FORENSIC DISCRIMINATION OF PAPERS USING ATR-FTIR SPECTROSCOPY AND CHEMOMETRICS TECHNIQUES

NURINA FATINI BINTI MAZLAN

SCHOOL OF HEALTH SCIENCES

UNIVERSITI SAINS MALAYSIA

2026

FORENSIC DISCRIMINATION OF PAPERS USING ATR-FTIR SPECTROSCOPY AND CHEMOMETRICS TECHNIQUES

by

NURINA FATINI BINTI MAZLAN

Dissertation submitted in partial fulfilment of the requirements for the degree of Bachelor of Science (Honours) Forensic Science

February 2026

ACKNOWLEDGEMENT

Firstly, all praise and thanks are due to Allah SWT for His blessings throughout this journey and for granting me good health, patience, and strength to complete this research until the end. Besides that, I would like to express my sincere appreciation and deepest gratitude to my supervisor, ChM Dr. Dzulkiflee bin Ismail, for his invaluable guidance, support, and dedication throughout the completion of my Final Year Project. I am truly grateful for his time and commitment, which greatly contributed to the successful completion of this project.

I would also like to extend my sincere thanks to all laboratory assistants involved and my fellow friends who conducted their projects under the same supervisor for their kind guidance and helpful support in using laboratory instruments, as well as for providing constructive feedback and suggestions throughout my research journey. In addition, words cannot fully express my gratitude to my best friend, Putri Junaidah, who has always been there to listen and offer unwavering support whenever I needed it.

To my parents and siblings, I am extremely grateful for your unconditional love, support, and motivation that encouraged me to complete my studies. Lastly, I would like to thank myself for successfully completing this research and for not giving up despite all the challenges and experiences, both good and bad, encountered along the way.

TABLE OF CONTENT

CERTIFICATE I
DECLARATION III
ACKNOWLEDGEMENT IV
TABLE OF CONTENT V
LIST OF TABLES VII
LIST OF FIGURES VIII
LIST OF SYMBOLS IX
LIST OF ABBREVIATIONS X
ABSTRAK XII
ABSTRACT XIV
CHAPTER 1: INTRODUCTION 16
1.1 Research Background 16
1.2 Problem Statement 19
1.3 Objectives 20
1.3.1 General Objective 20
1.3.2 Specific Objectives 20
1.4 Scope of study 20
1.5 Significance of Study 20
CHAPTER 2: LITERATURE REVIEW 22
2.1 Introduction 22
2.2 Paper Composition and Manufacturing Process 24
2.2.1 Manufacturing Conditions 26
2.3 Paper Analysis in Forensic Document Examination 27
2.4 Analytical Methods 29
2.4.1 General methods 29
2.4.2 Method Comparison 29
2.4.3 Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) Spectroscopy 31
2.5 Chemometric Analysis 33
2.5.1 Principal Component Analysis 34
2.5.2 Linear Discriminant Analysis 37
CHAPTER 3: METHODOLOGY 40
3.1 Research Design 40
3.1.1 Study scope 40
3.1.2 Sample Collection 41
3.2.2 Data collection 41

3.2.3 Data analysis 43
CHAPTER 4: RESULTS AND DISCUSSION 46
4.1 Repeatability and Reproducibility 46
4.2 Principal Component Analysis 54
4.3 Linear Discriminant Analysis 58
CHAPTER 5: CONCLUSION AND FUTURE RECOMMENDATIONS 62
5.1 Conclusion 62
5.2 Limitations of the study 63
5.3 Future Recommendations 64
REFERENCES 66

LIST OF TABLES

Table 1: Brand of exercise books used 40

Table 2: IR Absorption Band and its Assignment 51

Table 3: Classification Result of LDA for Paper Sample 58

LIST OF FIGURES

Figure 1: General process of manufacturing papers 24

Figure 2: Spectra of three types of white copy paper 35

Figure 3: PCA score plots of three types of white copy paper 36

Figure 4: Flow chart of methodology for this study 39

Figure 5: Repeatability of Spectra within the Book from Same Brand 45

Figure 6: Reproducibility of Spectra within Brand 46

Figure 7: IR Spectrum of Paper Sample from Brand CP 48

Figure 8: IR Spectrum of Paper Sample from Brand KM 49

Figure 9: IR Spectrum of Paper Sample from Brand SBS 49

Figure 10: IR Spectrum of Paper Sample from Brand SW 50

Figure 11: IR Spectrum of Paper Sample from Brand UP 50

Figure 12: PCA Score Plot 54

Figure 13: PCA 3D Scatter Plot 56

Figure 14: LDA Discrimination Plot for Paper Samples 59

LIST OF SYMBOLS

%	percentage
cm^{-1}	inverse centimeter
cm^2	square centimeter
mm	millimeters
n	number of samples
PC1, PC2, PC3	first, second, and third principal components

LIST OF ABBREVIATIONS

ATR	Attenuated Total Reflectance
ATR-FTIR	Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy
$CaCO_3$	Calcium Carbonate
CART	Classification and Regression Trees
CH_3	Methyl
CH_2	Methylene
ClO_2	Chlorine Dioxide
CP	CAMPAP Paper Brand
DP	Discriminating Power
FTIR	Fourier Transform Infrared Spectroscopy
GCC	Ground Calcium Carbonate
H_2O_2	Hydrogen Peroxide
IBM SPSS	IBM Statistical Package for the Social Sciences
IR	Infrared
KM	KAMI Paper Brand
LDA	Linear Discriminant Analysis
LIBS	Laser-Induced Breakdown Spectroscopy
LOO	Leave-One-Out
MCT	Mercury Cadmium Telluride
NaOH	Sodium Hydroxide
Na_2S	Sodium Sulfide
OPUS	Optical User Software

PCA	Principal Component Analysis
PC	Principal Component
PLS-DA	Partial Least Squares Discriminant Analysis
RSD	Relative Standard Deviation
SiC	Silicon Carbide
SBS	SBS Paper Products
SNV	Standard Normal Variate
SVM	Support Vectors Machines
SW	Smart Write Paper Brand
TiO_2	Titanium Oxide
ToF-SIMS	Time-of-Flight Secondary Ion Mass
UP	Uni Paper Products
USM	Universiti Sains Malaysia
VSC	Video Spectral Comparators
XRF	X-ray Fluorescence

DISKRIMINASI FORENSIK KERTAS MENGGUNAKAN SPEKTROSKOPI ATR-FTIR DAN TEKNIK KEMOMETRIK

ABSTRAK

Kertas merupakan satu jenis bahan bukti penting dalam pemeriksaan dokumen forensik yang mampu memberikan banyak maklumat berkaitan ketulenan dokumen serta asal-usulnya. Keupayaan untuk membezakan kertas berdasarkan asal jenama boleh dimanfaatkan dalam siasatan forensik yang melibatkan kes pengubahsuaian, penggantian atau penipuan dokumen. Walau bagaimanapun, pembezaan bahan kertas moden adalah sukar kerana komposisi kimianya yang hampir seragam dan kebanyakannya terdiri daripada selulosa. Oleh itu, kajian ini dijalankan bagi menilai keberkesanan teknik spektroskopi Attenuated Total Reflectance Fourier Transform Infrared (ATR-FTIR) yang digabungkan dengan kaedah kemometrik dalam membezakan kertas buku latihan daripada pelbagai jenama komersial. Dalam kajian ini, 5 jenama buku latihan telah dianalisis iaitu CAMPAP, Smart Write, KAMI, Uni Paper Products dan SBS Paper Products. Spektroskopi ATR-FTIR dijalankan pada beberapa titik berbeza pada sampel kertas-kertas ini bagi memperoleh spektrum inframerah dengan mengambil kira ketidakhomogenan permukaan serta variasi instrumen. Data spektrum dipraproses menggunakan kaedah Standard Normal Variate (SNV) dan dianalisis dalam kawasan cap jari antara 1500–600 cm^{-1}. Pengurangan dimensi dan pengecaman corak dilakukan menggunakan Analisis Komponen Prinsipal (PCA), manakala pengelasan terselia dilaksanakan menggunakan Analisis Diskriminan Linear (LDA). Keputusan PCA menunjukkan bahawa tiga komponen prinsipal utama (PCs) menerangkan sebanyak 93% daripada jumlah varians (PC1 61.6%, PC2 21.9% dan PC3 9.5%), yang membuktikan bahawa sebahagian besar maklumat spektrum telah dikekalkan dengan baik. Plot skor PCA memperlihatkan kecenderungan pengelompokan berdasarkan jenama kertas, walaupun terdapat beberapa sampel yang bertindih

secara separa. Model LDA seterusnya berjaya mengelaskan sampel dalam kumpulan asal dengan ketepatan 100%, manakala pengesahan silang menggunakan kaedah leave-one-out (LOU) menghasilkan ketepatan ramalan sebanyak 73.3%, yang menunjukkan keupayaan diskriminasi sederhana terhadap sampel tidak diketahui. Secara keseluruhannya, hasil kajian menunjukkan bahawa gabungan spektroskopi ATR-FTIR dengan PCA dan LDA merupakan teknik yang tidak memusnahkan, objektif dan boleh dipercayai untuk pembezaan serta pengelasan kertas berdasarkan asal jenama. Rangka kerja analisis ini mempunyai potensi tinggi sebagai pendekatan tambahan dalam pemeriksaan dokumen forensic melibatkan kertas.

FORENSIC DISCRIMINATION OF PAPERS USING ATR-FTIR SPECTROSCOPY AND CHEMOMETRICS TECHNIQUES

ABSTRACT

Paper is a significant type of forensic document evidence, which can reveal a lot of information about the authenticity of documents and their provenience. The fact that the paper can be distinguished on the basis of brand origin can be used in the forensic investigations in the cases of document alteration, substitution, or fraud. Discrimination of modern paper materials is however difficult because they are similar in chemical composition which is mostly composed of cellulose. The research was to assess how effective Attenuated Total Reflectance fourier transform infrared (ATR-FTIR) spectroscopy in conjunction with chemometric techniques is in distinguishing between exercise book paper and various commercial brands. In the study, the exercise books which were analyzed were five brands: CAMPAP, Smart Write, KAMI, Uni Paper Products and SBS Paper Products. ATR-FTIR spectroscopy was performed on different spots of the paper samples to collect infrared spectra in order to consider surface heterogeneity and instrumental variation. Standard Normal Variate (SNV) was used to preprocess spectral data and analyze the spectral data in the fingerprint region 1500-600 cm^{-1}. The dimensionality reduction and pattern recognition were implemented using Principal Component Analysis (PCA), and the supervised classification was implemented using Linear Discriminant Analysis (LDA). The PCA output indicated that the top three principal components explained 93% of the total variance (PC1 61.6%, PC2 21.9% and PC3 9.5%), which showed that the majority of the information on spectral was well preserved. PCA score plots indicated the trends of clustering based on paper brand, but there also were a few samples partially overlapping. The next LDA model was successful in classifying the original grouped

samples with 100% accuracy and cross-validation with the leave-one-out method was successful with 73.3% prediction accuracy, which is moderate discrimination ability on unknown samples. All in all, the results suggest that ATR-FTIR spectroscopy with PCA and LDA offers a non-destructive, objective and a reliable technique of discrimination and classification of paper according to brand origin. It is a powerful analytical framework that has high potential as an addition to forensic document examination.

CHAPTER 1: INTRODUCTION

1.1 Research Background

Forensic document examination (FDE) is a specialized discipline within forensic science responsible for the scientific analysis of questioned documents (QDs). Questioned documents include various types of documents including disputed signatures, fraudulent contracts, counterfeit currency, threatening letters, and forged security documents (Gorziza et al., 2022). These documents can be created manually by handwriting or by advanced modern technologies such as printers and photocopiers (Sharma et al., 2024). Paper is the basic substrate for printing and handwriting documents and paper analysis is frequently required in investigations, especially when authenticity of a document is questioned (Kumar et al., 2017; Lee et al., 2025; Zięba-Palus et al., 2020).

Identifying whether a questioned document page has a similar composition to the rest of the document is an important investigation topic among forensic investigators to justify or disprove fraud (Farid et al., 2021). Paper can show evidence of page replacement, addition or removal in documents. Moreover, paper analysis can be useful in case of matching or differing documents, including anonymous letters, with the items that have been obtained at the crime scene or their possible suspects (Lee et al., 2025). The information concerning the paper origin or manufacturer is a critical component of this process that can give hints to the further investigation (Qin et al., 2023). Given the importance of documents in legal proceedings, the techniques used by forensic examiners should be reliable, objective, and non-destructive in order to preserve the integrity of the evidence (Farid et al., 2021; Kumar et al., 2017; Sharma et al., 2021; Sharma et al., 2024).

Paper is a heterogeneous, multi-component material (Yuan et al., 2025; Zięba-Palus et al., 2020). Its major structural component is cellulose; a biopolymer consisting of linear units

of $\beta(1\rightarrow4)$ D-glucopyranosyl units. Interactions between and within cellulose chains, primarily controlled by the formation of inter- and intramolecular hydrogen bonds, are responsible for the properties of the paper (Zięba-Palus et al., 2020). However, modern paper also contains a complex of non-cellulose elements which are vital for the purpose of distinguishing one product from another. Many additives have been used by manufacturers with the purpose of attaining a particular effect, including inorganic fillers, such as calcium carbonate, kaolin, talc, gypsum, chalk, to meet required attributes, including the ability to be opaque, bright, and smooth (Song et al., 2022).

This compositional diversity, which is inherent to all paper products and is caused by differences in raw materials, manufacturing processes, and the amounts/qualities of additives, gives each paper product a characteristic physicochemical signature or fingerprint (Yuan et al., 2025). Forensic examination takes advantage of these differences to describe and distinguish paper materials (Sharma et al., 2024). However, it can be difficult to differentiate common white copy papers since manufacturers often use similar primary raw material (cellulose) and the differentiating additive materials are very often found in tiny amount (Lee, 2021). This complexity requires sophisticated analytical techniques that would identify slight chemical variations.

Vibrational spectroscopy, specifically the Infrared (IR) spectroscopy, have become the main analytical instruments to respond to the need of non-destructive analysis in FDE (Farid et al., 2021; Sharma et al., 2024; Zięba-Palus et al., 2020). Fourier Transform Infrared (FTIR) spectroscopy is a molecular analysis method applied to define and characterise chemical structures with great accuracy and sensitivity (Khofar et al., 2022). In particular, the ATR mode has become of interest in the characterization of paper since it is a non-destructive technique that is effective and allows investigating the surface of the sheet (Farid et al., 2021; Geminiani et al., 2022; Sharma et al., 2024). It also offers qualitative and quantitative data on organic and

inorganic components, which enables the identification of functional groups that arise due to the usage of various components of the paper (Khofar et al., 2022; Lee, 2021).

However, spectral interpretation by purely visual inspection may not suffice, especially in the case of comparison between physically and chemically similar white copy papers (Lee, 2021). The inherent problem is that the FTIR spectra often have multiple, incompletely resolved and broad absorbance peaks that overlap each other to a large extent, particularly the fingerprint region (2000 to 400 cm^{-1}) (Farid et al., 2021; Kumar et al., 2017; Xia et al., 2019). To overcome this difficulty and bring out meaningful differences in the complex spectral data, the use of chemometric tools is indispensable (Qureshi, 2022; Sharma et al., 2021). Chemometrics is the use of mathematical and statistical techniques applied to chemical data to obtain objective, accurate and reproducible data essential for legal proceedings (Gorziza et al., 2022; Kumar et al., 2017). In the context of forensic document analysis, chemometrics are applied to manage large data sets with multiple variables produced as a result of spectroscopic techniques (Qureshi, 2022; Sharma et al., 2021).

Principal Component Analysis (PCA) is one of the most frequently used unsupervised chemometric methods implemented in association with ATR-FTIR spectroscopy in document examination (Farid et al., 2021; Sharma et al., 2021; Sharma et al., 2024; Zięba-Palus et al., 2020). PCA works as a dimensionality reduction technique, which reduces the high-dimensionality of data with thousands of wavenumbers into a smaller set of orthogonal axes known as Principal Components (PCs) (Gawad et al., 2022; Lee, 2021; Lee et al., 2025). The use of PCA score plots makes it possible to visualize the underlying patterns and data structures (Lee, 2021; Lee et al., 2025). Samples having similar chemical characteristics are clustered closely together, allowing for the objective grouping and differentiation of different paper types (Gawad et al., 2022).

1.2 Problem Statement

The handwriting or ink analysis is usually used to determine whether documents are authentic and of the origin in forensic document examination (Devlin et al., 2024; Gawad et al., 2022). However, the physical paper itself can also be the source of information that is equally important, especially in revealing documents in which a page or more might have been replaced, added or removed to a contract (Farid et al., 2021; Lee et al., 2025; Lee et al., 2025). Papers produced by different manufacturers are expected to have varying chemical fingerprints, but discrimination is often challenging because most manufacturers use the same raw materials (cellulose) and the subtle variations contributed by additives are often minute (Lee, 2021). Although paper properties are the key to its investigation, there has been little research on the paper characterization, especially in consumer goods like exercise books, which do not have visible batch markings (Mamontov, 2020).

Generally, standard printing or copy paper cannot be differentiated just by a mere glance. Prior research has indicated that there are a number of issues that are experienced in the procedure of paper analysis such as overlay of large bands of numerous functional groups that are seen in ATR-FTIR spectrums (Farid et al., 2021; Silva et al., 2022). Others stated that the presence of cellulose components and inhomogeneities in the surface resulted in the difficulties in the analysis (Wertz et al., 2022). This makes a major problem in forensic investigations when the capability to identify paper and trace it to its origin would be used to confirm the authenticity of documents or identify substitution (Lee et al., 2025).

Thus, the proposed research would determine the efficiency of ATR-FTIR spectroscopy in the differentiation of paper belonging to various brands and batches of exercise books, and this could be used to develop a non-destructive and highly accurate system of objective comparison and analysis of forensic paper.

1.3 Objectives

1.3.1 General Objective

The general objective of this study is to test the ability of ATR-FTIR spectroscopy to distinguish between paper from various brands of exercise books.

1.3.2 Specific Objectives

- To describe the ATR-FTIR spectra of the paper samples of various brands of exercise books.
- To identify the characteristic absorption peaks and spectral differences among the paper samples.
- To evaluate the potential of ATR-FTIR spectroscopy coupled with chemometric analysis as a non-destructive analytical tool for paper discrimination in forensic document examination.

1.4 Scope of study

The exercise books selected for this study were new and unused, one example for each of the five different brands. Each sample is kept in a contamination-free storage environment prior to analysis. In order to retain the integrity of the sample, the paper sheets were not chemically treated, and were not sectioned physically or through extraction. The spectral data is collected using ATR-FTIR and processed to remove any noise. In order to ensure reliability and accuracy for statistical evaluation, a baseline shift is corrected and scattering is minimized. After that, Principal Component Analysis (PCA) is used for data analysis in order to visualize potential clustering and disparity among the five brands of exercise books, as well as to decrease the dimensionality of the spectral data set. Then, the paper samples are subjected to Linear Discriminant Analysis (LDA) to improve classification and differentiation.

1.5 Significance of Study

The main importance of the study is to provide an alternative to the existing forensic tools that cannot always deliver a certain answer to the cases related with the document forgery. The method facilitates the determination of functional groups and offers both qualitative and quantitative data of organic and inorganic paper components. ATR-FTIR is a fully non-destructive method, which is crucial in maintaining the integrity of the evidence in the future testing or trial in court. It is considered a rapid process that does not need extensive or no advanced sample preparation, thereby conserving time in the investigative research.

This study can play a major role in proving or ruling out fraud in situations where a page or pages are being added, replaced, or removed to multi-page documents such as contracts or wills. The effectiveness of this approach enables forensic specialists to make a conclusive decision on disputed documents. It can allow investigators to match or compare documents discovered at a crime scene, which are anonymous, with paper recovered at a crime scene or at the location of the suspect by establishing the chemical fingerprint of the paper.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

Forensic document examination (FDE) often entails the examination of paper, which forms the primary foundation of documents, including questioned and suspecting fraudulent ones (Lee et al., 2025; Zięba-Palus et al., 2020). The physical and chemical properties of the questioned paper, when compared to the other portions of the text or known samples, are decisive in the argument for or against falsification (Farid et al., 2021). In paper analysis, the study of cellulose, fillers, dyes, pigments and resins enables forensic scientists to get a distinctive chemical fingerprint of a particular paper, which is essential in the process of discrimination and verification (Gawad et al., 2022; Khofar et al., 2022; Sharma et al., 2024).

Paper is a multi-component material with a complex structure comprising a structural matrix and numerous additives (Mamontov, 2020; Yuan et al., 2025). The major component is cellulose fibers, which are usually extracted using the pulping and bleaching of wood or non-wood agricultural residues (Pavithra et al., 2015; Sharma et al., 2020). It contains inorganic components like calcium carbonate ($CaCO_3$), kaolin (clay), talc, and titanium dioxide (TiO_2) to enhance the opacification, brightness, and smoothness (Silva et al., 2022; Zięba-Palus et al., 2020). These elements form a specific physico-chemical signature, which is the key to forensic discrimination (Farid et al., 2021; Sharma et al., 2024).

The current methods of forensic investigations are based on a tier approach, in which non-destructive methods are applied initially, and resorting to micro-destructive ones is taken into account (Sharma et al., 2024; Yuan et al., 2025). Optical inspection is typically performed through the use of Video Spectral Comparators (VSC) to view transmitted light, UV luminescence, and fiber morphology (Farid et al., 2021; Lee et al., 2025; Tomar et al., 2021). In spectroscopic, ATR-FTIR spectroscopy can be used to give a surface-level organic

fingerprint, which is especially practical in determining cellulose, lignin, and sizing agents (Khofar et al., 2022; Wertz et al., 2022; Yuan et al., 2025). Other methods, including Laser-Induced Breakdown Spectroscopy (LIBS) and X-ray Fluorescence (XRF), can give fast elemental profiling of inorganic fillers, including Ca, Ti, and Al (Arbi et al., 2024; Yuan et al., 2025). In other respects, advanced imaging can be used to identify the periodic marks, including weave and drainage prints, that are transferred onto the paper surface by the forming fabrics of a paper-making machine, which are a sort of fingerprints of individual manufacturers (Lee et al., 2025).

The use of chemometric analysis is a necessary aspect of forensic science because current analytical tools produce large and complicated data that is frequently beyond the processing ability of the human eye (Qureshi, 2022; Sharma et al., 2024). Chemometrics offers an objective, unbiased, and reproducible method of analysis that is essential when presenting evidence in a court of law (Gorziza et al., 2022; S. Sharma et al., 2021). Paper is indistinguishable by the eye in most instances because of spectral overlap. Dimensionality reduction with the use of chemometrics is needed, reducing thousands of variables to a few meaningful components without losing the needed information (Lee, 2021; Xia et al., 2019).

Principal Component Analysis (PCA) is the most widely used in FDE (Xia et al., 2019). It transforms spectral data into Principal Components (PCs) to be represented as sample clusters and to determine outliers (Gorziza et al., 2022; Sharma et al., 2021). PCA is applied to a large set of variables so that smaller amounts of uncorrelated variables are obtained and the large majority of the variance in the data is explained (Lee et al., 2025; Sharma et al., 2024). The decreased dimensions and computational complexity improve the interpretation and performance of the analysis (Kapoor et al., 2024; Lee et al., 2025).

2.2 Paper Composition and Manufacturing Process

The structure of paper is multidimensional and mainly depends on the use of plant fibers. Its composition is heterogeneous and complex, so it provides the basis for the distinction between different products and manufacturers (Yuan et al., 2025). The main component that gives the structural integrity of the paper is cellulose which is a biopolymer made of linear $\beta(1\rightarrow4)$ D-glucopyranosyl units (Farid et al., 2021). The inherent characteristics of paper are mainly related to the existence of interactions, mainly of inter- and intramolecular hydrogen bonds, between and inside these cellulose chains (Zięba-Palus et al., 2020). The details of its properties depend strongly on the method of manufacturing and the type of additives (Mamontov, 2020).

The basic mix is comprised of a primary source of fibers, various types of mineral fillings and additives. In order to predict the spectrum of the paper, it is important to know the process involved in making a paper, including additives added for each step. The general process of manufacturing paper, which is applicable for the grades such as writing paper used for exercising books, can be broadly put in to three main stages as shown in Figure 1. The first stage is pulp making or pulping, then pulp processing, which includes process of screening, washing and bleaching, and lastly paper-making (Haile et al., 2021; Kamali and Khodaparast, 2015; Sharma et al., 2020).