Research Article

A New Approach Based on Ensemble Clustering for the Fabric Color Batching Problem

Yusuf Kuvvetli[1] , Ebru Çalışkan[2], Onur Balcı[3], Esra Tabaş Asiltürk[4]

[1] Cukurova University, Department of Industrial Engineering Sarıcam,Adana, https://orcid.org/0000-0002-9817-1371, ykuvvetli@cu.edu.tr
[2] Kahramanmaras Sutcu Imam University, Department of Textile Engineering. 12 Subat/K.Maras , https://orcid.org/0000-0002-2133-8241, clskn.ebruu@gmail.com
[3] Kahramanmaras Sutcu Imam University, Department of Textile Engineering. 12 Subat/K.Maras , https://orcid.org/0000-0001-6885-7391, obalci@ksu.edu.tr
[4] Baykan Denim,Malatya, https://orcid.org/0000-0001-8363-3070, esratabas.asilturk@baykandenim.com

[*] Correspondence: ykuvvetli@cu.edu.tr; Tel.: +90 5322559145

Received: 04 June 2025
Revised: 27 September 2025
2nd Revised: 01 December 2025
Accepted: 03 December 2025
Published: 05 December 2025

Reference: Kuvvetli, Y., Çalışkan, E., Balcı, O., & Tabaş Asiltürk, E. (2025). A new approach based on ensemble clustering for the fabric color batching problem. *The European Journal of Research and Development, 5*(1), 310–321.

Abstract

The fashion industry is one of the industries most influenced by aesthetics and quality. This necessitates that products manufactured for this industry possess high quality and aesthetic appeal. Denim products are among the most frequently used in this industry for various purposes. This study proposes an ensemble clustering approach for visually sorting batches to reliably classify color consistency in denim fabrics. First, separate batches were obtained using three common methods (DBSCAN, hierarchical clustering, and K-Means) with 800×800 pixel RGB images of fabric samples for each order. Then, an ensemble rule based on the majority principle was designed to reduce inconsistencies between methods and balance random initialization and parameter sensitivity. Each sample was assigned to the final batch according to the majority preference among the batches given by the three algorithms. It is evaluated that the proposed approach by comparing it with reference batch assignments predefined by experts. The outputs of the individual algorithms and the ensemble *results are compared each other. The findings show that the ensemble rule produces more stable batches that are closer to expert decisions. While preserving the strengths of the individual methods, the ensemble rule reduces the impact of their weaknesses.*

Keywords: Color batching, Ensemble clustering, DBSCAN algorithm, K-means algorithm, Agglomerative clustering

1. Introduction

In the textile industry, the nature of production necessitates color consistency in fabrics within garment systems. Numerous factors influence the outcome of the dyeing process of fabric. A plenty of studies have previously reported the significance of the dyeing process, the impact of dye concentration, and the quality of the fibers on color variation. Additionally, research has emphasized the importance of the effects of dye concentration, temperature, duration, and auxiliary materials on the outcome of the dyeing process [1-2]. A specific focus on denim has revealed that the impact of washing processes on the physical and mechanical properties, durability, and comfort of denim fabrics can be significant [3]. In the context of denim fabric production, the examination of color effect, with particular attention given to color consistency, as it is a pivotal factor in determining the quality of the final product and the satisfaction of the customer. The utilization of artificial intelligence methodologies has been employed in numerous studies to forecast these parameters [4-6]. Minor variations in the color tones of fabrics from different production batches have the potential to induce discernible inconsistencies in product assembly processes. These inconsistencies can have adverse ramifications on both production efficiency and brand reliability. Consequently, the accurate batching (i.e., the sorting of fabrics into batches) following the manufacturing process constitutes a pivotal element in the quality control chain.

Conventionally, the batching process has been executed manually, based on visual assessments conducted by expert specialists. However, given that this method depends on human perception and is subject to factors such as environmental conditions, eye fatigue, and individual differences, it is susceptible to variability. This condition can result in fabrics from the same batch being allocated to different batches or in similar color tones being designated to disparate classes by different specialists. In this context, digital image processing and machine learning-based approaches offer a significant alternative by providing objectivity, repeatability, and scalability in color analysis. A review of existing studies reveals a variety of works involving image processing-based decisions in denim. One study examined the feasibility of employing image processing techniques to predict the quality of slub denim [7]. In the context of color matching, the spectrophotometric response

of fabrics composed of diverse colored fibers from varied materials was predicted through the utilization of Kubelka-Munk and artificial neural networks [8]. Another study endeavored to predict physical properties from fabric images [9]. A review of the extant literature reveals the employment of a variety of methodologies in the context of color matching studies. Among these approaches, artificial neural networks and linear regression [10], ant colony optimization [11], and support vector machines [12] have been utilized.
This study proposes an innovative colour matching method that can be used when the output variable is undefined. In this respect, it attempts to present an approach that differs from existing studies.

2. Materials and Methods

This section provides a comprehensive overview of the images utilized in the study, the methodologies employed for their acquisition, and the estimation techniques employed.

2.1. Data collection and preprocessing

Initially, it is essential to note that the material utilized for the study comprises 800x800 pixel images. An exemplar image is presented in Figure 1. Consequently, sample fabrics from the same order are evaluated collectively, as illustrated. Each fabric has been digitally imaged under conditions of homogeneous lighting and with fixed white balance settings.

Figure 1: Example Color Batching Fabrics

In evaluating the efficacy of the developed approaches, the batching results obtained by an expert for denim fabric samples from various production batches belonging to the same order were compared. Predictions were made for 2,139 fabric images obtained from a total of 59 orders.

In the pre-processing of the images, the RGB image values were first normalized to the 0–1 range using min-max transformation. This approach served to mitigate the impact of outlier values. Subsequently, CLAHE (Curvelet Local Adaptive Histogram Equalization) histogram equalization was performed. The CLAHE (Constant Histogram Area) method [13] involves the initial segmentation of the image into smaller tiles, followed by histogram equalization for each individual piece. This approach enables the examination of potential color differences.

2.2. Prediction Methods

Four different clustering methods were used for batch allocation. These are the K-Means, Agglomerative and DBSCAN algorithms. As these methods are based on different mathematical approaches and model the similarities between fabric samples in different ways, a fourth method, the ensemble learning method, is also recommended.

2.2.1. K-means Algorithm

The K-Means method is a frequently utilized and robust approach in the extant literature [14]. In the K-means algorithm, observations are divided into k predefined clusters, with each cluster being optimised around its centroid [15]. Equations 1 and 2 are used to determine the cluster centroids.

$$d(x,c) = \sum_{\forall i \in N} (x_i - c_j)^2 \qquad \text{(Eq. 1)}$$

$$c_j = \frac{1}{m} \sum_{l=1}^{m} x_l \qquad \text{(Eq. 2)}$$

In this study, the k value was determined between two and the maximum number of lots in an order. Each alternative algorithm was entered into the ensemble learning algorithm as a distinct prediction outcome. The silhouette score value was utilized as the foundation for determining the optimal number of clusters for the k-means approach. The silhouette score value is calculated as in Equation 3 for the average inter-cluster distance value a and the average intra-cluster distance value b [16].

$$silhoutte\ score = \frac{b-a}{max\ (a,b)} \qquad \text{(Eq. 3)}$$

2.2.2. Agglomerative Clustering Algorithm

Hierarchical clustering initially treats each example as a separate cluster and gradually merges similar clusters. This approach is similar to the k-means method in that it utilizes calculations based on the Euclidean distance. The following summary outlines the general steps of the algorithm [17]:

- The calculation of a proximity matrix is performed for each element.
- The objective is to identify the element that has the minimum distance in the matrix.
- The two clusters that are closest to each other in terms of distance are then combined into a single cluster.
- The proximity matrix is subject to modification through the evaluation of the newly identified cluster as a singular element.
- In the event that multiple clusterings remain, the process is to be repeated.

2.2.3. DBSCAN Algorithm

The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) method is a density-based algorithm that does not require the number of clusters to be specified in advance. The algorithm performs clustering based on the specified radius (r) and minimum number of points ($minpts$) data. For a given point p, the points that satisfy the proximity criterion are determined according to Equation 4. If the condition $N_r(p) > minpts$ is satisfied for this point, they are grouped together (18). Clustering is performed according to varying r and $minpts$ values. The ideal values are determined by examining the silhouette score value, akin to the k-means method.

$$N_r = \{q \in N \mid d(p,q) < r\} \qquad \text{(Eq. 4)}$$

2.2.4. Majority based Ensemble Clustering Algorithm

Each clustering algorithm generates its own cluster labels for the dataset. However, these labels may vary depending on the different parameter sensitivities and distance definitions of the methods [19]. To reduce these differences and obtain more stable results, a majority rule-based ensemble approach was applied.

The steps of this approach can be summarised as follows:

1. Cluster labels from three different algorithms are obtained for each fabric sample.
2. In cases where the majority is the same, this label is assigned as the final batch label for the sample.
3. In rare cases where multiple modes are formed, the cluster with the highest average similarity score is preferred.

3. Results

https://doi.org/10.56038/ejrnd.v5i1.700

Clustering analyses were performed using feature vectors obtained from 800×800 RGB images for each fabric sample. The results of the K-means method are presented in Figure 2. The confusion matrix was created by comparing user-assigned batching with k-means predictions. The K-means method demonstrated a high degree of success in implementing batching on the dataset. However, it exhibited limited sensitivity in intermediate regions characterized by smooth tone transitions. It was observed that the system was unable to make successful predictions, particularly in order samples with a low number of batches.

Figure 3 presents the results obtained for agglomerative clustering. Preliminary findings suggest that the proposed method generally outperforms the k-means algorithm. Although it has been possible to group fabrics with similar undertones in a more detailed manner, lot errors have increased in some clusters with high similarity.

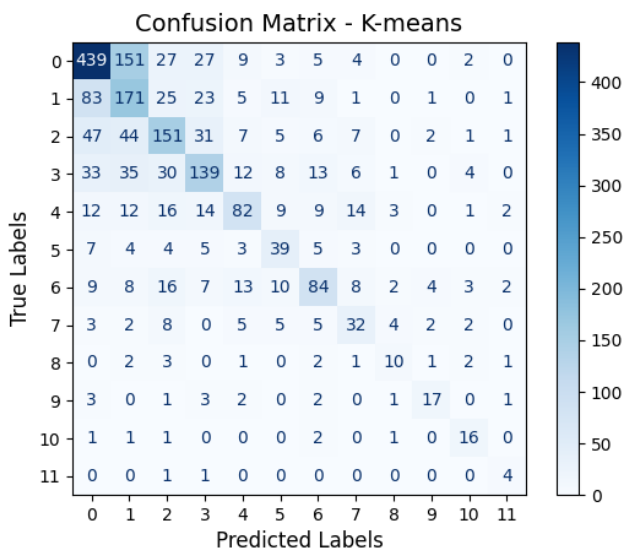

Figure 2: Confusion matrix for K-means algorithm

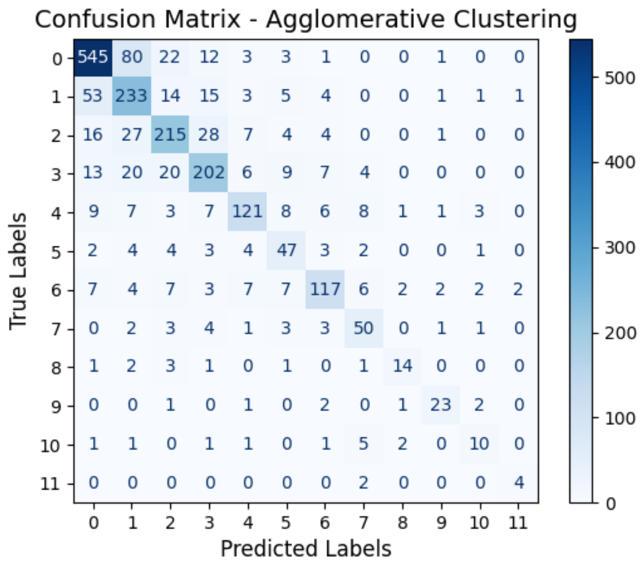

Figure 3: Confusion matrix for agglomerative clustering algorithm

As illustrated in Figure 4, the results obtained from the implementation of the DBSCAN algorithm are presented. A comparative analysis reveals that the DBSCAN algorithm consistently produces superior outcomes when utilized in comparison to alternative methodologies. The primary factor contributing to this phenomenon can be attributed to the substantial quantity of pixels and the observation that the DBSCAN algorithm demonstrates optimal performance when operating on data of this magnitude. However, in samples with a low number of lots, the program made erroneous selections by excluding some samples as noise.

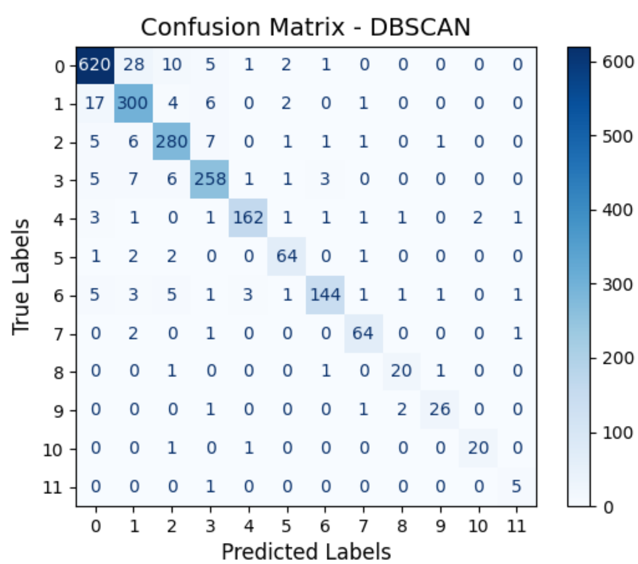

Figure 4: Confusion matrix for DBSCAN algorithm

Finally, when examining the ensemble clustering method in Figure 5, it can be said that it yields similar results to DBSCAN in many classes. It can be said that the ensemble method can produce more successful results, particularly when the number of lots increases.

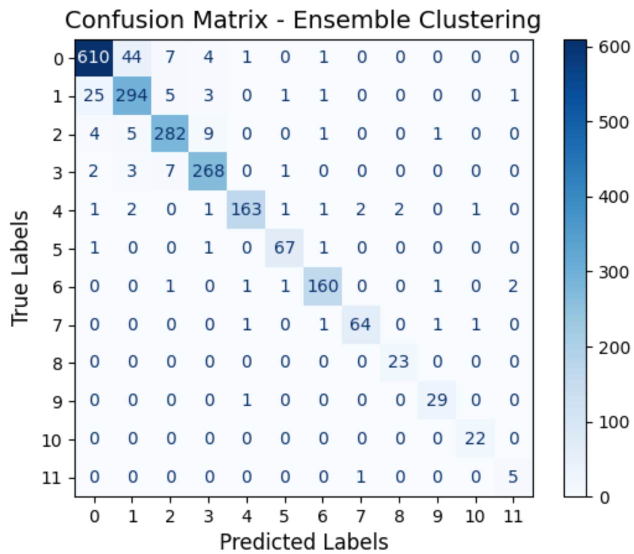

Figure 5: Confusion matrix for majority based ensemble algorithm

When performance is evaluated on an order basis (see Figure 6), it is evident that the most effective approaches are DBSCAN and ensemble methods. With the exception of a few substantial deviations observed in a limited number of orders, the ensemble clustering method generally produced satisfactory results. The ensemble outputs produced more clearly separated and homogeneous clusters in the color space, reducing the misclassification rate in the transition regions between clusters.

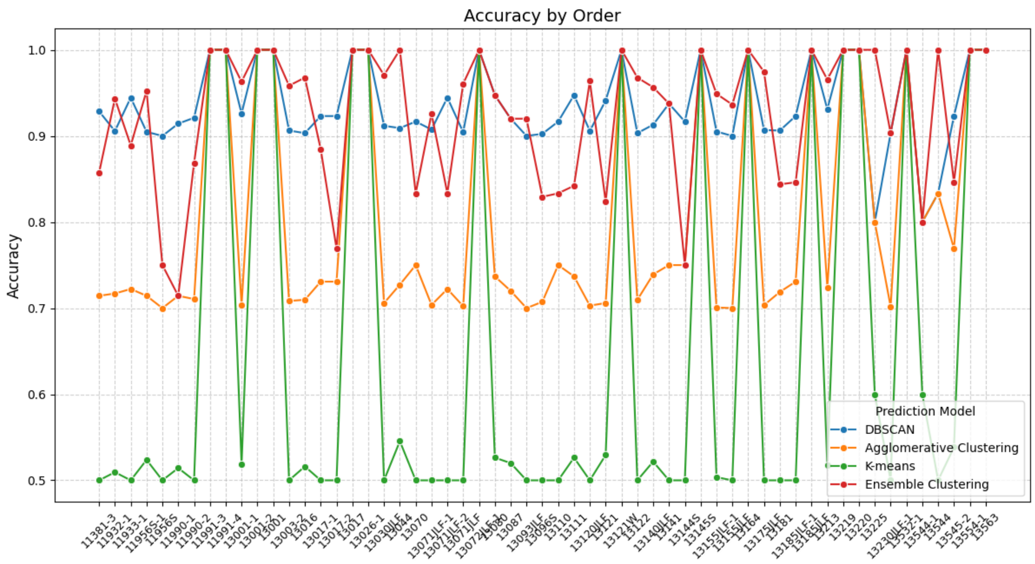

Figure 6: Comparing batch accuracies for orders

The mean values of precision, recall, and the F1-score for each batch in selecting the optimal method are presented in Table 1. As evidenced by the preceding graphs, DBSCAN and ensemble clustering exhibited analogous results; however, ensemble clustering demonstrated superior performance metrics, including the F1-score. The findings suggest that the sensitivity of each algorithm to distinct data distributions influences the interpretation of color variations.

Table 1: Comparison of the methods

Method	Mean precision	Mean recall	Mean f1-score
K-means	0.5123	0.5477	0.5209
Agglomerative Clustering	0.6849	0.6882	0.6843
DBSCAN	0.8845	0.9013	0.8913
Ensemble Clustering	0.9060	0.7744	0.9216

In the context of evaluating batch-based performance in the order, the F1 score demonstrated notable efficacy in a multitude of orders. The results, which are summarized in Table 2, demonstrate that when the lot number is particularly high (12) or low (2), the F1 score value is lower than in other examples.

Table 2: Batching results for majority based ensemble clustering algorithm

Method	Mean precision	Mean recall	Mean f1-score
K-means	0.5123	0.5477	0.5209
Agglomerative Clustering	0.6849	0.6882	0.6843
DBSCAN	0.8845	0.9013	0.8913
Ensemble Clustering	0.9060	0.7744	0.9216

The findings suggest that ensemble approaches, which integrate the outputs of multiple clustering methods rather than utilizing them separately, can generate more consistent results. The majority rule-based ensemble approach developed for color batch matching in denim fabrics has yielded stable and repeatable results that closely resemble expert opinions by combining the strengths of different clustering algorithms. These results establish a significant foundation for the implementation of artificial intelligence-supported quality assurance applications within the textile industry.

4. Discussion and Conclusion

The proposed model is a majority-based ensemble clustering model capable of performing automatic lot sorting based on digital images of denim fabric samples. Each fabric sample was entered as input in RGB format at a size of 800×800 pixels. The DBSCAN, Agglomerative clustering, K-Means, and ensemble approach based on the majority rule principle were then compared. The results obtained were then compared with the results of manual lot sorting performed by experts. The performance of both individual clustering methods and the proposed ensemble model was analyzed. The objective of this study is to attain more consistent results that approximate expert opinions in the color-based lot sorting process by integrating the strengths of various algorithms.

Consequently, this research presents a concrete step towards the digitalization of quality control processes in denim production and focuses on minimizing human-based errors. This finding underscores the potential of the proposed machine learning-based majority rule approach as a reliable alternative for the automatic detection of color variations. In subsequent studies, the adaptation of diverse ensemble and clustering methods to this problem can be explored. Furthermore, the adaptation of supervised learning methods to enhance prediction accuracy is a potential avenue for future research.

References

[1] Wang, Y. W., Yi, Q. Z., Ding, Y., Ji, F., & Wang, N. (2021). Study on the factors influencing the dyeing performance of cotton fabric with vat dyes based on principal component analysis. The Journal of the Textile Institute, 112(9), 1460-1466.

[2] Bhuiyan, M. R., Shaid, A., Bashar, M. M., & Sarkar, P. (2016). Investigation on dyeing performance of basic and reactive dyes concerning jute fiber dyeing. Journal of Natural Fibers, 13(4), 492-501.

[3] Ramratan, R. K., & Singh, S. (2020). Study of denim jeans fabric on finishing process and characteristic performances. Asian Textile Journal, 29, 26-30.

[4] Kan, C. W., Wong, W. Y., Song, L. J., & Law, M. C. (2013). Prediction of Color Properties of Cellulase-Treated 100% Cotton Denim Fabric. Journal of Textiles, 2013(1), 962751.

[5] Hung, O. N., Chan, C. K., Kan, C. W., Yuen, C. W. M., & Song, L. J. (2014). Artificial neural network approach for predicting colour properties of laser-treated denim fabrics. Fibers and Polymers, 15(6), 1330-1336.

[6] Zulfiqar, A., Manzoor, T., Ijaz, M. B., Nawaz, H. H., Ahmed, F., Akhtar, S., ... & Umar, M. (2024). Artificial-neural-network-based predicted model for seam strength of five-pocket denim jeans: a review. Textiles, 4(2), 183-217.

[7] Song, B., Pan, R., Zhang, N., Xiang, J., & He, W. (2024). Visual similarity simulation of slub denim based on image processing. Textile Research Journal, 94(1-2), 120-135.

[8] Furferi, R., Governi, L., & Volpe, Y. (2016). Color matching of fabric blends: Hybrid Kubelka-Munk+ artificial neural network based method. Journal of Electronic Imaging, 25(6), 061402-061402.

[9] Xiao, B., Bi, W., Jia, X., Wei, H., & Adelson, E. H. (2016). Can you see what you feel? Color and folding properties affect visual–tactile material discrimination of fabrics. Journal of vision, 16(3), 34-34.

[10] Hwang, J. P., Kim, S., & Park, C. K. (2015). Development of a color matching algorithm for digital transfer textile printing using an artificial neural network and multiple regression. Textile Research Journal, 85(10), 1076-1082.

[11] Chaouch, S., Moussa, A., Ben Marzoug, I., & Ladhari, N. (2019). Application of ant colony optimization to color matching of dyed cotton fabrics with direct dyestuffs mixtures. Color Research & Application, 44(4), 556-567.

[12] Yani, Z. H. A. N. G., Jianqiang, Z. H. U. A. N. G., Rong, H. U. A. N. G., Aihua, D. O. N. G., & Haodong, Y. U. A. N. (2022). Smart Clothing Fabric Color Matching with Reference to Popular Colors. Journal of Donghua University (English Edition), 39(4).

[13] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas and K. E. Muller, "Contrast-limited adaptive histogram equalization: speed and effectiveness," *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, Atlanta, GA, USA, 1990, pp. 337-345, doi: 10.1109/VBC.1990.109340.

[14]Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. Electronics, 9(8), 1295.

[15]Chong, B. (2021). K-means clustering algorithm: a brief review. Academic Journal of Computing & Information Science, 4(5), 37-40.

[16]Shahapure, K. R., & Nicholas, C. (2020, October). Cluster quality analysis using silhouette score. In 2020 IEEE 7th international conference on data science and advanced analytics (DSAA) (pp. 747-748). IEEE.

[17]Bouguettaya, A., Yu, Q., Liu, X., Zhou, X., & Song, A. (2015). Efficient agglomerative hierarchical clustering. Expert Systems with Applications, 42(5), 2785-2797.

[18]Khan, K., Rehman, S. U., Aziz, K., Fong, S., & Sarasvady, S. (2014, February). DBSCAN: Past, present and future. In The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014) (pp. 232-238). IEEE.

[19]Golalipour, K., Akbari, E., Hamidi, S. S., Lee, M., & Enayatifar, R. (2021). From clustering to clustering ensemble selection: A review. Engineering Applications of Artificial Intelligence, 104, 104388.