

Article

Not peer-reviewed version

Non-Destructive Species Discrimination of Japanese Bast Fibers: A Feasibility Study Using Micro-Hyperspectral Imaging and Chemometrics

[Yexin Zhou](#)*, [Yoichi Ohyanagi](#), Akiko Iwata, [Koji Shibazaki](#), Kazuhito Murakami

Posted Date: 25 March 2026

doi: 10.20944/preprints202603.2031.v1

Keywords: non-destructive testing (NDT); micro-hyperspectral imaging (Micro-HSI); paper fiber identification; chemometrics; cultural heritage conservation; bast fiber identification



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Non-Destructive Species Discrimination of Japanese Bast Fibers: A Feasibility Study Using Micro-Hyperspectral Imaging and Chemometrics

Yexin Zhou ^{1,*}, Yoichi Ohyanagi ², Akiko Iwata ², Koji Shibazaki ² and Kazuhito Murakami ³

¹ Life and Culture Department, Seirei Women's Junior College, Akita 011-0937, Japan

² Faculty of Fine Arts, Aichi University of the Arts, Nagakute 480-1194, Japan

³ School of Information Science and Technology, Aichi Prefectural University, Japan

* Correspondence: zhou@gml.akita-seirei.ac.jp

Abstract

Accurate paper fiber identification is crucial for cultural heritage conservation. To address the destructive nature of traditional staining and the “black-box” limitations of macroscopic AI models, this study explores the feasibility of a non-destructive testing paradigm using micro-hyperspectral imaging (Micro-HSI). Three traditional Japanese pure bast fibers (Kozo, Mitumata, and Gampi) were analyzed as standard samples. Raw relative reflectance spectra from microscopic regions of the fibers were extracted via Micro-HSI. Dynamic normalization and Savitzky–Golay first-derivative filtering were applied to suppress scattering and baseline drift. Principal component analysis (PCA) and linear discriminant analysis (LDA) were subsequently employed for dimensionality reduction and supervised classification. The results showed that while unsupervised PCA suffered from inter-class overlap due to shared cellulose-dominated structures, supervised LDA amplified weak chemical fingerprint differences, achieving complete class separation of the highly similar fibers. Analysis of the feature loadings confirmed that the classification relies on the visible-range reflectance baseline, lignin $\pi \rightarrow \pi^*$ transition absorption (400–450 nm), and near-infrared $O-H$ and $C-H$ overtone vibrations (~835 nm). This proof-of-concept study demonstrates that combining Micro-HSI with chemometrics enables high-precision, non-destructive fiber separation while retaining rigorous physicochemical interpretability, thereby providing an optical reference baseline for future historical paper analysis.

Keywords: non-destructive testing (NDT); micro-hyperspectral imaging (Micro-HSI); paper fiber identification; chemometrics; cultural heritage conservation; bast fiber identification

1. Introduction

1.1. The Materiality of Paper Artifacts and the Urgency of Conservation

Paper is both a primary carrier of historical records and a structurally vulnerable organic substrate. Moisture fluctuation, temperature cycling, light, and microbial activity degrade it over time—cleaving cellulose chains and oxidizing lignin—with measurable consequences for mechanical stability and visual legibility [1].

Effective restoration requires fiber compatibility between repair materials and the original substrate; mismatches in chemistry or physical behavior introduce secondary damage such as tearing or discoloration at the repair interface. For many historical papers, however, source fibers remain poorly documented. Broad labels—hemp, bark fiber, bamboo—are frequently the only available descriptors, and these categories do not resolve the species-level chemical and structural differences that vary with plant origin and pulping method [2]. At the identification resolution required for conservation work, that gap is a practical problem.

1.2. Limitations of Traditional Fiber Identification

Standard fiber identification relies on destructive staining—most commonly the JIS P8120 iodine-zinc chloride test—combined with optical microscopy [3]. Neither approach is well-suited to heritage objects. Physical sampling removes material that cannot be replaced, and the morphological route has its own ceiling: beating and aging routinely eliminate the surface features—lumens, pits, cross-markings—that identification depends on. For closely related bast fibers such as mulberry species, even intact samples present ambiguous morphology [4].

As a result, curators frequently have no choice but to label valuable historical papers as having an "undetermined fiber composition."

1.3. Spectroscopic Approaches for Fiber Discrimination

Spectroscopic techniques provide an alternative approach for identifying bast fibers based on their molecular composition. Reflectance and infrared spectroscopy have been widely applied to characterize natural fibers in textiles and plant materials. Previous studies demonstrated that spectral reflectance combined with multivariate analysis can effectively distinguish different natural textile fibers in a non-destructive manner [5].

Similarly, infrared spectroscopy has been used to discriminate traditional bast fibers employed in cultural artifacts. Combining spectral measurements with chemometric analysis enables the detection of subtle biochemical differences between bast fibers [6].

These studies demonstrate the potential of optical spectroscopy for fiber classification. However, most spectroscopic measurements rely on bulk samples or macroscopic measurements, which may include mixed materials or background interference from paper structures.

1.4. Hyperspectral Imaging and the Need for Micro-Scale Analysis

Hyperspectral imaging (HSI) has recently emerged as a powerful tool for the non-destructive analysis of cultural heritage materials. By recording spectral information for each pixel in an image, HSI enables simultaneous spatial and spectral analysis. Recent studies have successfully applied short-wave near-infrared hyperspectral imaging to predict chemical properties and visualize compositional variations in historical paper [7]. In addition, hyperspectral data combined with multivariate statistical models such as principal component analysis (PCA) can improve classification performance in complex spectral datasets [8].

Alongside chemical methods, macroscopic AI screening provides a rapid, non-destructive alternative for paper tracing. As we demonstrated in our previous study on patch-based classification [9], macro image networks handle high-throughput preliminary screening very well. However, this approach hits a distinct resolution threshold when applied to complex, mixed-material papers. Because macroscopic RGB sensors capture only morphological and textural features, their diagnostic power is severely compromised the moment fibers are physically degraded or heavily blended during manufacturing.

Despite these technological advances, most existing hyperspectral and AI studies focus on the macroscopic imaging of entire documents. To overcome macroscopic morphological ambiguity and accurately identify mixed components, it is imperative to shift the analytical focus from macro-texture to micro-chemistry. However, the intrinsic spectral signatures extracted purely from the microscopic regions of the fibers within paper structures remain poorly understood.

1.5. Research Aim

To address this limitation, this study explores the feasibility of using micro-hyperspectral imaging (Micro-HSI) for the identification of bast fibers at the microscopic level. The research is conducted as part of the Grant-in-Aid for Scientific Research project "Elucidation of the Paper Road by data science – Based on Quantitative, Qualitative research and AI Multidimensional analysis".

Three traditional Japanese bast fibers—Kozo, Mitsumata, and Gampi—were selected as standard samples. By extracting spectral signatures from microscopic regions of the fibers and applying chemometric analysis, this study aims to establish a microscopic spectral reference for fiber identification and to evaluate the potential of Micro-HSI as a non-destructive diagnostic tool for cultural heritage conservation.

2. Materials and Methods

2.1. Optical and Chemical Basis of Bast Fiber Spectroscopy

The VNIR optical response of bast fibers is governed by three biopolymers—cellulose, hemicellulose, and lignin—whose contributions are chemically distinct but spectrally overlapping. Cellulose and hemicellulose together dominate the bulk signal: cellulose is largely transparent across the visible range, while its $O-H$ and $C-H$ bonds generate overtone and combination absorption bands in the NIR [10]; hemicellulose adds a broadly similar hydroxyl and ether signature that blends into the cellulose baseline without introducing sharp discriminating features [10]. The chemical contrast needed for fiber identification comes primarily from lignin. Aromatic phenols in lignin undergo $\pi \rightarrow \pi^*$ electronic transitions that strongly attenuate reflectance in the UV and blue-violet region—a response that varies with lignin retention across species and pulping conditions [11]. Because cellulose composition is nearly identical across bast fiber types, residual lignin distribution is what makes spectral discrimination possible. Hyperspectral imaging in the VNIR range captures these subtle reflectance differences at the spatial resolution of individual filaments, converting species-level biochemical variation into measurable spectral contrast without any physical sampling.

Extracting these localized spectral fingerprints directly traces the original plant material without ever damaging the artifact.

2.2. Micro-Hyperspectral Imaging Principles

Micro-hyperspectral imaging (Micro-HSI) combines microscopic imaging with spectral acquisition to generate a three-dimensional data cube, $I(x, y, \lambda)$. Each spatial pixel contains a complete reflectance spectrum across the measured wavelength range. Unlike macroscopic spectrometers, Micro-HSI allows for the precise selection of microscopic regions of the fibers as regions of interest (ROIs), thereby effectively circumventing the diffuse reflection background interference caused by air voids in the paper.

In this study, relative reflectance spectra were calculated using a standard white reference panel according to:

$$R(x, y, \lambda) = \frac{I_{\text{sample}}(x, y, \lambda)}{I_{\text{white}}(x, y, \lambda)}$$

where I_{sample} represents the recorded intensity from the fiber sample and I_{white} represents the intensity from the reference panel under identical illumination conditions. (Note: The constant additive noise from the sensor's dark current was not physically subtracted here, as it is mathematically eliminated during the subsequent first-derivative preprocessing stage, detailed in Section 2.6).

2.3. Sample Selection and Preparation

Our analysis centers on Kozo (*Broussonetia papyrifera*), Mitsumata (*Edgeworthia chrysantha*), and Gampi (*Diplomorpha sikokiana*)—the three primary bast fibers in Japanese papermaking. We acquired all materials from a fully authenticated 1973 archive (Encyclopedia of Handmade Japanese Paper, The Mainichi Newspapers).

To establish a robust optical baseline, we utilized six 100% pure paper specimens per species, all free of fillers and chemical additives. By restricting the reference set to unprocessed materials, we ensured the resulting spectra accurately reflect intrinsic fiber chemistry rather than industrial variables—a prerequisite for extending this methodology to historically aged or mixed-pulp samples.

2.4. Experimental Setup and Hardware Configuration

Data acquisition relied on a custom Micro-HSI setup: a push-broom hyperspectral camera (NH-9, EBA Japan, Tokyo) mounted on an upright metallurgical microscope (ECLIPSE LV100ND, Nikon, Tokyo) through a standard C-mount. The optical path operated in reflection mode, lit by a stabilized 12 V–50 W halogen lamp fixed at a 45° angle. The detector covers the 350–1100 nm range at a 5 nm resolution. To cleanly separate the microscopic fiber body from the empty background voids, we captured every hyperspectral cube through a 50× objective lens. The technical specifications of the imaging platform are summarized in Table 1.

Table 1. Technical specifications of the Micro-HSI platform.

Parameter	Specification
Camera type	Push-broom HSI
Spectral range	350 nm–1100 nm (covering UV-Vis-NIR)
Spectral resolution	5 nm
Detector resolution	2048 (H) × 1080 (V) pixels
Microscope platform	Nikon ECLIPSE LV100ND
Objective lens	50× (NA 0.8)
Illumination	12 V-50 W halogen lamp

2.5. Data Acquisition and Standardization Protocol

All observations were standardized under invariant magnification, illumination, and exposure settings. During the calibration phase, hyperspectral data from a standard white panel was registered as the system baseline, allowing the software to automatically convert the raw fiber images into calibrated relative reflectance. ROI selection was performed manually on these calibrated files to isolate pure fiber signatures. Sampling was strictly confined to the paraxial region and the primary focal plane. We specifically selected the bulky, central segments of the fiber bodies to bypass edge-diffraction artifacts, while carefully dodging air voids and surface impurities (Figure 1).

Ten ROIs were sampled from each specimen and averaged for the final CSV export. For each fiber type, six independent paper specimens were analyzed, yielding a total of 60 spectra per class.

A specific quirk of the proprietary software (HSDAnalyzer) is that it scales 100% relative reflectance to a raw integer of 4000. However, the export automatically generates a %Average(/4000x100) column, which restores the true percentage. These pre-calibrated values were parsed directly into our chemometric models.

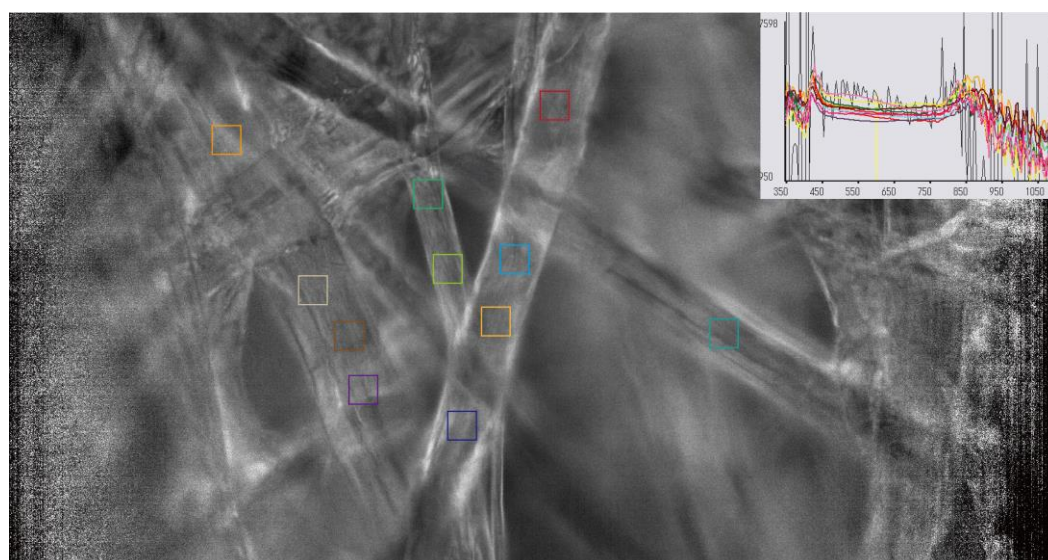


Figure 1. Micro-HSI data cube and ROI sampling workflow, using Kozo paper fibers at 50× magnification as a representative example. The image comprises a microscopic grayscale view of the fiber network, superimposed with a schematic map showing the 10 specific Regions of Interest (ROIs) manually selected from the bulky, central fiber segments. Integrated into the top-right corner of the figure is a plot displaying the unprocessed, raw relative reflectance profiles corresponding to these 10 ROIs, as directly obtained within the HSDAnalyzer software interface prior to any pre-processing.

2.6. Spectral Pre-Processing and Chemometrics

All pre-processing and chemometric steps were implemented in custom Python 3.9 scripts using SciPy and scikit-learn, with Processing 4.2 handling visualization. Raw relative reflectance spectra were normalized dynamically to a common scale. Coefficient of variation (CV) was then computed across all spectral bands to evaluate ROI stability and intra-class dispersion; this calculation identified the 400–1000 nm window as the range with the most favorable signal-to-noise ratio. Spectral data outside this range were excluded from further analysis.

The uneven thickness and surface topology of handmade paper introduce baseline drift unrelated to fiber chemistry. A Savitzky–Golay (S–G) first-derivative filter—11-point window, second-order polynomial—was applied to the truncated spectra to address this [12]. Differentiation removes any constant additive offset, eliminating both scattering-induced baseline shifts and dark-current contributions from the camera sensor. Resolution of faint chemical absorption features is also improved, as derivative transformation sharpens inflection points that are otherwise obscured by the background slope.

Derivative spectra were subsequently input into a sequential PCA–LDA pipeline. Principal Component Analysis (PCA) was applied without class labels, reducing the dataset to three principal components that captured the dominant spectral variance. These components served as input features for Linear Discriminant Analysis (LDA), which maximized between-class separation across the three fiber types. Unsupervised dimensionality reduction alone was insufficient to resolve Kozo, Mitsumata, and Gampi, given their shared cellulose-dominated variance structure; the supervised LDA step was required to project the data into a space where the three classes formed distinct, non-overlapping clusters.

3. Results

3.1. Raw Spectral Signatures and Baseline Calibration

Uncalibrated spectra showed broad amplitude variation and heavily overlapping SD envelopes across all three fiber types (Figure 2), consistent with light scattering introduced by the uneven thickness and surface roughness of the fiber network. Dynamic normalization reduced this dispersion and rescaled relative intensities, making species-specific spectral features resolvable.

Between 450 and 750 nm, the three fibers diverge visibly (Figure 3). Kozo (red) rises steadily across the visible range, reaching its relative maximum near 750 nm. Gampi (green) follows a flatter, slightly concave trajectory at moderate reflectance. Mitsumata (blue) records the lowest overall visible-range reflectance among the three, but is distinguished by higher-frequency fluctuation and a localized reflectance peak at 425 nm that exceeds both Kozo and Gampi at that wavelength—a feature not observed in the other two species.

Above 750 nm, all three fibers show a step-like reflectance increase peaking near 850 nm. Their relative ranking shifts beyond 900 nm: Mitsumata reflectance progressively exceeds the group mean, while Kozo drops to the lowest relative position in this region.

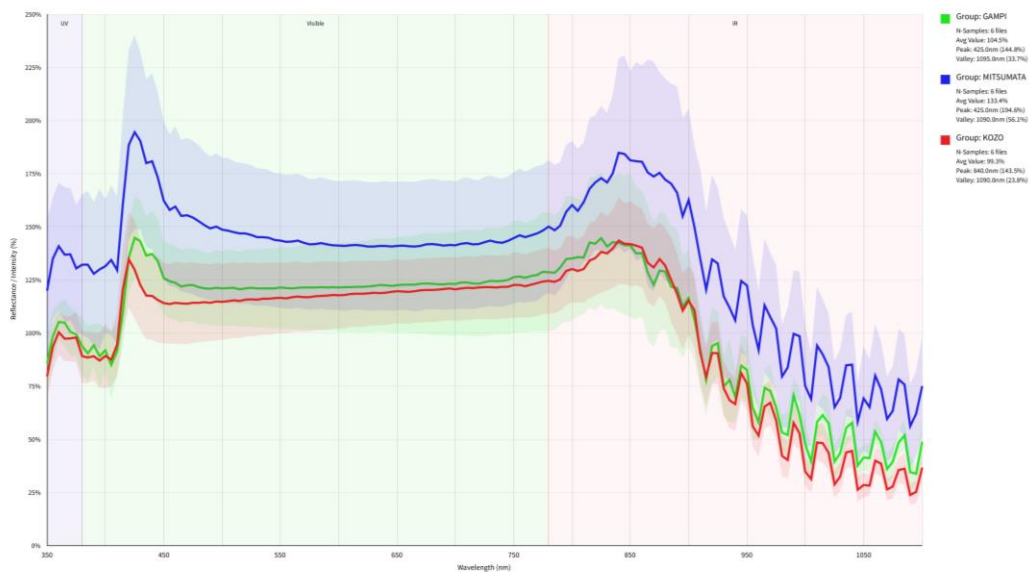


Figure 2. Uncalibrated relative reflectance spectra of Kozo (red), Mitsumata (blue), and Gampi (green) prior to dynamic normalization. Solid lines indicate mean reflectance; shaded regions indicate SD envelopes across sampled ROIs. In the visible range, SD envelopes overlap substantially across all three species, consistent with light scattering from fiber surface roughness. Kozo shows the narrowest intra-class dispersion; Mitsumata the widest; Gampi intermediate.

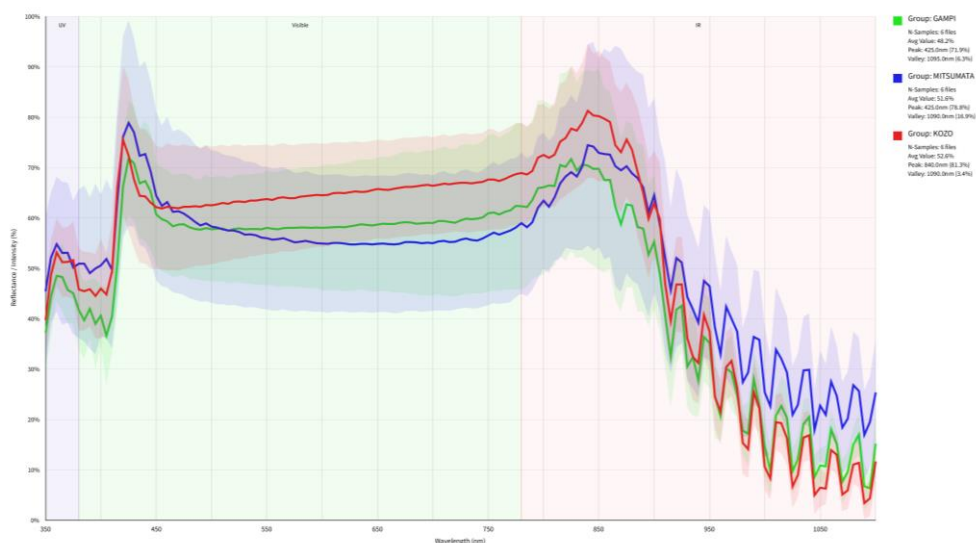


Figure 3. Dynamically normalized relative reflectance spectra of Kozo (red), Mitsumata (blue), and Gampi (green). Solid lines indicate mean reflectance; shaded regions indicate SD envelopes across sampled ROIs. Following normalization, the mean spectral profiles become more distinguishable—particularly the concave visible-range profile of Mitsumata and the steady linear rise of Kozo—while the SD envelopes show increased overlap compared to the uncalibrated spectra. This pattern reflects the intended effect of normalization: rescaling suppresses absolute intensity differences driven by physical scattering, exposing the underlying spectral shape of each fiber type at the cost of increased relative dispersion.

3.2. Physical Scattering Interference and Spectral Normalization

Coefficient of variation (CV) was computed across the full spectral range to evaluate data consistency (Figure 4). In the visible region (450–750 nm), CV remained between 13% and 24%, indicating that dynamic normalization effectively suppressed scattering-induced variance and preserved consistent intra-class spectral profiles.

Beyond 760 nm, CV increased sharply, exceeding 50% at several wavelengths. This elevation is attributed to reduced detector sensitivity combined with weak $C-H$ and $O-H$ overtone signals in this region. Inter-species variation in CV was also observed: Mitsumata showed the highest mean CV (27.53%), reflecting greater structural heterogeneity at the fiber level, while Kozo was the most uniform (19.74%). Based on these findings, subsequent chemometric modeling was confined to the 400–1000 nm range to maintain an acceptable signal-to-noise ratio while retaining the principal VNIR absorption features.

3.3. Unsupervised Dimensionality Reduction and Inter-Class Overlap

PCA was applied as an initial unsupervised dimensionality reduction step. The first three principal components accounted for 62.50% of total variance (PC1: 28.90%, PC2: 22.97%, PC3: 10.63%); however, 2D score plots revealed substantial inter-class mixing across all PC pairs, with 95% confidence ellipses overlapping for all three fiber types (Figure 5a–c).

The overlap is consistent with the dominant contribution of the shared cellulose matrix to spectral variance. Where bulk polysaccharide structure drives much of the signal, PCA has insufficient sensitivity to resolve the trace lignin and hemicellulose differences that distinguish the three species. This behavior is a recognized limitation of unsupervised methods applied to closely related cellulosic materials and indicates that supervised classification is necessary for species-level discrimination [12].

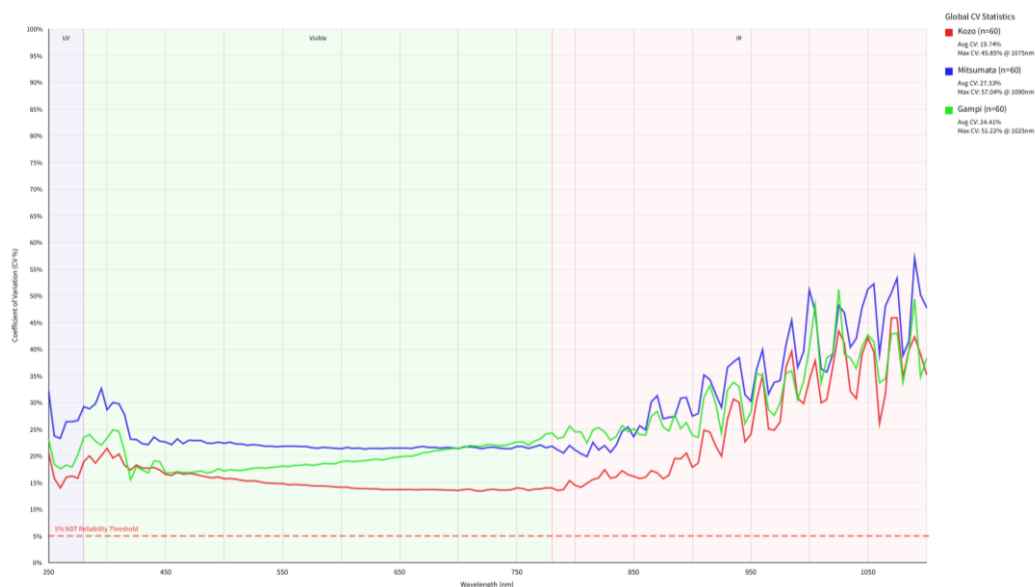
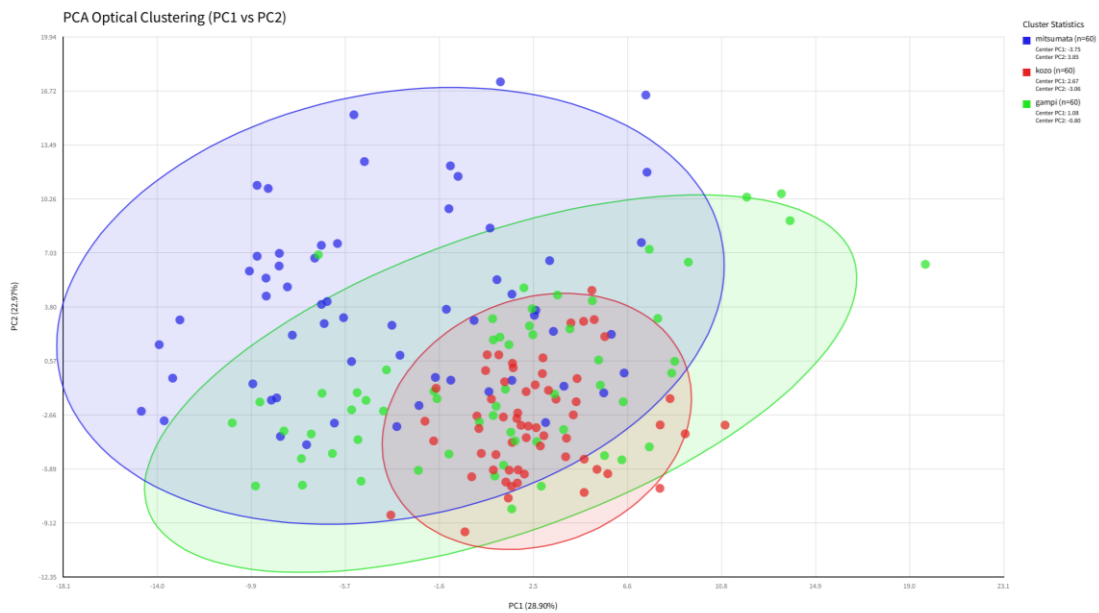
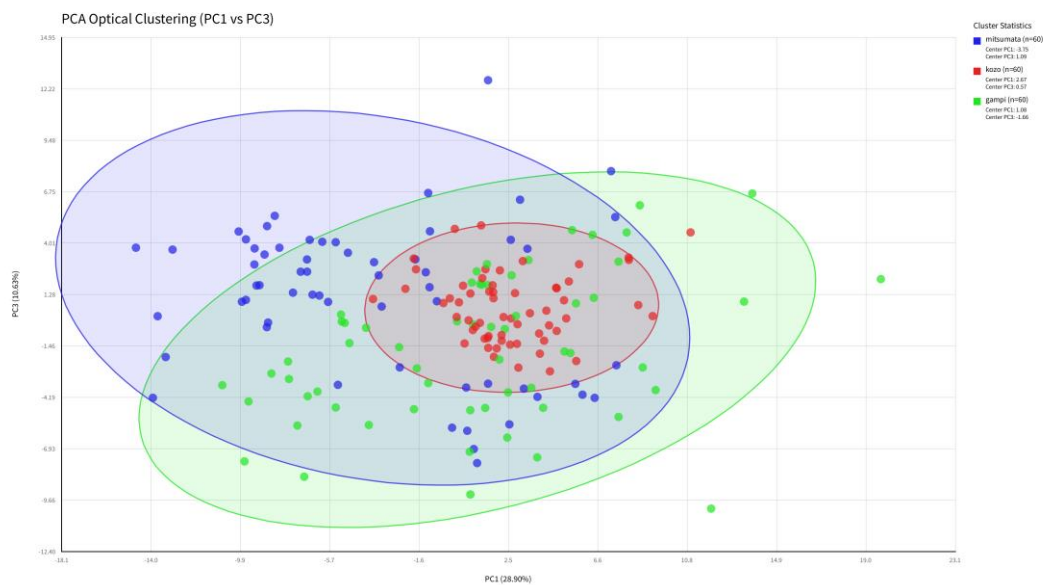


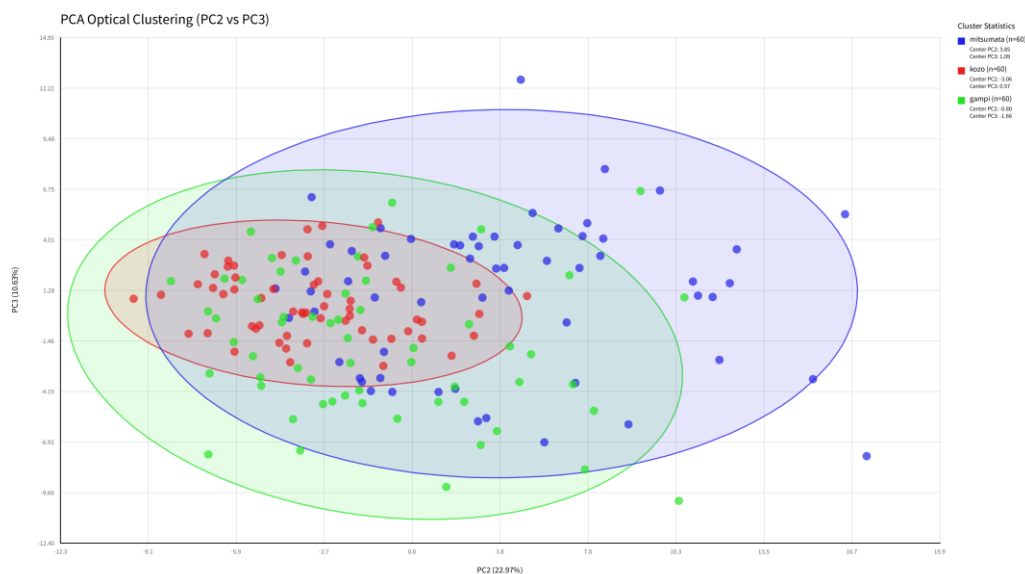
Figure 4. Coefficient of variation (CV) across the full spectral range for Kozo (red), Mitsumata (blue), and Gampi (green). Within the 450–750 nm window, CV curves remain low and stable for all three species, though with diverging trends: Gampi shows a gradual upward trajectory while Kozo and Mitsumata trend downward. Across the broader 400–1000 nm range, inter-species differences in CV behavior become more pronounced. Beyond 760 nm, CV rises sharply in all three fiber types, attributed to reduced detector sensitivity and weak overtone signals in this region; above 950 nm, partial overlap between species curves is observable at several wavelengths. These characteristics justify restricting subsequent chemometric modeling to the 400–1000 nm window.



(a)



(b)



(c)

Figure 5. PCA score plots with 95% confidence ellipses for Kozo (red), Mitsumata (blue), and Gampi (green): (a) PC1 vs. PC2, (b) PC1 vs. PC3, and (c) PC2 vs. PC3. Across all three projections, inter-class overlap is substantial, consistent with the SD envelope overlap observed in the normalized spectra. Intra-class dispersion differs markedly among species: Mitsumata shows the largest confidence ellipse and highest data scatter, indicating greater structural heterogeneity; Kozo is the most tightly clustered; Gampi falls intermediate. The compact distribution of Kozo may reflect more complete removal of non-cellulosic components—such as lignin and pectin—during pulping, resulting in a more chemically uniform fiber population.

3.4. Supervised Species Discrimination via LDA

LDA was applied to the standardized first-derivative spectra to address the inter-class overlap observed in PCA. Where PCA distributes variance globally without reference to class labels, LDA maximizes the ratio of between-class to within-class variance, directing the projection toward features that separate the three fiber types.

In the resulting discriminant space, Kozo, Mitsumata, and Gampi occupy distinct, non-overlapping regions (Figure 6). LD1 (73.16% of discriminant variance) provides the primary separation: Kozo is positioned at a centroid of LD1 = 5.90, well separated from both Mitsumata and Gampi, which cluster on the negative side of the axis. LD2 (26.84%) resolves the remaining overlap between Mitsumata (LD2 centroid: 3.03) and Gampi (LD2 centroid: -3.16), with intra-class clusters remaining compact throughout.

The separation achieved by LDA indicates that species-specific biochemical differences—attributed to residual lignin and hemicellulose retained after pulping—produce measurable spectral contrast even where raw reflectance profiles appear similar. Micro-HSI combined with first-derivative preprocessing and supervised LDA is therefore sufficient for species-level discrimination among these three bast fiber types under controlled reference conditions.

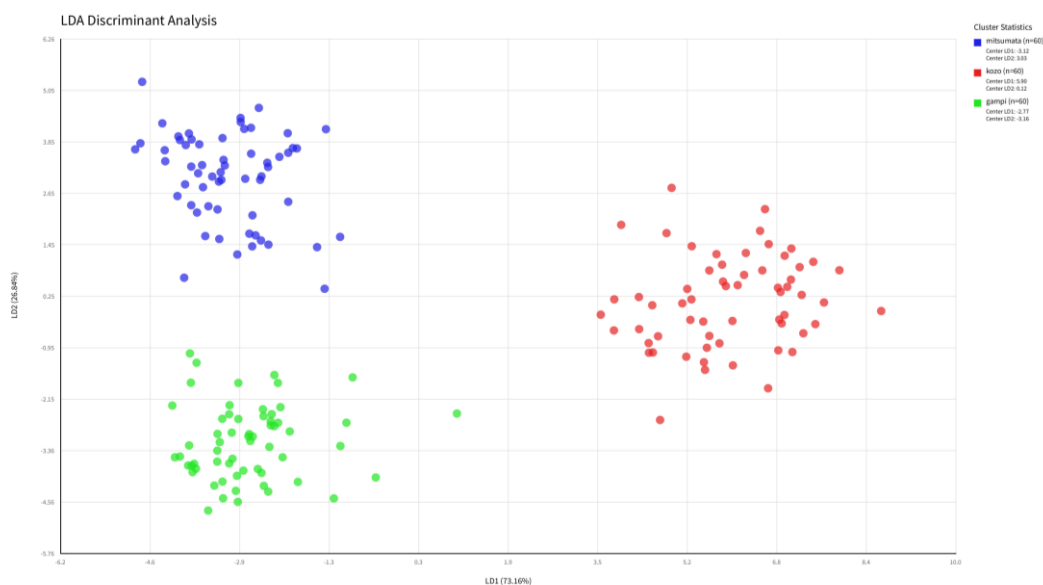


Figure 6. LDA score plot for Kozo (red), Mitsumata (blue), and Gampi (green), each $n = 60$. All three fiber types form fully separated clusters in the discriminant space with no overlapping data points. LD1 (73.16% of discriminant variance) provides the primary separation: Kozo is positioned at a centroid of $LD1 = 5.90$, well separated from Mitsumata ($LD1 = -3.12$) and Gampi ($LD1 = -2.77$), both of which cluster on the negative axis. This pronounced separation along LD1 is consistent with the tighter intra-class dispersion of Kozo observed in both the PCA projections and SD envelopes. Mitsumata and Gampi share similar LD1 positions, indicating overlap in their primary spectral characteristics; LD2 (26.84%) resolves the two vertically, with Mitsumata centered at $LD2 = 3.03$ and Gampi at $LD2 = -3.16$.

3.5. Interpretation of PCA Loading Features

Although LDA provides the final classification boundary, examining the PCA loading vectors helps reveal the spectral features responsible for the variance structure of the dataset.

PCA loading vectors (Figure 7) were examined to link statistical variance to specific spectral features.

PC1 (peak loading ≈ 0.153 at 585 nm) is dominated by broad visible-range reflectance, with elevated loadings spanning 500–700 nm. This distribution corresponds to the differing visible-range slopes among the three fiber types—particularly the steady linear rise characteristic of Kozo—rather than discrete chemical absorption bands. PC1 therefore reflects aggregate light-scattering behavior associated with fiber microstructure rather than specific molecular markers.

PC2 and PC3 capture the chemically diagnostic signals. PC2 loading peaks sharply at approximately 400 nm (≈ 0.156), consistent with $\pi \rightarrow \pi^*$ electronic transitions in lignin aromatic rings; this feature tracks residual lignin retained after pulping. PC3 is concentrated in the NIR, with the highest absolute loading in the dataset at 835 nm (≈ 0.226), a wavelength associated with vibrational modes in hydrogen-bonded polysaccharide networks and indicative of variation in hemicellulose or cellulose organization.

Taken together, the three components separate the spectral variance into distinct physical-chemical contributions: bulk light-scattering behavior (PC1), lignin chromophore absorption (PC2), and NIR polysaccharide vibrational response (PC3). This decomposition provides a mechanistic basis for the species-level discrimination achieved in the subsequent LDA step.

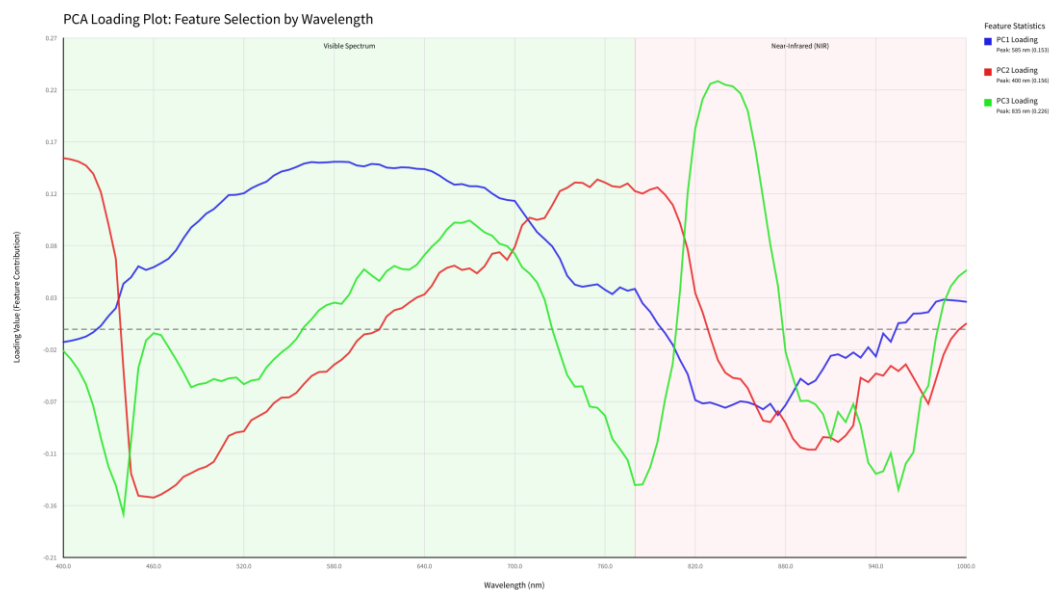


Figure 7. PCA loading vectors for the first three principal components—PC1 (blue), PC2 (red), and PC3 (green)—across the 400–1000 nm spectral range. Background shading distinguishes the visible region (light green) from the near-infrared region (light pink). PC1 loadings are concentrated in the visible range, peaking at 585 nm. PC2 shows a sharp maximum at 400 nm, adjacent to the UV boundary. PC3 exhibits the most pronounced feature in the dataset: a sharp positive loading peak between 820 and 850 nm (maximum at 835 nm), with markedly greater amplitude variation in this region compared to PC1 and PC2. Key loading maxima are annotated for each component.

4. Discussion

Material compatibility requirements in paper restoration create a practical need for fiber identification methods that do not involve physical sampling. Macroscopic hyperspectral approaches and deep-learning screening tools address part of this problem but remain sensitive to surface heterogeneity and aging-related optical noise. Shifting acquisition to the filament scale with Micro-HSI removes these confounding contributions; spectra extracted from individual fiber regions reflect biopolymer composition rather than surface condition.

At the microscopic scale, the shared cellulose backbone of Kozo, Mitsumata, and Gampi dominates the variance structure—an effect that becomes more pronounced when signals are integrated across the full paper matrix, as in document-wide hyperspectral mapping [7]. LDA reorients the projection toward between-class separation rather than total variance, and the resulting cluster boundaries align with established spectroscopic assignments: the 400–450 nm region corresponds to lignin aromatic transitions [11], and the 835 nm feature to polysaccharide network vibrations [10].

Among the three fibers, Mitsumata produced the most distinctive visible-range response—a concave reflectance profile and localized peak near 425 nm not present in Kozo or Gampi. This feature, linked to lignin chromophore distribution, was the most visually identifiable marker during spectral inspection and contributed to Mitsumata's higher mean CV (27.53%) relative to Kozo (19.74%), consistent with greater structural heterogeneity at the fiber level.

The reference dataset was deliberately restricted to pure, unprocessed specimens to isolate intrinsic fiber signatures. Historical documents introduce additional variables—mixed pulps, degradation products, surface coatings—that were excluded here by design. Extending the method to real artifacts will require spectral unmixing for multi-component systems and expanded databases incorporating naturally aged samples. Integration with white-light confocal microscopy may also improve spatial registration between chemical and structural data. Within the Paper Road project, the microscopic references established here are intended to provide fiber-level calibration for the

macroscopic Stage 2 models, where prediction bias from uncharacterized fiber composition is a known source of error.

5. Conclusions

Micro-HSI combined with dynamic normalization and S–G first-derivative filtering successfully extracted fiber-level spectral signatures from traditional Japanese bast paper specimens without physical sampling. Dynamic normalization suppressed scattering-induced baseline variation, and derivative transformation removed residual drift while sharpening chemical absorption features.

PCA of the processed spectra revealed substantial inter-class overlap among Kozo, Mitsumata, and Gampi, attributable to their shared cellulose variance structure. Supervised LDA resolved the three fiber types into non-overlapping clusters; discrimination was supported by lignin-related absorptions at 400–450 nm and polysaccharide vibrational features near 835 nm, consistent with established spectroscopic assignments.

These results provide a controlled microscopic spectral reference for pure Japanese bast fibers under the Paper Road research framework. Subsequent work will extend the database to aged and mixed-pulp samples and develop quantitative unmixing approaches for multi-component historical documents.

Author Contributions: Conceptualization, Y.Z.; methodology, Y.Z.; software, Y.Z.; formal analysis, Y.Z.; investigation, Y.Z., Y.O., A.I. and K.A.; writing—original draft preparation, Y.Z.; writing—review and editing, Y.Z., K.S. and N.K.; supervision, K.S. and N.K.; project administration, K.S. and N.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Grant-in-Aid for Scientific Research, Japan Society for the Promotion of Science (JSPS), grant number 22H00003 (project: "Elucidation of the Paper Road by data science – Based on Quantitative, Qualitative research and AI Multidimensional analysis"). The APC was waived by the journal.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: During the preparation of this manuscript, the author(s) used Claude (Anthropic) for the purposes of language editing and text revision. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

Micro-HSI	Micro-Hyperspectral Imaging
VNIR	Visible and Near-Infrared
NIR	Near-Infrared
UV	Ultraviolet
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
PC	Principal Component
LD	Linear Discriminant
CV	Coefficient of Variation
SD	Standard Deviation
ROI	Region of Interest

S–G	Savitzky–Golay
NDT	Non-Destructive Testing
JIS	Japanese Industrial Standard

References

1. Area, M.C.; Cheradame, H. Paper aging and degradation: Recent findings and research methods. *BioResources* **2011**, *6*, 5307–5337. [CrossRef]
2. Avataneo, C.; Sablier, M. New criteria for the characterization of traditional East Asian papers. *Environmental science and pollution research international* **2017**, *24*, 2166–2181. [CrossRef]
3. Japanese Industrial Standards. *Paper, board and pulps—Fiber analysis*; JIS P 8120:1998; Japanese Standards Association: Tokyo, Japan, 1998.
4. H. Lukesova, B. Holst. Identifying plant fibers in cultural heritage with optical and electron microscopy: how to present results and avoid pitfalls. *Heritage Science* **2024**, *12*, 12. [CrossRef]
5. Garside, P.; Wyeth, P. Identification of Cellulosic Fibres by FTIR Spectroscopy - Thread and Single Fibre Analysis by Attenuated Total Reflectance. *Studies in Conservation*. **2003**, *48*, 269–275. [CrossRef]
6. Okuyama, M.; Sato, M.; Akada, M. The Study on Excavated Bast Fibres Using Synchrotron Polarized FT-IR Micro-Spectroscopy. *Sen'i Gakkaishi* **2012**, *68*, 55–58. DOI:10.2115/fiber.68.55. [CrossRef]
7. Wu, Y.; Wang, B.; Chen, J.; Huang, X.; Xu, J.; Wei, W.; Chen, K. Non-destructive prediction and pixel-level visualization of polysaccharide-based properties in ancient paper using SWNIR hyperspectral imaging and machine learning. *Carbohydrate Polymers* **2025**, *352*, 123198. DOI: 10.1016/j.carbpol.2024.123198 [CrossRef]
8. Yagi, C.; Yoshimura, N.; Takayanagi, M.; Kikuchi, R.; Yasunaga, T.; Hayakawa, N. Discrimination of traditional plant fibers used in Japanese cultural artifacts by infrared spectroscopy. *Vibrational Spectroscopy*. **2022**, *123*, 103466. [CrossRef]
9. Kamiya, N.; Ashino, K.; Sakai, Y.; Zhou, Y.; Ohyanagi, Y.; Shibasaki, K. Non-Destructive Estimation of Paper Fiber Using Macro Images: A Comparative Evaluation of Network Architectures and Patch Sizes for Patch-Based Classification. *NDT* **2024**, *2*, 487–503. DOI: 10.3390/ndt2040028 [CrossRef]
10. Schwanninger, M.; Rodrigues, J.C.; Fackler, K. A review of band assignments in near infrared spectra of wood and wood components. *J. Near Infrared Spectrosc.* **2011**, *19*, 287–308. DOI: 10.1255/jnirs.955 [CrossRef]
11. Sadeghifar, H.; Ragauskas, A. Lignin as a UV light blocker—A review. *Polymers* **2020**, *12*, 1134. [CrossRef]
12. Rinnan, A.; van den Berg, F.; Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*. **2009**, *28*, 1201–1222. DOI: 10.1016/j.trac.2009.07.007 [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.