

---

Article

# Unsupervised Spectral Analysis of Bio-Dyed Textile Samples

Zong-Yue Li <sup>1,\*</sup>, Joni Hyttinen <sup>1</sup>, Riikka Räisänen <sup>2</sup>, Xiao-Zhi Gao <sup>1</sup>, and Markku Hauta-Kasari <sup>1</sup>

<sup>1</sup> School of Computing, Faculty of Science, Forestry and Technology, University of Eastern Finland, Joensuu, 80101, Kuopio, 70211, Finland

<sup>2</sup> Department of Education, Faculty of Educational Sciences, University of Helsinki, Helsinki, 00014, Finland

\* Correspondence: [winstonli711@gmail.com](mailto:winstonli711@gmail.com)

Received: 12 February 2023

Accepted: 23 March 2023

Published: 23 June 2023

**Abstract:** Natural compounds such as biological colorants (biocolorants) have long been employed as crucial ingredients for dyeing textile in the textile industry. As one part of the BioColour Consortium project, our goal is to take advantage of machine learning (in cluster analysis) to discover possible clusters of bio-dyed textile in the absence of ground truth labels or other knowledge of expert domains. Specifically, we use unsupervised learning methods of agglomerative clustering, fuzzy c-means, ordering points to identify the clustering structure (OPTICS) and self-organizing maps (SOMs), resulting in an investigation that combines data visualization and cluster analysis. In summary, we apply some selected data mining methods to 1) discover hidden clusters emerging among products that are colored with biocolorant (specifically bio-dyed textile samples), and 2) show the potentials of clustering techniques in the case study.

**Keywords:** Data mining; unsupervised learning; cluster analysis; self-organizing maps

---

## 1. Introduction

The natural dye extensively used in the modern industry has been brought back to the fore as the potential alternative for the synthetic dye. More intensive research has been undertaken on natural dyes because they are more eco-friendly and less allergenic and toxic (to humans) than synthetic dyes in chemical properties [1]. Undoubtedly, digital data is helpful to support the growth, health and standardization of the biocolorant industry. Moreover, it is a critical issue to refine and organize these data efficiently for characterization and quality analysis. Artificial intelligence techniques have been applied to these fields. For example, as a famous tool in machine learning, the support vector machine (SVM) has been applied to classify dyes in colored fibers [2].

Our paper aims to organize the database of BioColour (<https://biocolour.fi/en/about-the-project/>) where unsupervised learning methods are used for feature extraction and data clustering. Although the unsupervised learning methods have been applied in many areas, they have not yet been utilized in the cluster analysis of bio-dyed samples. Our open-ended and exploratory work is able to 1) support authentication and quality analysis of biocolorant sources along with their colored products; 2) serve companies and dye conservators refining and utilizing biocolorants; and 3) prompt large scale utilization of biocolorants in the future.

## 2. Data Preprocessing

For this paper, the input data consists of bio-dyed textile samples that are provided by the BioColour project. For each sample, its reflectance spectra are captured by the spectrophotometer under the CIE standard illuminant D65 light source, and measured in the range from ultraviolet (360 nm) to near-infrared (740 nm) with its interval being 10 nm. In data preprocessing, the data reduction method is applied to reduce computational costs and memory resources. In this section, we first describe the bio-dyed samples and then use two alternative methods to conduct data reduction, including the wavelength-color-space-based converting method (which is specific to this type of data) and the principal component analysis (PCA) method.

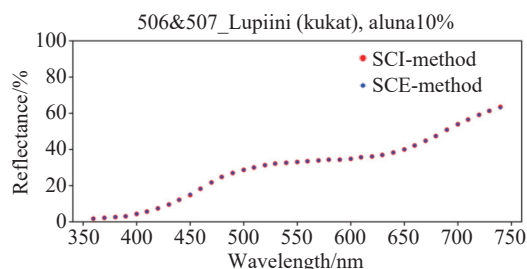
### 2.1. Database Specifics

In our experiments, the database consists of 611 bio-dyed textile samples in total. A small example subset of the

spectrum data and its graphs are shown in Table 1 and Figure 1. In Table 1, each row represents the spectral reflectance of a named sample in the wavelength range of 360-740 nm. Furthermore, two measurement methods are applied to the sample, i.e. the specular component included (SCI) and specular component excluded (SCE) methods.

**Table 1** Reflectance spectra of bio-dyed textile samples based on the SCI and SCE methods in the wavelength range of 360-740 nm

Reflectance (%)under different wavelength (nm)	360	370-730	740
506_Lupiini(kukat),aluna10%_SCI	1.72	...	63.26
507_Lupiini(kukat),aluna10%_SCE	1.74	...	63.15



**Figure 1.** Reflectance spectra of bio-dyed textile samples based on the SCI method (red) and SCE method (blue) in the wavelength range of 360-740 nm.

## 2.2. Data Conversion Between Different Color Spaces

By converting the data into different color spaces, we can perform data reduction based on data characteristics. Since the bio-dyed samples are based on the given spectral distribution, we can compute CIE XYZ values from data's reflectance spectrum. The relationship between CIE XYZ and spectrum data in the reflectance case can be summarized in Equations (1)–(4).

$$X = \frac{1}{N} \sum_i \bar{x}_i(\lambda) S_i(\lambda) I_i(\lambda) \Delta\lambda \tag{1}$$

$$Y = \frac{1}{N} \sum_i \bar{y}_i(\lambda) S_i(\lambda) I_i(\lambda) \Delta\lambda \tag{2}$$

$$Z = \frac{1}{N} \sum_i \bar{z}_i(\lambda) S_i(\lambda) I_i(\lambda) \Delta\lambda \tag{3}$$

$$N = \sum_i \bar{y}_i(\lambda) I_i(\lambda) \Delta\lambda \tag{4}$$

where  $S(\lambda)$  means the spectral reflectance of samples and  $I(\lambda)$  means the spectral power distribution of a reference illuminant locating in a visible spectrum  $\lambda$ .  $\bar{x}$ ,  $\bar{y}$  and  $\bar{z}$  are the CIE standard observer functions. Since the XYZ color spaces are perceptually non-uniform, we need to convert these spaces to other color spaces for cluster analysis.

### 2.2.1. Conversion from XYZ to LAB

In order to guarantee the reliability and validity of experimental results, we apply CIE LAB as a data reduction method. CIE LAB, as a CIE standard observer, has a gamut that is larger than human vision, thereby providing a color space that is more perceptually linear than other models. Specifically, perceptually linear refers to how similar the change of the color value is to the change of visual importance. In the CIE LAB color space, the color is expressed as three values:  $L^*$  is the perceptual brightness, and  $a^*$  and  $b^*$  are the chrominance components. Formally, CIE LAB is calculated from the XYZ coordinates as shown in Equations (5)–(7).

$$L^* = \begin{cases} 116 \sqrt[3]{\frac{Y}{Y_n}} - 16 \frac{Y}{Y_n} > 0.008856 \\ 903.3 \left(\frac{Y}{Y_n}\right) \frac{Y}{Y_n} < 0.008856 \end{cases} \tag{5}$$

$$a^* = 500 \left( \sqrt[3]{\frac{X}{X_n}} - \sqrt[3]{\frac{Y}{Y_n}} \right) \tag{6}$$

$$b^* = 200 \left( \sqrt[3]{\frac{Y}{Y_n}} - \sqrt[3]{\frac{Z}{Z_n}} \right) \tag{7}$$

In this work, the CIE standard D65 light source is used as the neutral point with values  $X_n = 95.04$ ,  $Y_n = 100.00$  and  $Z_n = 108.88$ . Thus, we can get a more perceivable three-dimension model than CIE XYZ so as to analyze samples' clustering characteristics.

### 2.2.2. Conversion from XYZ to Standard RGB

In general, most data visualization systems rely on the RGB color model instead of the CIE XYZ or CIE LAB models. We use the standard RGB color model as a visualization tool rather than an input in cluster analysis. In CIE standards, the linear normalized transformation from the standard RGB space to the XYZ space is given in Equation (8) as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{8}$$

The conversion of the XYZ space to the standard RGB color space is given as follows:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.2406 & -1.5372 & -0.4986 \\ -0.9689 & 1.8758 & 0.0415 \\ 0.0557 & -0.2040 & 1.0570 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{9}$$

### 2.3. Principal Component Analysis

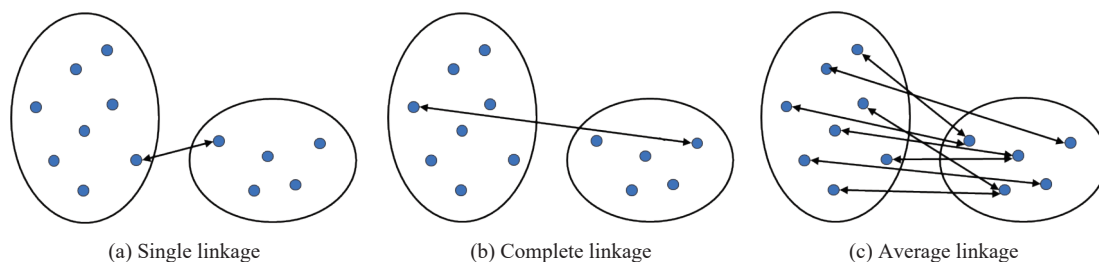
The PCA is one of the most commonly used dimensionality reduction methods [3,4]. In the PCA method, the covariance matrix is calculated for eigenvalues and eigenvectors. When analyzing bio-dyed samples with dozens of spectrum components, the projection values can describe the original data with reasonable accuracy.

## 3. Conventional Methods for Clustering

This section discusses three different clustering methods applied to the bio-dyed textile data. Based on the grouping rationale of data points, traditional clustering algorithms can be broadly divided into three main categories: hierarchical, partitional and density-based clustering [5]. After introducing more details about these categories, we introduce two evaluation criteria, i.e. the Davies-Bouldin index (DBI) [6] and Silhouette coefficient (SIL) [7], which are used as the basis for evaluating the following clustering results.

### 3.1. Hierarchical Clustering: Agglomerative Method

Hierarchical clustering can be divided into two major categories, i.e. the agglomerative and divisive methods. In most methods of hierarchical clustering, splitting or merging objects are decided and measured by the linkage metric which is adopted to calculate the pairwise distances of objects. Three main linkage criteria are shown in Figure 2, namely, the single linkage, complete linkage and average linkage criteria.



**Figure 2.** Different linkage criteria when calculating the distance of two clusters. **(a)** The minimum distance determined by nearest samples. **(b)** The maximum distance determined by furthest samples. **(c)** The average distance determined by all samples.

In agglomerative clustering, each sample is initially regarded as a single-element cluster, and any two clusters (that are the most similar) are merged into a new bigger cluster. This procedure is iterated until the required number of clusters is achieved or all points become members of one single big cluster.

### 3.2. Partitional Clustering: Fuzzy C-Means

In contrast to hierarchical clustering, partitional clustering algorithms divide the dataset into a pre-specified number of clusters and progressively refine the clusters to improve their quality. In experiments, we use the extension of the K-means [8] and fuzzy c-means (FCM) [9] algorithms, and this is useful in dealing with the situation of ambiguous clusters' boundaries. In FCM, each point has a degree of belonging to a distinct cluster and thus can be assigned to several clusters with different degrees.

### 3.3. Density-Based Clustering: OPTICS

In real-world databases, clusters may have non-spherical or linear, irregular shapes. Examples of such situations are shown in Figure 3. In these cases, hierarchical and partitional clustering methods will merge or break up actual clusters, which leads to an inaccurate result. In order to address this problem, density-based clustering is introduced to find these specially shaped clusters. In experiments, we mainly use ordering points to identify the clustering structure (OPTICS) for cluster analysis.

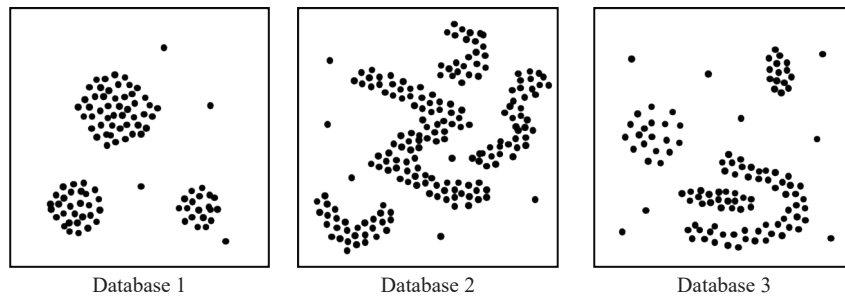


Figure 3. Arbitrarily shaped clusters in different databases.

The OPTICS is the extension of the DBSCAN [10], which was proposed by Ankerst et al. in 1999 [11]. The basic idea of the OPTICS is similar to that of the DBSCAN, but the OPTICS can detect clusters in data of varying densities, thus overcoming the difficulty of using one set of global density parameters in the DBSCAN. Take the dataset shown in Figure 4 as an example. Because A, B, C<sub>1</sub>, C<sub>2</sub> and C<sub>3</sub> have different densities and radii, the DBSCAN algorithm can characterize either A, B and C or C<sub>1</sub>, C<sub>2</sub> and C<sub>3</sub>, but cannot characterize all of them simultaneously.

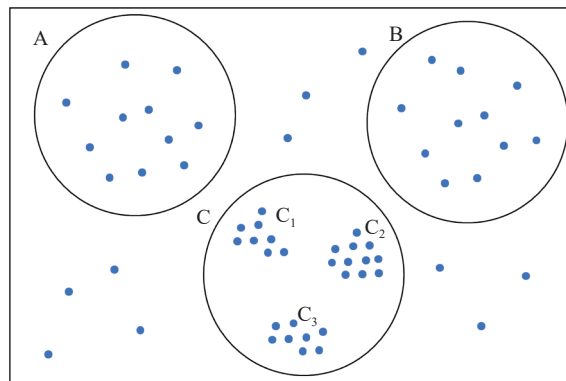


Figure 4. Detecting clusters with respect to different parameters.

### 3.4. Evaluation Criteria for Conventional Clustering Methods

Compared with supervised learning, data clustering (as an unsupervised learning task) does not typically come with a ground truth to evaluate the results. In this study, we apply two different approaches to measure the cohesion of intra-clusters and the separation of inter-clusters.

The Davies-Bouldin index (DBI) was firstly proposed by Davies and Bouldin [6]. This index is defined as a distance ratio between the intra-cluster and inter-clusters as follows:

$$DBI = \frac{1}{N} \sum_{i=1}^N \max_{j \neq i} \left( \frac{S_{C_i} + S_{C_j}}{M_{C_i, C_j}} \right) \quad (10)$$

where N is the number of clusters,  $S_{C_i}, S_{C_j}$  are the average distances between samples in clusters  $C_i$  and  $C_j$ , and  $M_{C_i, C_j}$  is the distance between the centroids of clusters  $C_i$  and  $C_j$ .

$$S_c = \frac{2}{|C| \cdot (|C| - 1)} \sum_{1 \leq a < b \leq |C|} dist(\mathbf{x}_a, \mathbf{x}_b) \tag{11}$$

$$M_{C_i, C_j} = dist\left(\frac{1}{|C_i|} \sum_{1 \leq a \leq |C_i|} \mathbf{x}_a, \frac{1}{|C_j|} \sum_{1 \leq b \leq |C_j|} \mathbf{x}_b\right) \tag{12}$$

where  $\mathbf{x}_a$  and  $\mathbf{x}_b$  are samples' vectors in their clusters. The smaller the value of DBI is, the more separated and compact the clusters are, which indicates better clustering accuracy.

In addition, we also use the mean Silhouette coefficient (SIL) of all samples to validate the consistency of the results from another perspective. The SIL, first introduced by Rousseeuw [7], is a measure of the similarity of samples within a cluster. The SIL of sample  $i$  is provided in Equation (13):

$$SIL(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{13}$$

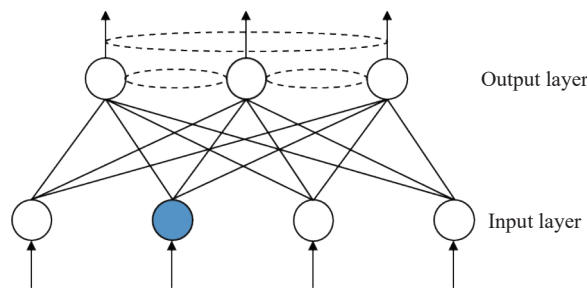
where  $a(i)$  is the average distance between sample  $i$  and all other samples in the same cluster, and  $b(i)$  is the average distance of sample  $i$  to all samples in its nearest cluster. The value of SIL ranges from -1 to 1. Values close to 1 mean that it is reasonable that sample  $i$  is in the current cluster and vice versa. While values near 0 indicate that sample  $i$  is on the boundary of two clusters. We can compute the mean SIL of all samples so as to evaluate the overall performance of a clustering algorithm.

#### 4. Self-Organizing Map for Clustering

Self-organizing maps (SOMs) were invented by Kohonen [12,13], which use competitive learning to produce the non-linear, low-dimensional projection of high-dimensional data, while retaining the topological structure of the input data. In this section, we explore two-dimensional SOM (2D SOM) and growing hierarchical SOM (GHSOM) [14] as unsupervised learning methods for clustering bio-dyed samples.

##### 4.1. Principles of Self-Organization in SOMs

SOM networks typically consist of two layers, i.e. the input layer and the output layer. Neurons and connections of two layers in a traditional SOM network are illustrated in Figure 5.

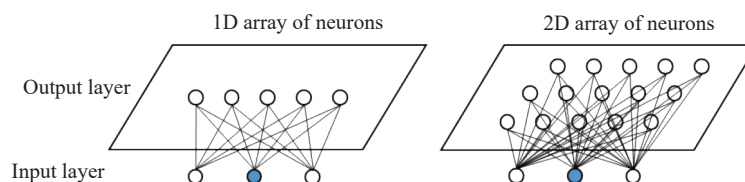


**Figure 5.** Neurons and connections in a traditional SOM structure, consisting of both input layer and output layer.

In general, self-organization aims to train a network that allows to approximate the data distribution by a smoothing process. To begin with, the neurons of output layer have a location on the plane of the network, and are connected to the ones located around them. In each iteration, every input vector selects one neuron in the network that matches best with itself. The winning neuron and its neighbors will move towards the sample vector for better matching. After several iterations, the network approaches the data distribution.

##### 4.2. Two-Dimensional SOM

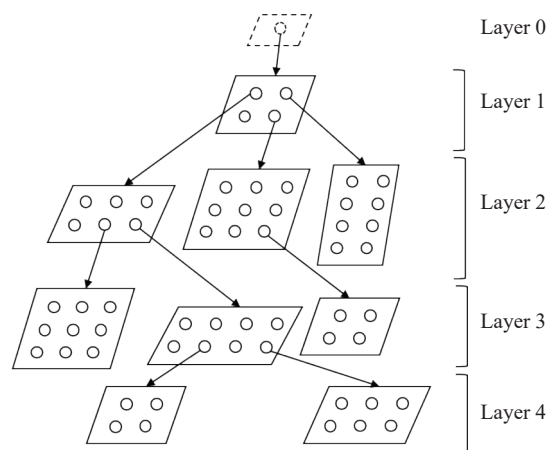
In the output layer, neurons' topological structures are usually one or two-dimensional as illustrated in Figure 6. In 2D SOM, neurons are organized with regular grids which consist of rows and columns.



**Figure 6.** Comparison between 1D and 2D SOM network in topological structure. The left uses 1D (linear) structure, while the right is 2D structure.

### 4.3. Growing Hierarchical SOM

In 2D SOM, a fixed network concerning the number and arrangement of neurons has to be determined prior to training. Obviously, in the case of unknown data characteristics, it is difficult to obtain satisfactory results when pre-defining the size of networks with a static architecture. In order to resolve this limitation, we use a model derived from the GHSOM (growing hierarchical). The basic idea of GHSOM is to use an adaptive hierarchical structure model, in which each layer is composed of an independent SOM. In particular, the GHSOM starts from a top-layer SOM with the entire data collection and zooms down to different low layers which have finer granularity. An example of the GHSOM architecture is shown in Figure 7.



**Figure 7.** Structure of a four-layer GHSOM.

## 5. Performance Evaluation

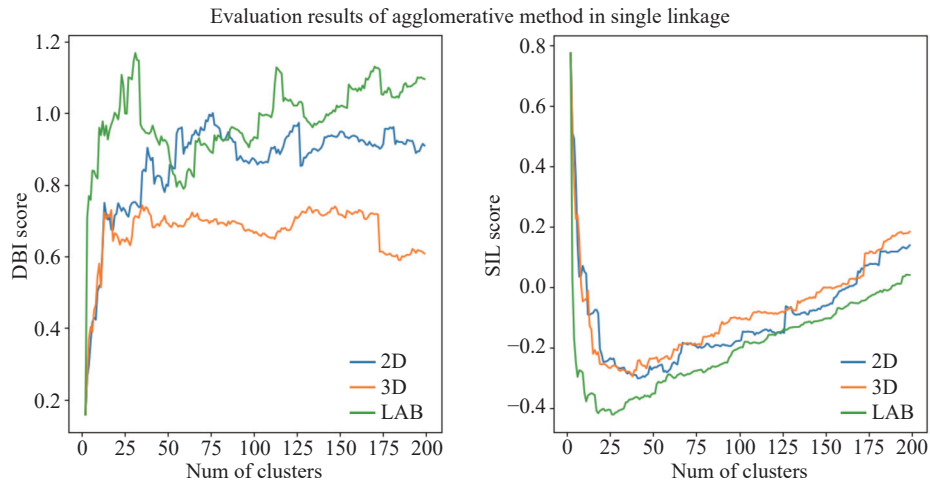
In this section, we carry out our experimental evaluation of different unsupervised learning algorithms for clustering bio-dyed textile samples. The source code of our Python scripts for running the experiments are publicly available (<https://github.com/4daJKong/Cluster-analysis-in-BioColour-project>).

### 5.1. Clustering Results with Conventional Methods

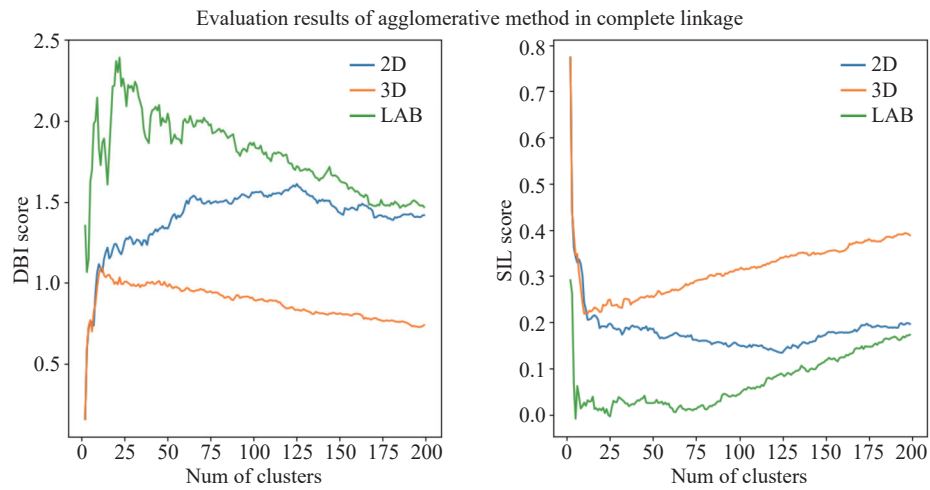
Our goal of this first experiment is to assess the performance of conventional clustering methods under different parameters. As a preprocessing step, we convert the reflectance spectrum of bio-dyed samples to the LAB color space. On the other hand, we also use PCA to project spectrum data into two-dimension and three-dimension data, respectively. The following clustering methods are applied for analysis: the agglomerative clustering, FCM and OPTICS, which are evaluated by the DBI and SIL criteria.

#### 5.1.1. Agglomerative Clustering

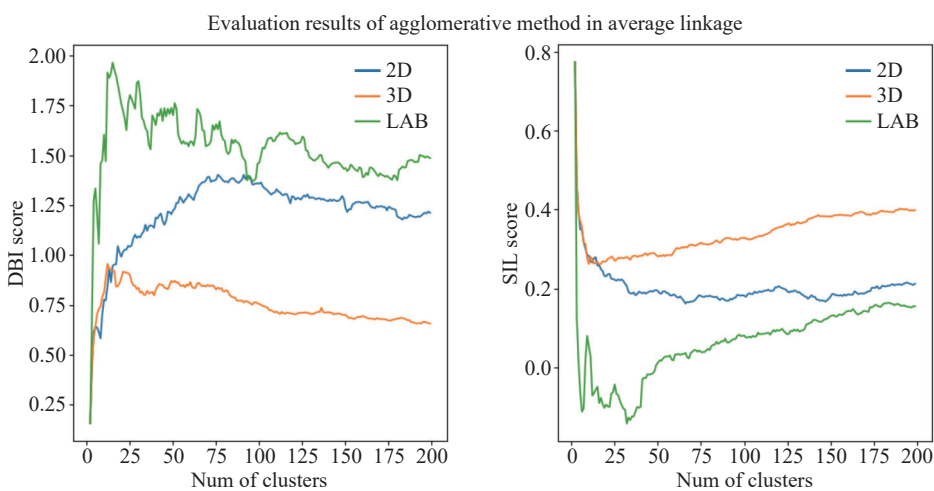
The evaluation results of the agglomerative method are illustrated in Figures 8 and 10, where the values of the DBI and SIL are given under different input functions, linkage functions and preset cluster numbers. Normally, when the value of the DBI is small or the value of the SIL is close to 1, good clustering results are achieved. In these figures, DBI increases and SIL decreases along with the increase of cluster numbers, which shows that the performance changes from good to bad at first. Then, when the preset cluster number is over 25, the curve of the DBI fluctuates and the curve of the SIL rises. Besides, for the data points in a 3D space after PCA, the best DBI scores are achieved for single linkage, and the best SIL scores are achieved for average linkage.



**Figure 8.** Evaluation results of agglomerative clustering in a single linkage function and different cluster numbers.



**Figure 9.** Evaluation results of agglomerative clustering in a complete linkage function and different cluster numbers.

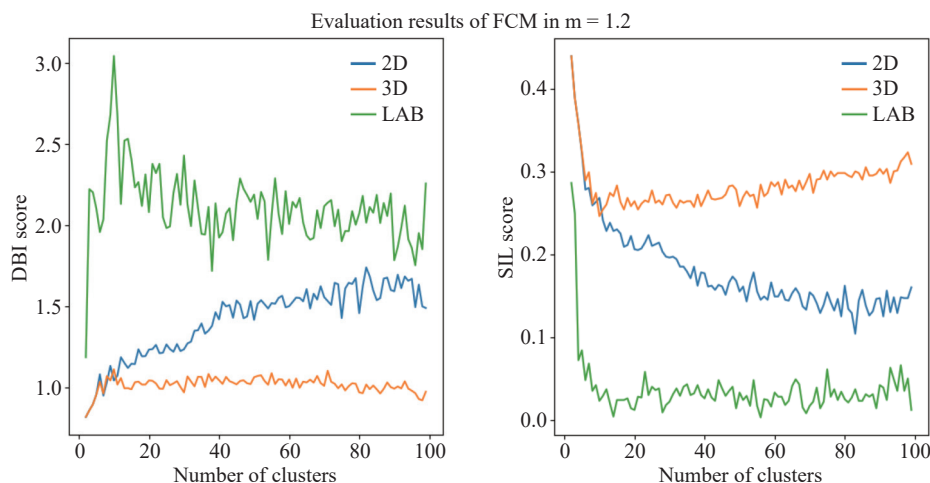


**Figure 10.** Evaluation results of agglomerative clustering in an average linkage function and different cluster numbers.

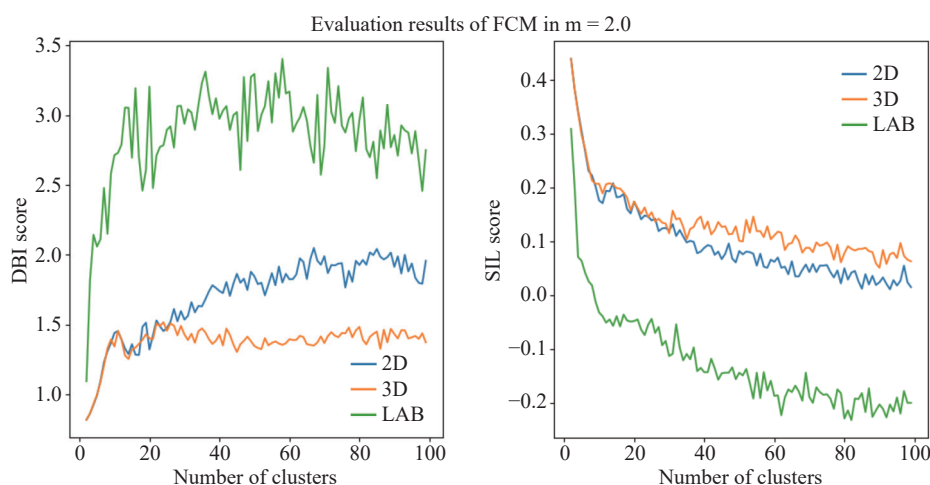
### 5.1.2. Fuzzy C-Means

The evaluation results obtained by FCM with various sets of parameters are presented in [Figures 11](#) and [12](#). We evaluate the scores of the DBI and SIL in two-dimensional, three-dimensional data spaces under different fuzziness parameters, i.e.  $m$ . As the number of clusters increases, the value of DBI increases in the 2D input space when  $m = 1.2$  and 2, which means that clustering results are bad. When input is in the LAB color space, the value of DBI

fluctuates, while the 3D input after PCA remains the same. On the other hand, the SIL curve in all situations decreases rapidly except for the case of  $m = 1.2$ , which shows that the intra-cluster similarity and inter-cluster dissimilarity decrease and the clustering performance gets bad.



**Figure 11.** Evaluation results of FCM in different cluster numbers and different fuzziness parameters  $m = 1.2$ .

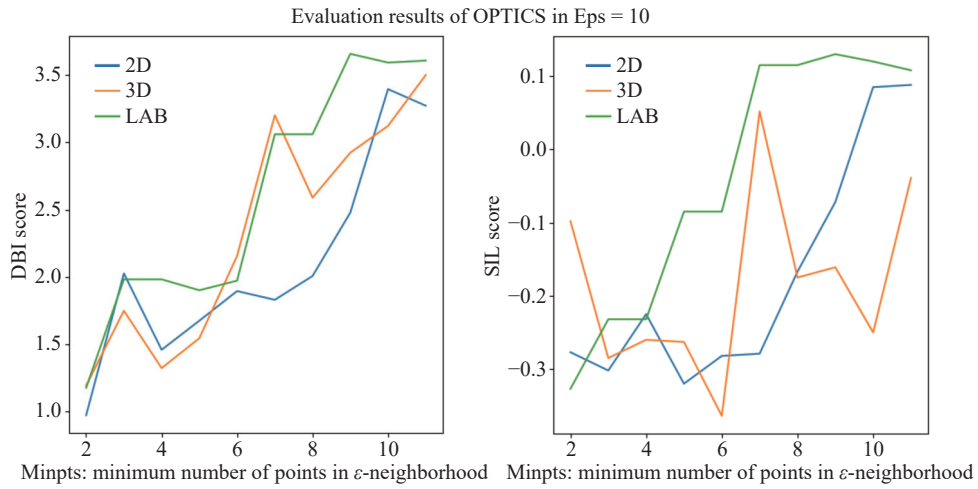


**Figure 12.** Evaluation results of FCM in different number of clusters, fuzziness parameter  $m = 2.0$ .

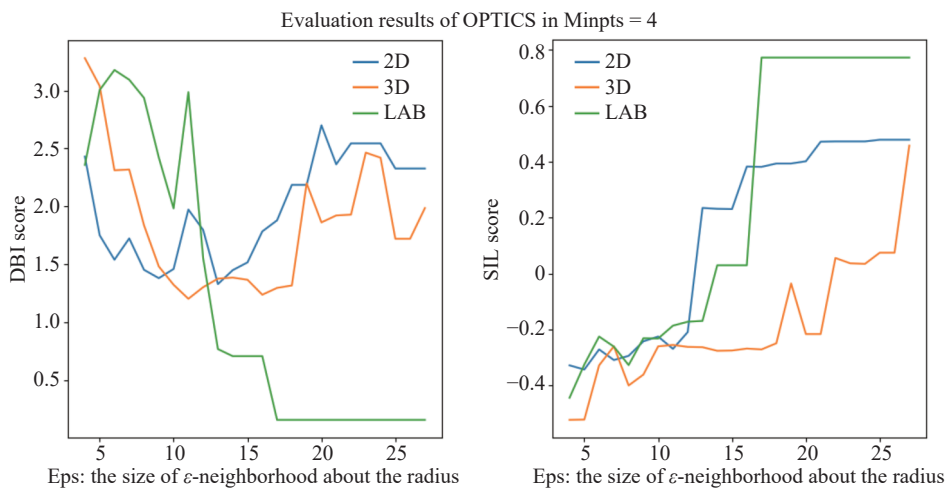
### 5.1.3. OPTICS

We mainly explore the influence of different distances of neighborhoods, the  $Eps$  ( $Minpts$ ), and the minimum number of points in  $Eps$ . We evaluate the DBI and SIL scores at values of  $Minpts$  2 to 11 when the  $Eps$  is a fixed value, see Figure 13. We also show the relationship between the evaluation criteria and  $Eps$  when  $Minpts = 4$ , see Figure 14. The two figures illustrate a great fluctuation in different parameters and evaluation results. This is related to the distribution of the bio-dyed samples and the density-based OPTICS algorithm.





**Figure 13.** Evaluation results of the OPTICS in different Minpts, Eps = 10.



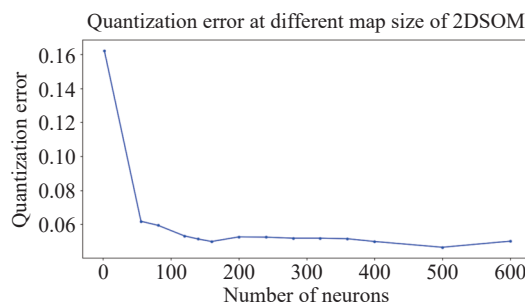
**Figure 14.** Evaluation results of the OPTICS in different Eps, Minpts = 4

### 5.2. Clustering Results in SOM

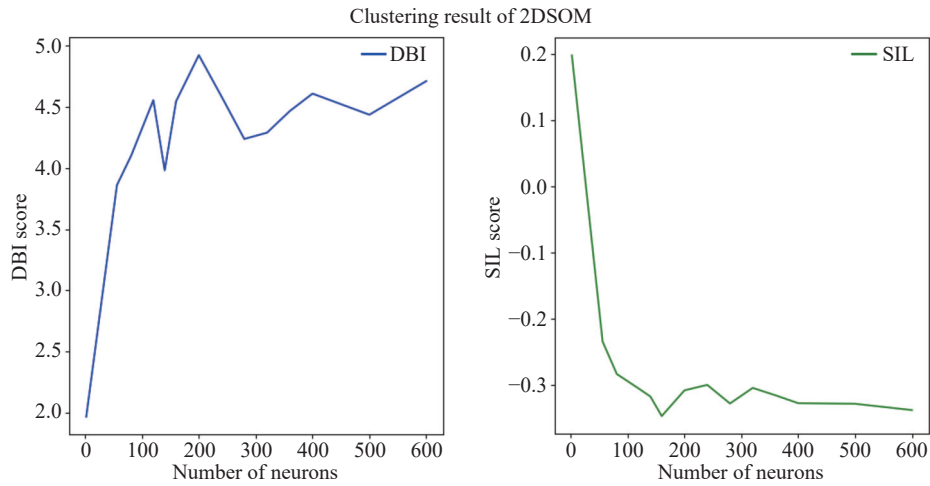
We experiment with two types of SOM, i.e. the 2D SOM and GHSOM, for analyzing spectrum data of bio-dyed textiles. Without dimension reduction, we use the full original reflectance vectors as inputs after normalization, test the performance of 2D SOMs with different map sizes, and evaluate the quality by the DBI measure, the SIL measures and the quantization error.

#### 5.2.1. Two-Dimensional SOM

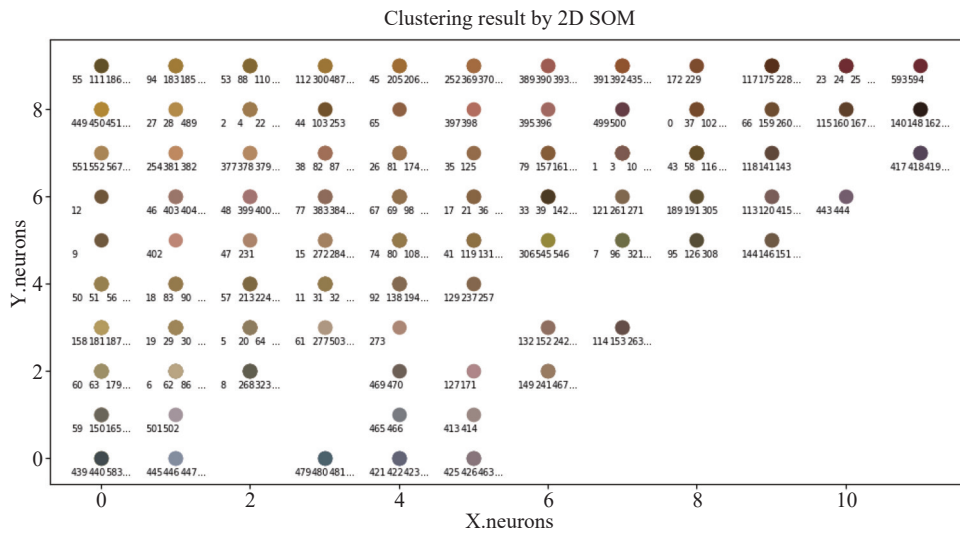
The values of the quantization error, DBI score and SIL score are obtained for different sizes of 2D SOMs in Figure 15 and Figure 16. We set the size of the map to vary from 2 to 600 neurons. Especially, the quantization error decreases along with the increase of the neuron number at first. However, when reaching a threshold of the map size, no obvious changes of quantization errors can be observed. In order to be able to display bio-dyed samples clearly, we mainly apply 2D SOM with a map size of  $14 \times 10$  including 140 neurons. The visualization result is shown in Figure 17, in which each neuron represents a cluster and the color is determined by one of the corresponding points of neurons.



**Figure 15.** Quantization errors at different map sizes of the trained two-dimensional SOM.



**Figure 16.** The curve of DBI and SIL scores at different map sizes of the trained two-dimensional SOM.



**Figure 17.** The distribution of data samples by 2D SOM in size of 140 neurons.

### 5.2.2. Growing Hierarchical SOM

In the experiment, different values of depth- and breadth-controlling parameters ( $\tau_u$  and  $\tau_m$ ) are used to test the performance of GHSOM. Based on different values of controlling parameters, the ratios of the numbers of sub-SOMs to the total number of neurons are presented in Table 2. Generally, the smaller  $\tau_u$  and  $\tau_m$  are, the larger the number of neurons is (in each layer of SOM) and the deeper the GHSOM is (in outputs).

**Table 2** The numbers of sub-SOMs  $N_s$  and the total number of neurons  $N_N$  as  $N_s/N_N$  with different controlling parameters in each layer

$\tau_m$	$\tau_u$	Layer1 SOM	Layer2 SOM	Layer3 SOM	Layer4 SOM	Total
0.2	0.02	1/6	6/16	6/24	8/42	21/86
0.3	0.03	1/4	4/16	5/20	2/8	12/48
0.4	0.04	1/4	4/16	4/16	1/4	10/40
0.5	0.05	1/4	4/16	3/12	—	8/32
0.6	0.06	1/4	3/12	4/16	—	8/32
0.7	0.07	1/4	3/12	2/8	—	6/24
0.8	0.07	1/4	2/8	2/8	—	5/20

### 5.3. Discussion

Based on unsupervised learning approaches, our experiments are conducted to research cluster analysis of bio-

dyed samples. Besides, based on different quantitative criteria, we evaluate the influence of different parameter values on clustering results. By using the existing implementation of five different unsupervised methods including the agglomerative clustering, fuzzy c-means, DBSCAN, 2D SOM and GHSOM, we obtain the performance of clustering results, and provide some visualization results to display the clusters of bio-dyed textile samples more effectively. In order to verify the performance of our methods, some criteria are used for evaluating the clustering experiments. Specifically, the scores of the DBI and SIL are mainly utilized in conventional clustering methods, and quantization errors are also computed in SOM methods. Figures 8-17 demonstrate the performance of these methods under different parameters. Overall, based on the scores of the DBI and SIL and when the input is 3D components (after conducting PCA in conventional methods), the intra-clusters' distance is small while the inter-clusters' distance is large, which indicates the good clustering performance. In addition, with the increase of the cluster number, the clustering performance tends to weaken greatly, especially when the cluster number exceeds 20.

Besides, we also use the quantization error as the criterion to assess the effect of the neuron number in SOM results, and try to find out adaptive methods to determine not only the neuron number, but also the hierarchical structure of the data in GHSOM. In next step, we will test more bio-dyed textile data in further research so as to verify the applicability of the proposed method.

## 6. Conclusions

In this paper, we have explored several unsupervised learning methods for the bio-dyed textile cluster analysis. Without the knowledge of taxonomy in chemistry and biology, machine learning approaches have been applied to analyze and cluster biocolorant samples. From the experiments, the clustering results have showed that unsupervised learning methods are indeed effective for a primary study in finding accurate clusters in the biocolorant database. Note that this study only focuses on three basic criteria for evaluating the performance of unsupervised learning methods, which is not comprehensive and adequate. In the future study, investigations will be performed with an increased number of samples in different types of the dyes database, and these samples include but are not limited to synthetic dyes samples. Furthermore, due to the partial availability of labeled data, we also plan to apply semi-supervised learning methods to biocolorants cluster analysis.

**Author Contributions:** **Zong-Yue Li:** conceptualization, data curation, formal analysis, investigation, methodology, writing—original draft and writing—review & editing; **Joni Hyttinen:** writing—review & editing, supervision; **Riikka Räisänen:** supervision, resources, project administration; **Xiao-Zhi Gao:** writing—review & editing, supervision; **Markku Hauta-Kasari:** writing—review & editing; supervision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was partially supported by BioColour (Bio-Based Dyes and Pigments for Colour Palette), one of three consortiums funded by the Strategic Research Council in IMPRES program for the years 2019-2025.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** All authors declare no conflicts of interest in this paper.

## References

- Lohtander, T.; Arola, S.; Laaksonen, P. Biomordanting willow bark dye on cellulosic materials. *Color. Technol.*, **2020**, *136*: 3–14.
- Rahaman, G.M.A.; Parkkinen, J.; Hauta-Kasari, M. A novel approach to using spectral imaging to classify dyes in colored fibers. *Sensors (Basel)* **2020**, *20*, 4379. doi:10.3390/s20164379
- Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, **1901**, *2*: 559–572.
- Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*, **1933**, *24*: 498–520.
- Arabie, P.; Hubert, L.; De Soete, G. *Clustering and Classification*; World Scientific: River Edge, 1996.
- Davies D. L.; Bouldin, D.W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *PAMI-1*, 224–227. doi:10.1109/TPAMI.1979.4766909
- Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, **1987**, *20*: 53–65.
- MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, 21 June-18 July 1965 and 27 December 1965-7 January 1966*; Statistical Laboratory of the University of California: Berkeley, 1967; pp. 281–297.
- Bezdek, J.C. A convergence theorem for the fuzzy ISODATA clustering algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **1980**, *PAMI-2*, 1–8. doi:10.1109/TPAMI.1980.4766964
- Ester, M.; Kriegel, H.P.; Sander, J.; et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, 2–4 August 1996*; AAAI Press: Portland, 1996; pp. 226–231.

11. Ankerst, M.; Breunig, M.M.; Kriegel, H.P.; *et al.* OPTICS: Ordering points to identify the clustering structure. *ACM SIGMOD Rec.*, **1999**, 28: 49–60.
12. Kohonen, T. The self-organizing map. *Proc. IEEE*, **1990**, 78: 1464–1480.
13. Kohonen, T.; Oja, E.; Simula, O.; *et al.* Engineering applications of the self-organizing map. *Proc. IEEE*, **1996**, 84: 1358–1384.
14. Merkl D.; Rauber, A. Uncovering the hierarchical structure of text archives by using an unsupervised neural network with adaptive architecture. In *Knowledge Discovery and Data Mining. Current Issues and New Applications*; Terano, T.; Liu, H.; Chen, A.L.P., Eds.; Springer: Berlin Heidelberg, 2000; pp. 384–395.

**Citation:** Li, Z.-Y.; Hyttinen, J; Räisänen, R.; *et al.* Unsupervised spectral analysis of bio-dyed textile samples. *International Journal of Network Dynamics and Intelligence*. 2023, 2(2), 100001. doi: [10.53941/ijndi.2023.100001](https://doi.org/10.53941/ijndi.2023.100001)

**Publisher’s Note:** Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2023 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license <https://creativecommons.org/licenses/by/4.0/>.