.5936	0.6530	0.5872	0.5939
	0.5584	0.5079	0.6919	0.5648	0.3362	0.5874	0.5924	0.6444	0.5846	0.5275
	0.5339	0.5017	0.6917	0.5526	0.3348	0.5869	0.5555	0.6390	0.5788	0.5071
	0.4573	0.4997	0.6301	0.5406	0.3344	0.5388	0.5239	0.6117	0.5753	0.4910
	0.4444	0.4644	0.6221	0.5361	0.3307	0.5363	0.5087	0.6039	0.5530	0.4800
	0.4334	0.4130	0.6111	0.5348	0.3250	0.5172	0.4946	0.5842	0.5481	0.4795
	0.4209	0.4006	0.6080	0.5236	0.3146	0.5168	0.4834	0.5751	0.5233	0.4719
	0.4168	0.3993	0.5993	0.5061	0.3057	0.5032	0.4739	0.5550	0.5202	0.4654
	0.4042	0.3934	0.5945	0.4959	0.2986	0.4780	0.4540	0.5476	0.5036	0.4613
	0.3940	0.3933	0.5723	0.4850	0.2926	0.4418	0.4356	0.5415	0.4992	0.4587
	0.3896	0.3876	0.5704	0.4678	0.2904	0.4415	0.4349	0.5412	0.4990	0.4445
	0.3765	0.3770	0.5666	0.4586	0.2867	0.4329	0.4279	0.5347	0.4833	0.4416
	0.3753	0.3626	0.5495	0.4483	0.2753	0.4310	0.4270	0.5244	0.4729	0.4409
	0.3627	0.3599	0.5391	0.4393	0.2690	0.4273	0.4224	0.5119	0.4628	0.4394
	0.3566	0.3491	0.5364	0.4340	0.2668	0.4078	0.4056	0.5061	0.4445	0.4388
	0.3551	0.3230	0.5328	0.4170	0.2660	0.4072	0.4046	0.5021	0.4404	0.4333
	0.3418	0.3217	0.5205	0.4102	0.2571	0.4069	0.4033	0.4970	0.4394	0.4111
	0.3224	0.3171	0.5190	0.4082	0.2553	0.3984	0.3872	0.4902	0.4300	0.3938
	0.3220	0.3134	0.5072	0.3846	0.2542	0.3977	0.3868	0.4740	0.4282	0.3913
	0.3148	0.3127	0.5062	0.3805	0.2529	0.3932	0.3694	0.4657	0.4282	0.3908
	0.3118	0.3122	0.5058	0.3755	0.2528	0.3905	0.3687	0.4544	0.4102	0.3824
	0.3103	0.3097	0.4977	0.3683	0.2511	0.3879	0.3683	0.4450	0.4059	0.3748
	0.2986	0.3013	0.4802	0.3655	0.2466	0.3631	0.3582	0.4151	0.3915	0.3706
	0.2839	0.2922	0.4796	0.3446	0.2403	0.3557	0.3571	0.4109	0.3864	0.3660
	0.2817	0.2896	0.4695	0.3310	0.2381	0.3543	0.3539	0.4072	0.3860	0.3614
	0.2753	0.2655	0.4674	0.3303	0.2246	0.3421	0.3398	0.3934	0.3846	0.3513
	0.2679	0.2598	0.4634	0.3167	0.2137	0.3394	0.3392	0.3744	0.3768	0.3482
	0.2669	0.2594	0.4573	0.2944	0.2119	0.3225	0.3299	0.3738	0.3757	0.3362
	0.2608	0.2503	0.4479	0.2938	0.2113	0.3163	0.3287	0.3715	0.3741	0.3341
	0.2570	0.2500	0.4369	0.2915	0.1978	0.3144	0.3279	0.3521	0.3676	0.3326
	0.2564	0.2431	0.4081	0.2833	0.1939	0.3059	0.3142	0.3456	0.3548	0.3319
	0.2519	0.2299	0.4065	0.2444	0.1898	0.2908	0.3130	0.3430	0.3458	0.3285
	0.2443	0.2287	0.4045	0.2320	0.1847	0.2775	0.3122	0.3330	0.3175	0.3248
	0.2427	0.2274	0.3489	0.2289	0.1738	0.2544	0.2946	0.3221	0.3012	0.3165
	0.2313	0.2224	0.3242	0.2242	0.1543	0.2401	0.2925	0.2743	0.2969	0.3034
	0.2275	0.2153	0.3180	0.2192	0.1483	0.2300	0.2062	0.2316	0.2796	0.2889
r'	**2**	**1**	**3**	**10**	**13**	**9**	**2**	**7**	**11**	**16**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group b and c	0.7605	0.7855	0.7917	0.7368	0.7365	0.5567	0.6091	0.7205	0.6854	0.7677
	0.7107	0.6943	0.7342	0.6778	0.6198	0.4898	0.4752	0.7172	0.6568	0.7251
	0.7051	0.6911	0.7197	0.6759	0.5537	0.4737	0.4598	0.6542	0.5768	0.7181
	0.6992	0.6207	0.6831	0.6639	0.5318	0.4642	0.4251	0.6479	0.5507	0.6710
	0.6486	0.5954	0.6830	0.6310	0.5129	0.4572	0.4228	0.6451	0.5238	0.6267
	0.6132	0.5778	0.6439	0.5994	0.5121	0.4434	0.4222	0.5867	0.5235	0.6262
	0.6011	0.5257	0.6264	0.5642	0.5015	0.4209	0.4008	0.5772	0.4500	0.6091
	0.5905	0.5045	0.6172	0.5542	0.4959	0.4174	0.3997	0.5678	0.4094	0.6054
	0.5610	0.5021	0.6165	0.5421	0.4954	0.4157	0.3995	0.5533	0.4025	0.6041
	0.5596	0.4997	0.6094	0.5069	0.4949	0.4100	0.3957	0.5483	0.4020	0.5992
	0.5583	0.4948	0.6021	0.4480	0.4936	0.4027	0.3918	0.5429	0.3941	0.5755
	0.5533	0.4923	0.5751	0.4447	0.4893	0.3976	0.3917	0.5254	0.3911	0.5753
	0.5519	0.4763	0.5677	0.4430	0.4869	0.3912	0.3901	0.5249	0.3887	0.5743
	0.5146	0.4748	0.5663	0.4379	0.4807	0.3855	0.3895	0.5127	0.3769	0.5554
	0.5093	0.4704	0.5653	0.4115	0.4798	0.3820	0.3811	0.5066	0.3714	0.5520
	0.4834	0.4515	0.5586	0.4053	0.4731	0.3740	0.3685	0.4775	0.3637	0.5479
	0.4833	0.4459	0.5380	0.4051	0.4490	0.3729	0.3614	0.4585	0.3611	0.5421
	0.4750	0.4379	0.5365	0.4011	0.4414	0.3725	0.3588	0.4518	0.3459	0.5412
	0.4691	0.4126	0.5196	0.3986	0.4349	0.3700	0.3456	0.4517	0.3434	0.5373
	0.4546	0.4123	0.5053	0.3734	0.4233	0.3685	0.3374	0.4340	0.3354	0.5345
	0.4467	0.4002	0.5008	0.3549	0.4131	0.3676	0.3344	0.4257	0.3325	0.5183
	0.4462	0.3907	0.4982	0.3497	0.4099	0.3611	0.3276	0.4197	0.3295	0.4991
	0.4460	0.3829	0.4966	0.3476	0.4087	0.3560	0.3259	0.4126	0.3294	0.4866
	0.4342	0.3632	0.4940	0.3333	0.4064	0.3444	0.3255	0.4118	0.3273	0.4860
	0.4327	0.3622	0.4906	0.3225	0.4052	0.3330	0.3056	0.4102	0.3251	0.4769
	0.4272	0.3592	0.4888	0.3213	0.3989	0.3299	0.2981	0.4075	0.3131	0.4622
	0.4248	0.3478	0.4846	0.3095	0.3961	0.3245	0.2962	0.3984	0.3003	0.4508
	0.4184	0.3463	0.4836	0.3093	0.3914	0.3084	0.2888	0.3923	0.2989	0.4493
	0.4126	0.3395	0.4809	0.3003	0.3907	0.3077	0.2817	0.3849	0.2977	0.4470
	0.4120	0.3344	0.4764	0.2919	0.3587	0.3074	0.2810	0.3835	0.2892	0.4449
	0.4074	0.3288	0.4712	0.2872	0.3562	0.3041	0.2737	0.3730	0.2832	0.4393
	0.4049	0.3204	0.4690	0.2845	0.3490	0.2907	0.2605	0.3636	0.2547	0.4361
	0.3814	0.3063	0.4583	0.2805	0.3426	0.2852	0.2546	0.3248	0.2503	0.4283
	0.3793	0.3051	0.4525	0.2758	0.3339	0.2693	0.2447	0.3225	0.2428	0.4210
	0.3688	0.3048	0.4494	0.2629	0.3275	0.2658	0.2427	0.3218	0.2405	0.4180
	0.3631	0.3023	0.4233	0.2622	0.3170	0.2433	0.2366	0.3183	0.2368	0.4078
	0.3564	0.2840	0.4104	0.2579	0.2985	0.2374	0.2366	0.3098	0.2342	0.3778
	0.3514	0.2791	0.3818	0.2469	0.2941	0.2318	0.2221	0.2769	0.2242	0.3717
	0.2911	0.2217	0.3222	0.2427	0.2685	0.2294	0.2164	0.2518	0.2114	0.3615
r'	**1**	**4**	**8**	**1**	**10**	**1**	**20**	**4**	**14**	**2**

	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}
All samples in the group b and c	0.4953	0.6140	0.7390	0.6485	0.5922	0.8291	0.5368	0.8331	0.8514
	0.4925	0.6044	0.6342	0.5972	0.4843	0.7965	0.5211	0.7765	0.5885
	0.4860	0.5457	0.5592	0.5139	0.4773	0.5979	0.4986	0.7626	0.5867
	0.4741	0.5341	0.5461	0.5125	0.4720	0.5796	0.4160	0.7526	0.4910
	0.4003	0.5157	0.5136	0.4563	0.4554	0.5643	0.4158	0.7325	0.4513
	0.4001	0.4864	0.4979	0.4504	0.4493	0.5549	0.4028	0.6863	0.4354
	0.3816	0.4857	0.4932	0.4442	0.4451	0.5293	0.3882	0.6769	0.4201
	0.3695	0.4841	0.4875	0.4427	0.4444	0.5106	0.3754	0.6596	0.4162
	0.3656	0.4780	0.4786	0.4290	0.4437	0.4905	0.3743	0.6500	0.4064
	0.3538	0.4760	0.4777	0.4283	0.4159	0.4803	0.3726	0.6368	0.4027
	0.3508	0.4707	0.4569	0.4262	0.4117	0.4778	0.3605	0.6267	0.4020
	0.3438	0.4701	0.4423	0.4103	0.4094	0.4510	0.3515	0.6246	0.4011
	0.3339	0.4623	0.4394	0.3978	0.4079	0.4427	0.3448	0.6031	0.3942
	0.3142	0.4531	0.4358	0.3854	0.4053	0.4350	0.3440	0.5994	0.3928
	0.3126	0.4501	0.4172	0.3713	0.3935	0.4276	0.3329	0.5978	0.3904
	0.3125	0.4436	0.4164	0.3671	0.3902	0.4246	0.3231	0.5930	0.3829
	0.3085	0.4292	0.4026	0.3625	0.3859	0.4241	0.3114	0.5900	0.3815
	0.3066	0.4115	0.4002	0.3590	0.3610	0.4108	0.3041	0.5837	0.3679
	0.3032	0.4027	0.3980	0.3552	0.3570	0.4084	0.2999	0.5762	0.3618
	0.3002	0.4010	0.3941	0.3500	0.3529	0.4075	0.2864	0.5531	0.3529
	0.2944	0.3933	0.3896	0.3453	0.3497	0.3996	0.2824	0.5486	0.3471
	0.2863	0.3607	0.3781	0.3442	0.3283	0.3987	0.2745	0.5476	0.3361
	0.2838	0.3581	0.3726	0.3425	0.3174	0.3987	0.2735	0.5472	0.3320
	0.2788	0.3535	0.3670	0.3347	0.3170	0.3985	0.2689	0.5462	0.3318
	0.2780	0.3382	0.3596	0.3218	0.3137	0.3936	0.2678	0.5447	0.3280
	0.2705	0.3356	0.3593	0.3144	0.3109	0.3913	0.2668	0.5443	0.3114
	0.2689	0.3334	0.3593	0.3140	0.3078	0.3893	0.2651	0.5337	0.3106
	0.2688	0.3325	0.3586	0.3074	0.3062	0.3826	0.2648	0.5288	0.2983
	0.2659	0.3292	0.3585	0.3004	0.3037	0.3671	0.2600	0.5138	0.2862
	0.2634	0.3285	0.3458	0.2815	0.2869	0.3666	0.2536	0.5095	0.2836
	0.2616	0.3240	0.3391	0.2774	0.2743	0.3631	0.2533	0.5068	0.2786
	0.2581	0.3240	0.3384	0.2721	0.2714	0.3631	0.2498	0.5002	0.2761
	0.2556	0.3173	0.3288	0.2568	0.2467	0.3563	0.2485	0.4984	0.2756
	0.2455	0.3145	0.3275	0.2527	0.2452	0.3545	0.2470	0.4825	0.2671
	0.2454	0.3064	0.3097	0.2424	0.2378	0.3463	0.2386	0.4792	0.2666
	0.2444	0.3063	0.3089	0.2345	0.2299	0.3392	0.2365	0.4654	0.2619
	0.2227	0.3008	0.2897	0.2263	0.2279	0.3363	0.2199	0.4544	0.2600
	0.2129	0.2734	0.2622	0.2243	0.2221	0.3334	0.1980	0.4468	0.2417
	0.1938	0.2678	0.2178	0.2169	0.2161	0.3318	0.1909	0.4411	0.1897
r'	**3**	**8**	**2**	**6**	**3**	**8**	**11**	**1**	**12**

TABLE A.30: Ranked *sim* values for $(b \vee \{a, c\})$ in VHHS dataset using FVR.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and c	0.5859	0.5913	0.6681	0.5148	0.6605	0.8208	0.8182	0.5143	0.7187	0.7927
	0.5551	0.5347	0.6590	0.4726	0.6269	0.7440	0.6276	0.4886	0.6213	0.7036
	0.5131	0.5181	0.6361	0.4405	0.6060	0.5555	0.6200	0.4748	0.5370	0.6921
	0.5124	0.4983	0.6130	0.4338	0.6038	0.5162	0.6141	0.4565	0.5370	0.6816
	0.5055	0.4690	0.6072	0.4243	0.5864	0.5069	0.6117	0.4497	0.5100	0.6669
	0.4888	0.4634	0.5902	0.4139	0.5727	0.4512	0.6108	0.4361	0.5080	0.6540
	0.4510	0.4568	0.5743	0.4111	0.5715	0.4284	0.5765	0.4300	0.5042	0.6467
	0.4425	0.4550	0.5197	0.3951	0.5692	0.3913	0.5689	0.4089	0.4648	0.6421
	0.4402	0.4016	0.5174	0.3800	0.5664	0.3872	0.5333	0.3992	0.4640	0.5999
	0.4368	0.3824	0.5077	0.3704	0.5653	0.3825	0.5313	0.3895	0.4614	0.5999
	0.4289	0.3718	0.4881	0.3685	0.5583	0.3798	0.5167	0.3855	0.4469	0.5910
	0.4255	0.3698	0.4754	0.3613	0.5505	0.3638	0.5160	0.3781	0.4429	0.5874
	0.4219	0.3657	0.4698	0.3501	0.5356	0.3631	0.5111	0.3726	0.4333	0.5737
	0.4187	0.3635	0.4686	0.3387	0.5232	0.3587	0.5054	0.3616	0.4311	0.5725
	0.4176	0.3339	0.4682	0.3357	0.5228	0.3505	0.5001	0.3589	0.4263	0.5629
	0.4144	0.3262	0.4660	0.3325	0.5221	0.3482	0.4786	0.3413	0.4177	0.5624
	0.4134	0.3196	0.4655	0.3314	0.5034	0.3324	0.4676	0.3389	0.4154	0.5520
	0.4043	0.3178	0.4608	0.3201	0.4919	0.3249	0.4635	0.3377	0.4138	0.5432
	0.3959	0.3154	0.4550	0.3158	0.4775	0.3216	0.4615	0.3373	0.4106	0.5419
	0.3819	0.3034	0.4502	0.3048	0.4747	0.3190	0.4547	0.3363	0.4012	0.5419
	0.3780	0.3024	0.4493	0.3034	0.4656	0.3162	0.4518	0.3278	0.3941	0.5310
	0.3765	0.3009	0.4439	0.2981	0.4579	0.3144	0.4517	0.3198	0.3930	0.5306
	0.3737	0.2964	0.4314	0.2796	0.4550	0.3113	0.4433	0.3114	0.3827	0.5182
	0.3684	0.2958	0.4303	0.2745	0.4515	0.3101	0.4377	0.3081	0.3825	0.4915
	0.3557	0.2955	0.4263	0.2667	0.4326	0.3064	0.4174	0.2883	0.3740	0.4763
	0.3495	0.2853	0.4157	0.2649	0.4234	0.3049	0.4134	0.2831	0.3723	0.4700
	0.3321	0.2752	0.4133	0.2613	0.4163	0.2965	0.4109	0.2717	0.3699	0.4519
	0.3293	0.2709	0.3980	0.2604	0.4009	0.2950	0.4051	0.2707	0.3664	0.4502
	0.3261	0.2587	0.3921	0.2548	0.3996	0.2937	0.3987	0.2538	0.3636	0.4317
	0.3246	0.2566	0.3916	0.2506	0.3945	0.2906	0.3986	0.2484	0.3582	0.4314
	0.3040	0.2440	0.3799	0.2474	0.3907	0.2879	0.3900	0.2423	0.3576	0.4274
	0.2964	0.2243	0.3620	0.2377	0.3779	0.2828	0.3864	0.2403	0.3533	0.3970
	0.2955	0.2235	0.3493	0.2230	0.3760	0.2762	0.3840	0.2357	0.3527	0.3889
	0.2884	0.2194	0.3481	0.2203	0.3724	0.2583	0.3776	0.2314	0.3401	0.3865
	0.2681	0.2086	0.3382	0.2027	0.3683	0.2361	0.3643	0.1960	0.3266	0.3797
	0.2584	0.2063	0.3313	0.1966	0.3547	0.2295	0.3515	0.1802	0.3106	0.3715
	0.2300	0.1827	0.3148	0.1829	0.3037	0.2260	0.3352	0.1782	0.2931	0.3556
	0.2216	0.1581	0.3086	0.1483	0.2934	0.2040	0.3295	0.1638	0.2649	0.3418
	0.2128	0.1083	0.2975	0.1426	0.2737	0.2005	0.3291	0.1632	0.2309	0.3111
r'	**3**	**4**	**15**	**8**	**3**	**8**	**3**	**22**	**5**	**1**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and c	0.7757	0.5258	0.8118	0.5601	0.5341	0.5674	0.6414	0.8041	0.7666	0.6888
	0.7434	0.5135	0.7506	0.5407	0.4952	0.5535	0.5885	0.6430	0.5702	0.6209
	0.7350	0.4900	0.7408	0.5365	0.4607	0.5384	0.5851	0.6369	0.5298	0.6116
	0.7343	0.4798	0.7380	0.4560	0.4380	0.5307	0.5383	0.5779	0.5142	0.5985
	0.6907	0.4755	0.7179	0.4178	0.4158	0.5298	0.5245	0.5562	0.4942	0.5975
	0.6784	0.4614	0.7163	0.4072	0.3715	0.5189	0.5008	0.5191	0.4876	0.5621
	0.6456	0.4340	0.7051	0.4025	0.3475	0.4972	0.4835	0.5084	0.4759	0.5526
	0.6418	0.4272	0.6856	0.3987	0.3430	0.4970	0.4833	0.5066	0.4414	0.5503
	0.6393	0.4261	0.6803	0.3870	0.3169	0.4829	0.4789	0.4949	0.4201	0.5463
	0.5954	0.4254	0.6646	0.3825	0.3156	0.4781	0.4733	0.4844	0.3984	0.5337
	0.5953	0.4063	0.6291	0.3689	0.3045	0.4735	0.4723	0.4731	0.3961	0.5257
	0.5751	0.3957	0.6188	0.3676	0.3039	0.4680	0.4547	0.4728	0.3961	0.5049
	0.5614	0.3709	0.6095	0.3613	0.3012	0.4522	0.4512	0.4392	0.3935	0.5021
	0.5478	0.3703	0.6051	0.3592	0.2908	0.4462	0.4501	0.4192	0.3890	0.4969
	0.5448	0.3690	0.6047	0.3581	0.2907	0.4425	0.4422	0.4141	0.3884	0.4939
	0.5268	0.3621	0.5939	0.3522	0.2876	0.4319	0.4419	0.4119	0.3876	0.4722
	0.5219	0.3613	0.5613	0.3496	0.2828	0.4155	0.4406	0.4045	0.3800	0.4636
	0.5177	0.3543	0.5556	0.3439	0.2747	0.4078	0.4308	0.4039	0.3655	0.4610
	0.4883	0.3479	0.5527	0.3364	0.2736	0.4075	0.4296	0.3991	0.3628	0.4580
	0.4871	0.3431	0.5519	0.3243	0.2598	0.4058	0.4209	0.3925	0.3597	0.4435
	0.4593	0.3398	0.5490	0.3173	0.2559	0.3635	0.4189	0.3782	0.3588	0.4282
	0.4517	0.3277	0.5444	0.3159	0.2523	0.3631	0.4127	0.3541	0.3565	0.4272
	0.4463	0.3214	0.5401	0.3126	0.2520	0.3602	0.4119	0.3472	0.3559	0.4129
	0.4340	0.3188	0.5366	0.3005	0.2493	0.3591	0.4102	0.3256	0.3448	0.4100
	0.4325	0.3070	0.5359	0.2787	0.2462	0.3523	0.3984	0.3203	0.3332	0.4081
	0.4288	0.3017	0.5357	0.2778	0.2459	0.3507	0.3966	0.2995	0.3279	0.4013
	0.4228	0.2893	0.5328	0.2770	0.2449	0.3492	0.3896	0.2976	0.3251	0.4003
	0.4146	0.2867	0.5277	0.2768	0.2437	0.3408	0.3727	0.2963	0.3207	0.3936
	0.4054	0.2579	0.5110	0.2726	0.2393	0.3363	0.3678	0.2881	0.3195	0.3830
	0.4009	0.2442	0.5020	0.2663	0.2332	0.3346	0.3622	0.2622	0.3081	0.3708
	0.3961	0.2409	0.4917	0.2635	0.2331	0.3323	0.3542	0.2607	0.3031	0.3551
	0.3897	0.2344	0.4861	0.2604	0.2320	0.3220	0.3372	0.2515	0.2948	0.3509
	0.3712	0.2284	0.4651	0.2486	0.2313	0.3198	0.3299	0.2500	0.2818	0.3491
	0.3588	0.2225	0.4642	0.2471	0.2165	0.3111	0.3283	0.2427	0.2785	0.3447
	0.3532	0.2116	0.4585	0.2264	0.2111	0.3003	0.3037	0.2345	0.2683	0.3319
	0.3457	0.1960	0.4396	0.1835	0.2057	0.2933	0.2983	0.2282	0.2395	0.3148
	0.3300	0.1883	0.4126	0.1801	0.1721	0.2843	0.2764	0.2234	0.2130	0.3062
	0.2732	0.1744	0.3805	0.1709	0.1651	0.2802	0.2528	0.1986	0.1938	0.2829
	0.2306	0.1698	0.3649	0.1413	0.1238	0.2472	0.1728	0.1798	0.1486	0.2533
r'	**2**	**3**	**2**	**12**	**5**	**9**	**2**	**4**	**1**	**4**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and c	0.7388	0.7594	0.7315	0.6269	0.8763	0.7793	0.8571	0.7615	0.8074	0.8176
	0.7353	0.7205	0.6903	0.5763	0.7275	0.6235	0.6982	0.7475	0.7726	0.5016
	0.6972	0.6644	0.6770	0.5199	0.5598	0.6043	0.6950	0.6911	0.7692	0.4762
	0.6562	0.6085	0.6712	0.5062	0.4957	0.5869	0.6393	0.6104	0.7640	0.3959
	0.6393	0.5835	0.6495	0.4538	0.4852	0.5674	0.6295	0.5664	0.7012	0.3805
	0.6257	0.5687	0.6123	0.3911	0.4551	0.5649	0.6121	0.5438	0.6503	0.3709
	0.6149	0.5155	0.6030	0.3851	0.4541	0.5309	0.6004	0.5322	0.6343	0.3465
	0.5951	0.4964	0.5897	0.3785	0.4514	0.5095	0.5867	0.5299	0.6291	0.3340
	0.5853	0.4917	0.5895	0.3731	0.4481	0.5001	0.5719	0.5140	0.6185	0.3334
	0.5578	0.4897	0.5856	0.3668	0.4463	0.4941	0.5590	0.5078	0.5919	0.3311
	0.5491	0.4822	0.5568	0.3553	0.4405	0.4930	0.5352	0.4965	0.5913	0.3282
	0.5233	0.4639	0.5491	0.3509	0.4147	0.4928	0.5146	0.4842	0.5876	0.3277
	0.5229	0.4476	0.5419	0.3434	0.4133	0.4588	0.5134	0.4721	0.5821	0.3200
	0.5195	0.4341	0.5305	0.3404	0.4045	0.4434	0.4882	0.4650	0.5724	0.3128
	0.4809	0.4224	0.5300	0.3177	0.3814	0.4428	0.4872	0.4603	0.5651	0.3033
	0.4778	0.4189	0.5241	0.3110	0.3689	0.4384	0.4856	0.4596	0.5600	0.3020
	0.4629	0.4162	0.5220	0.3011	0.3672	0.4246	0.4846	0.4474	0.5401	0.2892
	0.4623	0.4121	0.5193	0.2908	0.3667	0.4237	0.4824	0.4462	0.5334	0.2888
	0.4604	0.4030	0.5004	0.2827	0.3606	0.4136	0.4742	0.4441	0.5305	0.2884
	0.4524	0.3855	0.5003	0.2824	0.3478	0.4026	0.4705	0.4282	0.5220	0.2883
	0.4429	0.3801	0.4884	0.2799	0.3458	0.4010	0.4653	0.4254	0.5209	0.2778
	0.4416	0.3757	0.4860	0.2774	0.3319	0.3999	0.4601	0.4218	0.5185	0.2738
	0.4387	0.3735	0.4853	0.2732	0.3235	0.3964	0.4595	0.4051	0.5160	0.2649
	0.4387	0.3419	0.4438	0.2582	0.3201	0.3862	0.4568	0.4012	0.5149	0.2617
	0.4378	0.3341	0.4411	0.2502	0.3186	0.3819	0.4369	0.3830	0.4948	0.2592
	0.4127	0.3340	0.4409	0.2347	0.3145	0.3819	0.4361	0.3795	0.4947	0.2532
	0.4005	0.3242	0.4303	0.2260	0.3069	0.3727	0.4331	0.3775	0.4919	0.2482
	0.3984	0.3224	0.4244	0.2253	0.3054	0.3597	0.4161	0.3452	0.4764	0.2383
	0.3878	0.3183	0.4081	0.2008	0.3041	0.3577	0.3854	0.3389	0.4577	0.2229
	0.3840	0.3094	0.4040	0.1900	0.2960	0.3562	0.3848	0.3319	0.4407	0.2160
	0.3764	0.3091	0.3985	0.1755	0.2956	0.3508	0.3752	0.3306	0.4300	0.2150
	0.3680	0.3040	0.3924	0.1743	0.2951	0.3392	0.3727	0.3261	0.4250	0.2135
	0.3663	0.2953	0.3775	0.1707	0.2779	0.3350	0.3559	0.3231	0.4045	0.2119
	0.3597	0.2901	0.3713	0.1671	0.2635	0.3134	0.3477	0.3228	0.3962	0.2081
	0.3250	0.2871	0.3629	0.1632	0.2533	0.3103	0.3438	0.3218	0.3957	0.1959
	0.3152	0.2654	0.3542	0.1628	0.2346	0.2977	0.3327	0.2667	0.3791	0.1779
	0.3066	0.2620	0.3145	0.1522	0.2134	0.2705	0.3129	0.2636	0.3751	0.1673
	0.3047	0.2591	0.3009	0.1349	0.2049	0.2471	0.3043	0.2226	0.3745	0.1530
	0.2784	0.2340	0.2559	0.1247	0.1911	0.2432	0.2976	0.2198	0.2259	0.1124
r'	**1**	**6**	**5**	**1**	**12**	**2**	**2**	**4**	**13**	**1**

	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}
All samples in the group a and c	0.7181	0.7146	0.5475	0.8359	0.6940	0.7996	0.8084	0.8096	0.6321
	0.7074	0.6980	0.5397	0.7932	0.6606	0.7409	0.7801	0.6642	0.5646
	0.6999	0.6621	0.5386	0.6732	0.6606	0.7170	0.7225	0.6181	0.5470
	0.6564	0.6485	0.5076	0.6327	0.5541	0.7051	0.7067	0.6112	0.5337
	0.6465	0.6286	0.4524	0.5743	0.5163	0.6871	0.6784	0.5897	0.5061
	0.6449	0.5930	0.4427	0.5384	0.5000	0.6784	0.6687	0.5698	0.4408
	0.6384	0.5928	0.4384	0.5100	0.4921	0.6714	0.6648	0.5653	0.4162
	0.5757	0.5777	0.4232	0.4964	0.4750	0.6668	0.6634	0.5439	0.4153
	0.5555	0.5713	0.4147	0.4940	0.4733	0.6603	0.6532	0.5343	0.4129
	0.5183	0.5608	0.4087	0.4894	0.4470	0.6464	0.6388	0.5278	0.4057
	0.5105	0.5566	0.3993	0.4847	0.4390	0.6086	0.6357	0.5239	0.3977
	0.4984	0.5379	0.3869	0.4752	0.4286	0.6070	0.6218	0.5227	0.3842
	0.4934	0.5280	0.3761	0.4569	0.4196	0.6039	0.6200	0.5012	0.3825
	0.4762	0.5237	0.3750	0.4412	0.4114	0.5942	0.5776	0.4919	0.3716
	0.4699	0.5090	0.3681	0.4318	0.4077	0.5938	0.5772	0.4783	0.3677
	0.4689	0.5083	0.3604	0.4307	0.4052	0.5913	0.5728	0.4734	0.3571
	0.4660	0.4931	0.3574	0.4306	0.3987	0.5889	0.5702	0.4691	0.3556
	0.4621	0.4718	0.3554	0.4198	0.3986	0.5813	0.5682	0.4657	0.3495
	0.4548	0.4645	0.3081	0.4157	0.3966	0.5703	0.5667	0.4637	0.3446
	0.4545	0.4386	0.3039	0.4134	0.3956	0.5648	0.5646	0.4631	0.3258
	0.4528	0.4345	0.3009	0.4024	0.3932	0.5549	0.5644	0.4419	0.3173
	0.4513	0.4118	0.2942	0.4019	0.3906	0.5360	0.5528	0.4339	0.2921
	0.4500	0.4100	0.2768	0.3782	0.3825	0.5332	0.4959	0.4331	0.2866
	0.4449	0.4071	0.2694	0.3642	0.3726	0.5245	0.4955	0.4180	0.2644
	0.4343	0.4009	0.2660	0.3627	0.3570	0.5163	0.4937	0.4166	0.2622
	0.4262	0.3962	0.2607	0.3577	0.3562	0.5107	0.4783	0.3934	0.2551
	0.4188	0.3907	0.2585	0.3528	0.3369	0.5068	0.4771	0.3915	0.2529
	0.3902	0.3825	0.2573	0.3370	0.3341	0.4959	0.4676	0.3856	0.2438
	0.3893	0.3732	0.2411	0.3305	0.3147	0.4938	0.4627	0.3854	0.2388
	0.3878	0.3610	0.2367	0.3285	0.3082	0.4740	0.4584	0.3723	0.2342
	0.3574	0.3580	0.2345	0.3226	0.2920	0.4665	0.4564	0.3628	0.2219
	0.3530	0.3530	0.2249	0.3191	0.2829	0.4636	0.4551	0.3590	0.2217
	0.3509	0.3488	0.2240	0.3191	0.2728	0.4451	0.4490	0.3548	0.2121
	0.3491	0.3082	0.2226	0.3169	0.2684	0.4093	0.4324	0.3468	0.2093
	0.3467	0.2986	0.2197	0.3057	0.2353	0.4078	0.4258	0.3393	0.2002
	0.3202	0.2858	0.2099	0.2816	0.2338	0.4020	0.4192	0.3253	0.1973
	0.3070	0.2553	0.2020	0.2634	0.2200	0.3991	0.4076	0.3039	0.1876
	0.3042	0.2517	0.1956	0.2459	0.2098	0.3834	0.3629	0.3015	0.1761
	0.3029	0.2316	0.1844	0.2207	0.1761	0.3676	0.3615	0.2833	0.1209
r'	**4**	**9**	**2**	**8**	**3**	**6**	**1**	**2**	**11**

Table A.31: Ranked *sim* values for $(c \vee \{a, b\})$ in VHHS dataset using U-KTS FVR.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
All samples in the group a and b	0.8400	0.5744	0.6605	0.5121	0.6849	0.5500	0.5611	0.6441	0.7436	0.7375
	0.7505	0.5297	0.6287	0.4861	0.6478	0.5067	0.5444	0.5148	0.6789	0.5863
	0.6145	0.5009	0.6016	0.4825	0.6324	0.5053	0.5172	0.4382	0.6775	0.5708
	0.6039	0.4886	0.5243	0.4799	0.6203	0.4682	0.4900	0.4323	0.6506	0.5052
	0.5938	0.4732	0.4990	0.4187	0.5931	0.4479	0.4837	0.4206	0.6374	0.4774
	0.5902	0.4386	0.4951	0.4058	0.5737	0.4216	0.4809	0.4185	0.6271	0.4498
	0.5865	0.4142	0.4918	0.3927	0.5677	0.4054	0.4654	0.4148	0.6178	0.4432
	0.5667	0.4101	0.4779	0.3825	0.5585	0.3966	0.4378	0.4067	0.5873	0.4276
	0.5385	0.3459	0.4755	0.3812	0.5420	0.3859	0.4054	0.4055	0.5814	0.4219
	0.5381	0.3454	0.4729	0.3725	0.5256	0.3829	0.3876	0.3939	0.5812	0.4053
	0.5267	0.3439	0.4584	0.3621	0.5033	0.3814	0.3655	0.3890	0.5790	0.4009
	0.5048	0.3407	0.4357	0.3510	0.5006	0.3629	0.3586	0.3853	0.5785	0.3767
	0.4992	0.3334	0.4340	0.3460	0.4940	0.3618	0.3464	0.3795	0.5742	0.3715
	0.4845	0.3302	0.4282	0.3402	0.4802	0.3569	0.3312	0.3783	0.5741	0.3632
	0.4831	0.3233	0.4206	0.3329	0.4801	0.3513	0.3172	0.3706	0.5648	0.3527
	0.4726	0.3161	0.4104	0.3285	0.4789	0.3349	0.3078	0.3706	0.5589	0.3406
	0.4663	0.3113	0.4079	0.3267	0.4787	0.3285	0.3030	0.3595	0.5540	0.3281
	0.4593	0.3045	0.3999	0.3247	0.4725	0.3241	0.2985	0.3418	0.5540	0.3261
	0.4401	0.3040	0.3917	0.3196	0.4647	0.3192	0.2917	0.3410	0.5473	0.3236
	0.4392	0.2997	0.3620	0.3184	0.4554	0.3188	0.2894	0.3341	0.5360	0.3176
	0.4342	0.2884	0.3609	0.3012	0.4417	0.3116	0.2875	0.3289	0.5131	0.3101
	0.4301	0.2882	0.3453	0.2869	0.4414	0.3086	0.2831	0.3269	0.5028	0.3054
	0.4277	0.2835	0.3429	0.2822	0.4284	0.3025	0.2788	0.3268	0.4969	0.2847
	0.4253	0.2711	0.3411	0.2689	0.4256	0.3007	0.2720	0.3111	0.4870	0.2737
	0.4195	0.2640	0.3398	0.2667	0.4209	0.2819	0.2700	0.3107	0.4716	0.2668
	0.4026	0.2492	0.3333	0.2542	0.4110	0.2747	0.2695	0.3054	0.4652	0.2626
	0.3947	0.2449	0.3170	0.2505	0.4032	0.2713	0.2679	0.3005	0.4534	0.2606
	0.3826	0.2356	0.3105	0.2443	0.4027	0.2695	0.2660	0.2885	0.4488	0.2584
	0.3798	0.2294	0.3092	0.2411	0.3985	0.2640	0.2626	0.2804	0.4397	0.2504
	0.3717	0.2226	0.2969	0.2355	0.3914	0.2595	0.2470	0.2752	0.4258	0.2342
	0.3686	0.2050	0.2922	0.2225	0.3895	0.2549	0.2435	0.2749	0.4145	0.2299
	0.3512	0.2005	0.2857	0.2221	0.3746	0.2530	0.2424	0.2739	0.4073	0.2267
	0.3464	0.1977	0.2654	0.2215	0.3705	0.2528	0.2284	0.2696	0.4071	0.2208
	0.3399	0.1970	0.2644	0.2195	0.3639	0.2405	0.2184	0.2535	0.3825	0.2154
	0.3183	0.1964	0.2593	0.1944	0.3489	0.2401	0.2036	0.2461	0.3715	0.2060
	0.3159	0.1946	0.2416	0.1938	0.3370	0.2390	0.2033	0.2302	0.3576	0.1976
	0.2911	0.1735	0.2133	0.1925	0.3271	0.2382	0.1904	0.2302	0.3454	0.1885
	0.2852	0.1643	0.1818	0.1862	0.3019	0.1779	0.1902	0.2064	0.3100	0.1880
	0.2315	0.1552	0.1589	0.1627	0.2721	0.1602	0.1481	0.2053	0.2256	0.1516
r'	**4**	**6**	**10**	**8**	**1**	**8**	**2**	**4**	**2**	**2**

	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
All samples in the group a and b	0.6444	0.7287	0.7444	0.6661	0.4888	0.6398	0.8002	0.7366	0.8038	0.8066
	0.5590	0.5254	0.7444	0.6394	0.4611	0.5749	0.7385	0.7123	0.6804	0.6806
	0.5070	0.4841	0.7287	0.5948	0.4577	0.5716	0.6473	0.6699	0.6783	0.6426
	0.4724	0.4448	0.7055	0.5435	0.4149	0.5599	0.4999	0.6526	0.6155	0.6094
	0.4717	0.4385	0.6705	0.5387	0.3778	0.5189	0.4938	0.6476	0.6026	0.5867
	0.4546	0.4362	0.6642	0.5198	0.3766	0.5093	0.4937	0.6314	0.5923	0.5753
	0.4427	0.4248	0.6632	0.4954	0.3689	0.5010	0.4808	0.6261	0.5854	0.5367
	0.4364	0.3978	0.6598	0.4915	0.3393	0.4996	0.4759	0.6036	0.5836	0.5228
	0.4195	0.3975	0.6335	0.4885	0.3379	0.4970	0.4652	0.6029	0.5747	0.5139
	0.4156	0.3842	0.6233	0.4785	0.3356	0.4816	0.4595	0.5961	0.5703	0.5130
	0.4122	0.3790	0.6171	0.4619	0.3167	0.4812	0.4228	0.5956	0.5664	0.5121
	0.4006	0.3632	0.6170	0.4380	0.3116	0.4788	0.4138	0.5732	0.5446	0.5016
	0.3987	0.3629	0.6019	0.4356	0.3090	0.4773	0.4110	0.5718	0.5217	0.4645
	0.3950	0.3624	0.6018	0.4214	0.2966	0.4757	0.3975	0.5195	0.5123	0.4636
	0.3862	0.3605	0.5980	0.4186	0.2851	0.4748	0.3969	0.5184	0.5053	0.4414
	0.3837	0.3563	0.5910	0.4035	0.2798	0.4571	0.3940	0.4744	0.5043	0.4389
	0.3786	0.3538	0.5547	0.4029	0.2767	0.4557	0.3879	0.4718	0.5037	0.4236
	0.3681	0.3456	0.5460	0.3964	0.2756	0.4528	0.3851	0.4670	0.4979	0.4150
	0.3667	0.3399	0.5399	0.3806	0.2660	0.4521	0.3825	0.4636	0.4943	0.4084
	0.3601	0.3372	0.5206	0.3798	0.2656	0.4298	0.3810	0.4542	0.4837	0.4008
	0.3539	0.3304	0.5105	0.3741	0.2600	0.4203	0.3647	0.4518	0.4755	0.3971
	0.3518	0.3271	0.5044	0.3640	0.2479	0.4116	0.3537	0.4461	0.4706	0.3902
	0.3451	0.3139	0.4891	0.3446	0.2457	0.3751	0.3532	0.4362	0.4492	0.3897
	0.3357	0.3069	0.4865	0.3410	0.2311	0.3659	0.3487	0.4303	0.4488	0.3636
	0.3347	0.3061	0.4772	0.3370	0.2310	0.3650	0.3470	0.4223	0.4336	0.3578
	0.3194	0.2939	0.4725	0.3238	0.2292	0.3614	0.3424	0.4159	0.4322	0.3555
	0.3188	0.2859	0.4678	0.3113	0.2197	0.3599	0.3349	0.4038	0.4171	0.3477
	0.3132	0.2768	0.4614	0.3025	0.2147	0.3599	0.3307	0.4019	0.4135	0.3426
	0.3094	0.2658	0.4506	0.2963	0.2113	0.3300	0.3194	0.3613	0.4117	0.3419
	0.3083	0.2641	0.4495	0.2916	0.2007	0.3299	0.3088	0.3408	0.4079	0.3274
	0.2906	0.2591	0.4485	0.2786	0.1963	0.3273	0.3001	0.3114	0.4030	0.3274
	0.2896	0.2495	0.4467	0.2780	0.1871	0.3253	0.3000	0.3022	0.3845	0.3205
	0.2854	0.2466	0.4381	0.2747	0.1739	0.3231	0.2942	0.3007	0.3674	0.3102
	0.2693	0.2269	0.4372	0.2666	0.1736	0.3114	0.2789	0.2949	0.3649	0.3039
	0.2692	0.2227	0.4273	0.2600	0.1533	0.3105	0.2699	0.2843	0.3343	0.2990
	0.2382	0.2187	0.4020	0.2517	0.1497	0.2791	0.2615	0.2840	0.3100	0.2978
	0.2371	0.1943	0.4019	0.2106	0.1269	0.2688	0.2222	0.2641	0.3009	0.2861
	0.2078	0.1931	0.3495	0.2086	0.1222	0.2610	0.2108	0.2480	0.2799	0.2677
	0.2034	0.1668	0.2850	0.2024	0.1216	0.2325	0.2105	0.2255	0.2487	0.2630
r'	**9**	**1**	**1**	**10**	**15**	**8**	**2**	**9**	**4**	**3**

	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}	S_{27}	S_{28}	S_{29}	S_{30}
All samples in the group a and b	0.8058	0.6377	0.8029	0.6102	0.7158	0.6426	0.7150	0.7822	0.6505	0.7073
	0.6660	0.5615	0.7825	0.5691	0.6973	0.6374	0.6441	0.7030	0.4744	0.7064
	0.6312	0.5509	0.7571	0.5266	0.6156	0.6132	0.5969	0.6341	0.4612	0.6846
	0.6300	0.5508	0.7223	0.4841	0.5909	0.5744	0.5729	0.6317	0.4446	0.6617
	0.6281	0.5494	0.6903	0.4747	0.5855	0.5525	0.5188	0.6185	0.4319	0.6603
	0.6281	0.5494	0.6774	0.4667	0.5616	0.5150	0.5070	0.6180	0.4272	0.6482
	0.6124	0.5442	0.6551	0.4549	0.5290	0.4857	0.4828	0.5777	0.4169	0.6457
	0.6008	0.5281	0.6222	0.4435	0.5110	0.4407	0.4768	0.5734	0.4127	0.6311
	0.5831	0.5279	0.6145	0.4417	0.5091	0.4318	0.4692	0.5720	0.4119	0.6299
	0.5742	0.5085	0.5817	0.4274	0.5081	0.4253	0.4662	0.5647	0.4056	0.6290
	0.5625	0.4989	0.5767	0.4243	0.5032	0.4150	0.4649	0.5422	0.4030	0.6284
	0.5568	0.4854	0.5645	0.4088	0.4919	0.4119	0.4347	0.5215	0.3880	0.6038
	0.5551	0.4788	0.5540	0.4051	0.4863	0.4051	0.4329	0.5115	0.3860	0.5964
	0.5467	0.4670	0.5532	0.3849	0.4749	0.3875	0.4097	0.5078	0.3857	0.5776
	0.5376	0.4546	0.5530	0.3636	0.4361	0.3840	0.3952	0.5006	0.3835	0.5728
	0.5270	0.4263	0.5515	0.3528	0.4358	0.3801	0.3881	0.4906	0.3712	0.5701
	0.5137	0.4241	0.5499	0.3501	0.4329	0.3796	0.3819	0.4743	0.3641	0.5643
	0.5063	0.4171	0.5490	0.3434	0.4305	0.3782	0.3567	0.4673	0.3639	0.5464
	0.4949	0.4143	0.5489	0.3395	0.4223	0.3763	0.3484	0.4618	0.3625	0.5328
	0.4938	0.3977	0.5449	0.3361	0.4178	0.3519	0.3452	0.4378	0.3527	0.5277
	0.4907	0.3925	0.5396	0.3277	0.4175	0.3486	0.3370	0.4362	0.3507	0.5200
	0.4696	0.3869	0.5250	0.3264	0.4144	0.3475	0.3292	0.4293	0.3495	0.5076
	0.4639	0.3796	0.5167	0.3108	0.4137	0.3428	0.3261	0.4260	0.3385	0.5044
	0.4587	0.3514	0.4962	0.3066	0.4108	0.3416	0.3202	0.4205	0.3290	0.5017
	0.4197	0.3443	0.4952	0.3002	0.3954	0.3339	0.3026	0.4134	0.3214	0.5005
	0.4195	0.3426	0.4928	0.2977	0.3946	0.3286	0.2938	0.4066	0.3035	0.4887
	0.4163	0.3057	0.4865	0.2954	0.3914	0.3159	0.2920	0.3780	0.2962	0.4701
	0.4049	0.3000	0.4860	0.2786	0.3913	0.3144	0.2799	0.3735	0.2947	0.4686
	0.4007	0.2990	0.4724	0.2621	0.3524	0.3056	0.2789	0.3709	0.2870	0.4569
	0.3999	0.2887	0.4680	0.2619	0.3364	0.2971	0.2659	0.3661	0.2861	0.4541
	0.3976	0.2794	0.4613	0.2566	0.3276	0.2895	0.2596	0.3553	0.2771	0.4442
	0.3916	0.2701	0.4530	0.2485	0.3274	0.2662	0.2580	0.3543	0.2770	0.3973
	0.3876	0.2658	0.4362	0.2479	0.3190	0.2658	0.2542	0.3529	0.2691	0.3962
	0.3738	0.2618	0.4345	0.2444	0.3129	0.2639	0.2517	0.3289	0.2671	0.3868
	0.3575	0.2530	0.4302	0.2398	0.3049	0.2483	0.2445	0.3173	0.2658	0.3841
	0.3551	0.2506	0.4205	0.2160	0.2951	0.2272	0.2438	0.3152	0.2644	0.3687
	0.3351	0.2402	0.3967	0.2060	0.2945	0.2249	0.2250	0.2999	0.2137	0.3581
	0.3323	0.2297	0.3805	0.2050	0.2850	0.2030	0.2200	0.2757	0.1974	0.3464
	0.3013	0.2143	0.2161	0.1809	0.2259	0.1927	0.1934	0.2444	0.1500	0.2865
r'	**1**	**6**	**8**	**2**	**13**	**1**	**12**	**4**	**3**	**1**

	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	S_{36}	S_{37}	S_{38}	S_{39}
All samples in the group a and b	0.5529	0.6356	0.5289	0.6189	0.4630	0.5957	0.6861	0.7842	0.8525
	0.5520	0.6035	0.5211	0.6126	0.4311	0.5877	0.6215	0.7555	0.5501
	0.5389	0.5882	0.5150	0.5324	0.4166	0.5780	0.5677	0.7362	0.5390
	0.4843	0.5753	0.5097	0.5024	0.4134	0.5548	0.4646	0.7083	0.4991
	0.4568	0.5580	0.4930	0.4846	0.4119	0.5458	0.4491	0.7069	0.4972
	0.4369	0.5561	0.4659	0.4472	0.4064	0.5453	0.4409	0.6886	0.4929
	0.4194	0.5319	0.4608	0.4354	0.4036	0.5414	0.4311	0.6704	0.4699
	0.4186	0.5282	0.4430	0.4298	0.4033	0.5338	0.4005	0.6315	0.4411
	0.4078	0.5129	0.4291	0.4296	0.3996	0.5307	0.3729	0.6254	0.4330
	0.4036	0.5045	0.4176	0.4221	0.3870	0.5200	0.3696	0.6148	0.4325
	0.3950	0.4980	0.4161	0.3913	0.3756	0.5130	0.3559	0.6078	0.4294
	0.3827	0.4869	0.4074	0.3884	0.3727	0.4944	0.3555	0.6061	0.4279
	0.3764	0.4722	0.3939	0.3790	0.3717	0.4936	0.3531	0.5953	0.4272
	0.3734	0.4568	0.3928	0.3721	0.3647	0.4808	0.3528	0.5914	0.4023
	0.3716	0.4361	0.3905	0.3695	0.3621	0.4791	0.3470	0.5866	0.3913
	0.3664	0.4306	0.3876	0.3678	0.3586	0.4762	0.3429	0.5852	0.3834
	0.3631	0.4246	0.3795	0.3654	0.3558	0.4750	0.3368	0.5771	0.3745
	0.3584	0.3856	0.3789	0.3623	0.3461	0.4669	0.3217	0.5769	0.3548
	0.3465	0.3849	0.3759	0.3517	0.3341	0.4498	0.3209	0.5692	0.3519
	0.3458	0.3786	0.3692	0.3332	0.3260	0.4453	0.3185	0.5687	0.3330
	0.3407	0.3576	0.3658	0.3312	0.3238	0.4260	0.3000	0.5591	0.3109
	0.3307	0.3534	0.3613	0.3223	0.3205	0.4251	0.2944	0.5557	0.3026
	0.3212	0.3422	0.3570	0.3192	0.3193	0.4189	0.2928	0.5556	0.3001
	0.3074	0.3320	0.3534	0.3127	0.3185	0.4162	0.2914	0.5494	0.2907
	0.3012	0.3266	0.3468	0.2956	0.3038	0.4082	0.2897	0.5305	0.2871
	0.2995	0.3250	0.3393	0.2940	0.3025	0.4081	0.2859	0.5237	0.2794
	0.2968	0.3244	0.3319	0.2783	0.2988	0.4062	0.2842	0.5169	0.2709
	0.2888	0.3235	0.3199	0.2770	0.2957	0.4023	0.2828	0.5019	0.2654
	0.2800	0.3198	0.3187	0.2765	0.2794	0.3719	0.2816	0.4869	0.2636
	0.2648	0.3130	0.3091	0.2692	0.2772	0.3637	0.2767	0.4806	0.2556
	0.2616	0.2907	0.2972	0.2493	0.2692	0.3553	0.2684	0.4724	0.2523
	0.2557	0.2870	0.2970	0.2490	0.2647	0.3507	0.2566	0.4636	0.2508
	0.2479	0.2803	0.2828	0.2369	0.2608	0.3503	0.2463	0.4573	0.2422
	0.2418	0.2757	0.2683	0.2367	0.2602	0.3394	0.2357	0.4557	0.2409
	0.2403	0.2749	0.2679	0.2214	0.2516	0.3056	0.2315	0.4524	0.2326
	0.2229	0.2746	0.2612	0.2178	0.2496	0.3040	0.2293	0.4368	0.2312
	0.2206	0.2633	0.2589	0.2125	0.2384	0.2794	0.2048	0.4225	0.2260
	0.2186	0.2402	0.2115	0.2086	0.2306	0.2672	0.2042	0.3913	0.2245
	0.1957	0.2101	0.1384	0.1848	0.2141	0.2656	0.1836	0.3507	0.2073
r'	**6**	**9**	**2**	**1**	**4**	**7**	**1**	**1**	**11**

Appendix B

Further Experimental Results Concerning IKCA Approach Presented in Chapter 6

B.1 Overview

In this appendix additional results concerning the experiments presented in Chapter 6 are presented. In particular the results of IKCA evaluation with respect to accuracy, FMR and FNMR for the $\mathcal{KH}^t$ feature. The evaluation of IKCA with $\mathcal{KH}^t$ was conducted using a range of ω values, similar to the that used concerning the experiments reported on in Chapter 6 using $\mathcal{F}^t$. Authentication performance was evaluated as described in Sub-section 6.4.3. Tables B.1 and B.2 present the accuracy, FMR and FNMR results obtained using the ACB and VHHS data sets respectively. Form the tables, it can be observed that best values were recorded when $\omega = 125$; this was also noted with respect to earlier experiments reported on in Chapter 6.

For a comparison purpose with IKCA using $\mathcal{F}^t$, Tables B.3 and B.4 present the best values obtained, in both cases, using the ACB and VHHS data sets.

TABLE B.1: Obtained results for IKCA coupled with from $\mathcal{KH}^t$.

ω	Acc.	FMR	FNMR
25	72.02	1.780	3.020
50	86.16	1.651	2.890
75	92.54	1.630	2.885
100	93.67	1.632	2.871
125	94.18	1.600	2.610
150	94.09	1.610	2.672

Table B.2: Obtained results using the IKCA approach with $\mathcal{KH}^t$.

ω	Acc.	FMR	FNMR
25	71.25	1.981	3.060
50	82.45	1.903	3.050
75	83.88	1.890	3.051
100	91.97	1.849	2.972
125	90.78	1.850	2.980
150	90.88	1.850	2.980

Table B.3: Comparison between $\mathcal{KH}^t$ and $\mathcal{F}^t$ results using the IKCA approach.

ω	IKCA with $\mathcal{KH}^t$			IKCA with $\mathcal{F}^t$		
	Acc.	FMR	FNMR	Acc.	FMR	FNMR
25	72.02	1.780	3.020	80.11	0.598	2.025
50	86.16	1.651	2.890	88.02	0.595	1.990
75	92.54	1.630	2.885	95.41	0.591	1.990
100	93.67	1.632	2.871	96.12	0.580	1.970
125	94.18	1.600	2.610	96.22	0.599	1.979
150	94.09	1.610	2.672	95.04	0.586	1.978

Table B.4: Comparison between $\mathcal{KH}^t$ and $\mathcal{F}^t$ results using the IKCA approach.

ω	IKCA with $\mathcal{KH}^t$			IKCA with $\mathcal{F}^t$		
	Acc.	FMR	FNMR	Acc.	FMR	FNMR
25	71.25	1.981	3.060	70.41	1.048	2.052
50	82.45	1.903	3.050	85.24	1.045	2.051
75	83.88	1.890	3.051	83.33	1.047	2.052
100	91.97	1.849	2.972	96.01	1.045	2.050
125	90.78	1.850	2.980	94.06	1.040	2.049
150	90.88	1.850	2.980	93.11	1.047	2.050

Appendix C

Further Experimental Results Concerning KCASA approach presented in Chapter 7

C.1 Overview

This appendix presents further results concerning the evaluation given in Chapter 7 with respect to KCASA approach. More specifically, the results presented here are concerned with the evaluation of KCASA with $\mathcal{KH}^t$ in the context of U-KTS. The evaluation of the KCASA approach with $\mathcal{KH}^t$, in the context of U-KTS, was conducted in terms of accuracy, FMR and FNMR. Tables C.1 and C.2 show the obtained results for KCASA with $\mathcal{KH}^t$ using the ACB and VHHS data sets respectively. Recall that for the GP data set $\mathcal{KH}^t$ information was not available, so it was not possible to use the GP data set in this set of experiments. The table presents the performance of KCASA with $\mathcal{KH}^t$ in the context of DFT and DWT respectively. From the tables, it can be observed that, as in the case of $\mathcal{F}^t$, DWT outperform DFT. The best recorded results, with respect to ACB data set, were recorded when $\omega = 128$; $Acc = 95.85\%$, $FMR = 0.047$ and $FNMR = 1.091$ for DWT. Whereas the best result with respect to VHHS data set was recorded when $\omega = 64$; $Acc = 97.09\%$, $FMR = 0.561$ and $FNMR = 1.194$ in favour of, again, DWT KCASA representation.

TABLE C.1: The obtained results for KCASA coupled with $\mathcal{KH}^t$ applied to the ACB data set.

ω	DFT			DWT		
	Acc.	FMR	FNMR	Acc.	FMR	FNMR
16	81.78	0.37	1.721	86.45	0.197	1.741
32	90.77	0.235	1.608	91.84	0.098	1.104
64	94.64	0.141	1.589	95.29	0.047	1.092
128	94.92	0.139	1.588	**95.85**	**0.046**	**1.091**
256	95.73	0.138	1.561	95.64	0.046	1.092
512	95.10	0.138	1.565	95.14	0.047	1.092

TABLE C.2: The obtained results for KCASA coupled with $\mathcal{KH}^t$ applied to the VHHS data set.

ω	DFT			DWT		
	Acc.	FMR	FNMR	Acc.	FMR	FNMR
16	73.83	0.981	1.971	70.19	1.072	1.947
32	87.15	0.102	1.747	88.63	0.997	1.254
64	94.28	0.087	1.120	**97.09**	**0.561**	**1.194**
128	95.43	0.067	1.112	96.06	0.581	1.199
256	95.26	0.067	1.114	94.89	0.583	1.201
512	94.47	0.080	1.119	96.79	0.579	1.197

For completeness, Tables C.3 and C.4 present a comparison of the performance results for KCASA coupled $\mathcal{KH}^t$ against KCASA coupled $\mathcal{F}^t$, and DFT and DWT respectively, when applied to the ACB data set. In a similar manner, Tables C.5 and C.6 compares the operation of the KCASA approach applied to the VHHS data set.

TABLE C.3: Comparison of KCASA+DFT, coupled with $\mathcal{KH}^t$ and $\mathcal{F}^t$, applied to the ACB data set.

ω	$\mathcal{KH}^t$			$\mathcal{F}^t$		
	Acc.	FMR	FNMR	Acc.	FMR	FNMR
16	81.78	0.37	1.721	83.46	0.278	1.601
32	90.77	0.235	1.608	92.61	0.138	1.588
64	94.64	0.141	1.589	97.10	0.130	1.500
128	94.92	0.139	1.588	96.00	0.131	1.510
256	95.73	0.138	1.561	96.73	0.132	1.511
512	95.10	0.138	1.565	96.55	0.131	1.511

Table C.4: Comparison of KCASA+DWT, coupled with $\mathcal{KH}^t$ and $\mathcal{F}^t$, applied to the ACB data set.

ω	$\mathcal{KH}^t$			$\mathcal{F}^t$		
	Acc.	FMR	FNMR	Acc.	FMR	FNMR
16	81.78	0.37	1.721	83.46	0.278	1.601
32	90.77	0.235	1.608	92.61	0.138	1.588
64	94.64	0.141	1.589	97.10	0.130	1.500
128	94.92	0.139	1.588	96.00	0.131	1.510
256	95.73	0.138	1.561	96.73	0.132	1.511
512	95.10	0.138	1.565	96.55	0.131	1.511

Table C.5: Comparison of KCASA+DFT, coupled with $\mathcal{KH}^t$ and $\mathcal{F}^t$, applied to the VHHS data set.

ω	$\mathcal{KH}^t$			$\mathcal{F}^t$		
	Acc.	FMR	FNMR	Acc.	FMR	FNMR
16	73.83	0.981	1.971	76.95	0.791	1.81
32	87.15	0.102	1.747	93.10	0.087	1.097
64	94.28	0.087	1.120	97.42	0.045	1.085
128	95.43	0.067	1.112	96.49	0.048	1.088
256	95.26	0.067	1.114	97.22	0.046	1.087
512	94.47	0.080	1.119	96.04	0.047	1.089

Table C.6: Comparison of KCASA+DWT, coupled with $\mathcal{KH}^t$ and $\mathcal{F}^t$, applied to the VHHS data set.

ω	$\mathcal{KH}^t$			$\mathcal{F}^t$		
	Acc.	FMR	FNMR	Acc.	FMR	FNMR
16	70.19	1.072	1.947	80.19	0.978	1.871
32	88.63	0.997	1.254	88.63	0.0947	1.102
64	97.09	0.561	1.194	97.09	0.059	1.098
128	96.06	0.581	1.199	96.06	0.064	1.101
256	94.89	0.583	1.201	94.89	0.072	1.101
512	96.79	0.579	1.197	96.79	0.061	1.099