European Journal of Wood and Wood Products (2025) 83:131
https://doi.org/10.1007/s00107-025-02280-6

ORIGINAL ARTICLE

Insights into a new overclustering technique using machine learning for a self-selecting bin-restricted colour sorting setup for light red meranti (*Rubroshorea* spp.)

Chiat Oon Tan[1,2,6] · Shigenobu Ogata[1] · Hwa Jen Yap[2] · Ichiro Nakamoto[1] · Zuriani Usop[5] · Mohd 'Akashah Fauthan[5] · Shaer Jin Liew[6] · Siew-Cheok Ng[3,4]

Received: 4 April 2025 / Accepted: 27 May 2025

Abstract

Timber colour sorting is an important woodworking process in producing a homogeneously coloured and pleasant looking product. However, for multispecific timber such as light red meranti, LRM (*Rubroshorea* spp.), which spans a wide gamut of colours, there is an antagonistic compromise between having good separability of colour and the number of bins. This research attempts to solve this by intentionally overclustering the intensity gamut and then automating the selection of ideal colour sorting bins (CSB) for a given batch size to produce high-similarity coloured sorting. 178,327 unique LRM wood samples collected over 8 months of production were used. Machine learning clustering algorithms such as k-means and Otsu multithresholding were tested against percentile and equal spacing methods. Batch sizes of 250 (B250) and 1,000 (B1000) pieces were evaluated. Maximum likelihood estimation was tested against statistical methods to select the CSB, and ideal overcluster setups were determined using the average delta E (ΔE^*_{00}) assessment. The 'burn-in rates' of 3–30 pieces were then evaluated. For the B250 four-bin setup, six overclusters (6C4) performed best, with a recommended 'burn-in rate' of 12 pieces. For B1000, 5C4 performed best with a 'burn-in rate' of 10 pieces. The 4C3 configuration and the 'burn-in rate' of 10 pieces were found to be the best for three-CSB for both B250 and B1000. This study shows the feasibility of using machine learning to automate the bin selection process when the overclustering technique is used to improve colour sorting in situations with a restricted number of bins.

1 Introduction

Wood colour sorting is a process by which timber that are determined to be visually similar are clustered together to be processed and assembled into some wood product. This process is crucial in producing a pleasingly homogeneous appearance, which has been shown to be preferred by consumers (Høibø and Nyrud 2010). The sorting of wood by colour also improved density estimation algorithms (Tan et al. 2025) by clustering wood with similar density profiles. This allows manufacturers to create products, particularly laminated boards made of wood pieces with similar density, thus improving their long-term stability during their service life. The colour of the wood was also used to track and

✉ Shigenobu Ogata
ogata@me.es.osaka-u.ac.jp

✉ Siew-Cheok Ng
siewcng@um.edu.my

[1] Department of Mechanical Science and Bioengineering, Graduate School of Engineering Science, The University of Osaka, 1-3, Machikaneyama, Toyonaka, Osaka 560-8531, Japan

[2] Department of Mechanical Engineering, Universiti Malaya, Jalan Profesor Diraja Ungku Aziz, 50603 Kuala Lumpur, Malaysia

[3] Department of Biomedical Engineering, Universiti Malaya, Jalan Profesor Diraja Ungku Aziz, 50603 Kuala Lumpur, Malaysia

[4] Centre for Applied Biomechanics, Universiti Malaya, Jalan Profesor Diraja Ungku Aziz, 50603 Kuala Lumpur, Malaysia

[5] Machinery Technology Centre, SIRIM Berhad, 1A, Persiaran Zurah, Kawasan Perindustrian Rasa, 44200 Hulu Selangor, Selangor, Malaysia

[6] Hexagon World Sdn Bhd, C3-30-01, Centrus SOHO, Persiaran Cyberpoint Timur, Cyber 12, 63000 Cyberjaya, Selangor, Malaysia

Published online: 25 June 2025

characterise the change in colour in Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*) undergoing thermal modification (Torniainen et al. 2021b), giving an indication of the level of heat treatment and the quality of the treated wood (Torniainen et al. 2021a).

Therefore, to produce high-quality wood products, manufacturers must dedicate a significant proportion of their workforce to the selection and sorting of materials. This is traditionally performed manually (Wang et al. 2021), which is highly susceptible to human fallibility due to the psychophysical nature of colour perception and fatigue after long hours of work. There are industrial scanners available, such as the MiCROTEC WoodEye Parquet, Weinig CombiScan Sense, and ATB Blank Spectra. In fact, colour sorting is a very common industrial requirement, particularly in agriculture and food production, where colour is a key indicator of quality (ripeness, freshness, etc.). These include products such as salmon fillet (Ranjan et al. 2024), apples (Meenu et al. 2024), and rice (Cha et al. 2025).

Past studies on colour sorting in wood include the use of soft colour descriptors and k-Nearest Neighbours for parquet sorting (Bianconi et al. 2013) using the RGB colour space and comparing it with other colour spaces. Another study attempted to grade oak boards by colour (Defoirdt et al. 2012). This is done by comparing the CIEDE2000 distances to their respective sub-space centres and using human observers as yardsticks. K-means clustering method was also used for the colour sorting of beech wood (*Fagus sylvatica*) (Wang et al. 2021) using soft descriptors and colour histograms as features. Lin et al. (2020b) also used k-means to group wooden boards (unspecified species), but added an image preprocessing step to remove dark stains from the image which improved performance. An older but more direct method was to compare the histograms of the wood with those of the colour clusters and to set appropriate thresholds (Lu et al. 1997). A combination of the textural (using grey-level co-occurrence matrix, GLCM) and colour features of wood was also tested for oak parquet that had large variability in terms of colour and texture (Rožman et al. 2006). GLCM has also been recently used to classify textures on veneers (Savolainen 2023). More advanced methods would be to use deep learning algorithms, such as convolutional neural network (Liu et al. 2020) for rubberwood, XGBoost (Zhuang et al. 2021) for an unspecified flooring species, and Vision Transformer coupled with Densenet121 (Zhuang et al. 2022) for *Pometia pinnata* hardwood flooring.

KayuSort is an industrial prototype colour sorting system developed to handle multiple types of species, based on the algorithm developed in a previous study for red oak (*Quercus rubra*), yellow poplar (*Liriodentron tulipifera*) and maple (*Acer* spp.) veneers (Liew 2024). KayuSort was also tested for light red meranti [LRM, previously *Shorea* spp., but recently reclassified as *Rubroshorea* spp. per Ashton and Heckenhauer (2022); Kew Gardens (2024)] (Tan et al. 2025). However, a minor but persistent issue is the fact that LRM is a multispecific wood, encompassing 11 (Lim et al. 2016) different species from the *Rubroshorea* genus in Peninsular Malaysia, and having a wide range of densities (385–755 kgm^{-3}). Several LRM species overlap with those of dark red meranti (DRM), namely rambai daun (*Rubroshorea acuminata*), daun besar (*R. hemleyana*), kepong hantu (*R. macrantha*), tengkawang ayer (*R. palembanica*) and paya (*R. platycarpa*) (Gan and Lim 2004), differentiated only by their density. Due to this, the colour range of commercially available LRM is similarly broad. Therefore, one common practical difficulty faced by colour sorting algorithms is this under-clustering issue which none of the past research addressed.

It is ideal to train any colour sorting algorithm to cover the entire gamut of colours available for any given species, like what is being done in previous studies. However, in reality, timber arrives at a processing facility in batches, with a smaller breadth of colour distribution, compared to the total possible colour for the species, particularly one with a huge variation in colour like LRM. This means that a universal colour sorting binning system will result in poorly resolved separation of the timber with certain batches. To complicate things, most production facilities do not want the wood sorted into too many bins due to space and handling constraints. Also, they are often not concerned about the timber's absolute colour parameters but are more concerned about their colour homogeneity within the sorted bins, which when assembled will produce a pleasant looking product.

The latest upgraded version of KayuSort involved training separate colour swatches using the same dataset used for this study for various shades of darkness and lightness (for four bins, which were the system requirements desired by the facility) using six intentionally overclustered groups (hereinafter known as 'overclusters' in this study), and the sorting maps can be manually selected by the user depending on the perceived lightness or darkness of the batch of timber. The six overclusters were clustered using the quantile method and were arbitrarily chosen to allow the user to have three options (dark, medium, and light) given four desired bins each. However, manual selection of the correct set of bins proved to be tedious, whereby the timber in that batch must be manually combed through to provide adequate samples representative of all the available colours in the batch.

In this study, we first determined the ideal number of overclusters for a given number of bins using the long-term dataset gathered from the field. The scope of this study

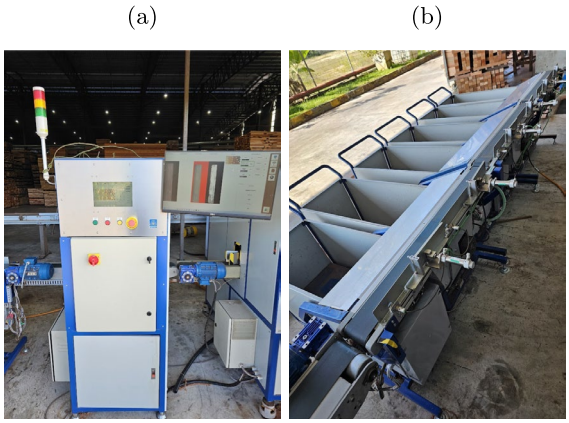

Fig. 1 KayuSort colour sorting system consisting of **a** the colour scanner, and **b** the sorting mechanism

covers three and four colour sorting bins. Having established the best overcluster setup, we then explored methods for automatically selecting the ideal bin configuration to achieve the best sorting performance (smallest colour variation between members in the bin). This involves sampling some timber and making a prediction of the colour distribution of the batch and then evaluating their efficacy.

2 Materials and methods

This section shall be divided into five subsections. The first subsection describes the samples and the equipment used to obtain the wood images. The second subsection explicates the feature extraction process. The third and fourth subsections explain the two salient algorithms used in this study, which are maximum likelihood estimation and the proposed new ΔE^*_{00} based performance metric. Lastly, the final subsection details the three steps to optimise the overclustered solution and discover the ideal colour sorting setup. These steps are (a) choosing the batch size, (b) selecting the ideal number of overclusters and bin configuration, and (c) determining the recommended 'burn-in rate'.

Statistical and machine learning tools and functions from the MATLAB R2024b Update 1 64-bit (24.2.0.2740171) Statistics and Machine Learning Toolbox (MathWorks Incorporated 2024) were used to perform the analyses.

2.1 Image sampling

The KayuSort industrial colour sorting prototype machine was used to acquire timber images. The imaging system consisted of an industrial line scan camera system (Hikrobot MV-CL022-40GC coupled with an MVL-KF1224M-25MP lens) with a CCS LNSP2-200SW line light (6600 K colour temperature), as shown in Fig. 1. The entire mechanical conveyor and sorting system was managed by an Industrial Shields MDuino 58+ industrial programmable logic controller, while a Dell PC (where KayuSort software ran) performed image capture, storage and colour processing for colour sorting. The current setup was configured with seven sorting bins (3 x modular double bins and 1 x bin at the end of the conveyor) and can be expanded in the future. The line scan camera was triggered by a proximity sensor on the input side of the conveyor. All autogain, auto white balance, and gamma correction settings were turned off. A Calibrite Colour Checker Passport 2 Macbeth colour calibration chart was used as a reference target for colour constancy adjustments.

KayuSort was coded in Visual Studio 2022 64-bit version 17.9.2 (Microsoft Corporation 2022) and is capable of sorting into 10 bins (software limit, and can be increased). Its core sorting algorithm is the self-organising maps (SOM) artificial neural network. The network training was designed to be semi-supervised and flexible, whereby the colour classes can be assigned by the workers based on u-matrix observation after training, manual feeding of pre-sorted labelled samples during image acquisition, or off-site clustering and training using acquired wood images.

A total of 178,685 unique images of LRM timber were collected between December 2023 and July 2024 in five separate periods when personnel were available to supervise data collection (detailed in Table 1), representing an eight-month sample of timber that arrived at Sim Seng Huat Industries Sdn Bhd (SSH). The colour distribution of these data was assumed to be a good representation of all possible

Table 1 The period when data from the original and superpopulation datasets used in this study were collected, the number of images per period, and the basic statistics upper L^*, a^*, and b^* values, and range of upper L^* values (s.d. = standard deviation)

Period	Date span	No. of images	Average upper L^* [s.d.]	Upper L^* range	Average upper a^* [s.d.]	Average upper b^* [s.d.]
Original	23/10*	392	75.1 [4.8]	62.6–85.8	7.3 [0.67]	– 0.1 [0.87]
1	04/12–18/12*	27,018	75.4 [5.96]	55.4–94.8	7.5 [0.71]	0.1 [0.96]
2	29/01–07/02	27,498	74.4 [5.92]	55.4–94.8	7.2 [0.77]	0.3 [0.96]
3	22/03–29/03	24,533	68.3 [5.24]	54.6–94.1	7.4 [0.68]	0.8 [0.92]
4	29/05–05/07	82,232	68.1 [5.65]	41.0–93.9	7.1 [0.65]	0.8 [0.90]
5	11/07–17/07	17,404	66.2 [5.17]	53.6–90.4	7.2 [0.58]	1.2 [0.89]
Overall†		178,685	70.0 [6.53]	41.0–94.8	7.2 [0.69]	0.6 [0.98]

Data collected in 2023, others were in 2024; † Excluding the original dataset

LRM colour variations in Peninsular Malaysia, as the mill sources their timber from several different forest concessions throughout Peninsular Malaysia. The wood pieces were of various sizes (the machine accepts lengths of 150 to 900 mm, widths of 75 to 150 mm and thicknesses of 25 to 75 mm), surfaced on two sides, and predominantly free of major defects such as knots, wane, and huge cracks. Minor defects such as live knots, pinholes, hairline cracks, and mineral streaks were not excluded. The wood pieces were already filtered by the workers for major defects, such as dead knots, stains, or rot, during the cross-cutting process prior to imaging. In this study, it was also assumed that the sequential feeding of the timber into the colour sorting machine was performed randomly, which, from observations, appeared to be so (as prior processes had mixed the timber from the incoming batch onto the pallet that fed into the colour sorting machine). In this study, this dataset shall be referred to as the 'superpopulation'.

The older, initial KayuSort training dataset consisted of 392 images of samples that were free of defects, including all minor defects, and of various sizes (dimension range similar to the dataset above). This dataset was obtained within a short period of time due to time constraints during the initial setup of the system for deployment and appeared to work well within the prototype testing phase. This dataset was separately evaluated to provide a baseline comparison for this study. Details of this dataset are also reported in Table 1. In this study, this dataset shall be referred to as the 'original dataset'.

Since colour sorting was the sole focus of this study, only the wood image was obtained and other parameters were not measured to avoid impeding regular production. The typical moisture content of wood on the production floor was 11.6% (standard deviation, s.d. = 1.3%, measured using Delmhorst JX-30 moisture meter) based on a previous study that involved the estimation of density in the same facility (Tan et al. 2025). It was assumed that the wood that arrived at the facility was properly kiln dried and that any variations in the moisture content did not influence the colour of the wood.

2.2 Feature extraction

The features extracted and used in this study were the average RGB values (red, green and blue) of the images of the wood samples, and their CIELAB converted values. Firstly, the raw images were cropped so that only the wood was visible. KayuSort's standard Otsu bisected dual-stratum colour features were obtained (Liew et al. 2023). The average RGB values of the lower and upper strata were then converted to CIELAB (Commission Internationale de l'Eclairage $L^*a^*b^*$) colour space using the D65 standard reference illuminant configuration (closest standard white point to the 6600 K lighting used in this study) (International Color Consortium 2004). This was performed by first applying gamma correction to the RGB values as per Eqs. (1) and (2), and subsequently converting them to CIEXYZ colour space per Eq. (3).

$$f_\gamma(u) = \begin{cases} -(1.055 \cdot |u^\gamma| - 0.055), & u \le -a \\ 12.92 \cdot u, & -a < u < a \\ 1.055 \cdot u^\gamma - 0.055, & u \ge a \end{cases} \quad (1)$$

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} f_\gamma\left(\frac{R}{255}\right) \\ f_\gamma\left(\frac{G}{255}\right) \\ f_\gamma\left(\frac{B}{255}\right) \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124564 & 0.3575761 & 0.1804375 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \quad (3)$$

where

$f_\gamma(u)$ gamma correction function for colour value u, where $0 \le u \le 1$

γ gamma correction factor $= \frac{1}{2.4}$

a conditional parameter with value of 0.0031308

R, G, B red (R), green (G) and blue (B) values, values from 0 to 255

R', G', B' gamma corrected RGB values

X, Y, Z transformed RGB values in CIEXYZ colour space

Using D65 reference illuminant white point values shown in Eq. (4) (International Color Consortium 2004), the CIEXYZ values were subsequently converted to their CIELAB values using Eqs. (5) and (6). All definitions and some of the conversions are also defined in ISO 13655:2017 (International Organization for Standardization 2017).

$$\begin{bmatrix} X_n \\ Y_n \\ Z_n \end{bmatrix} = \begin{bmatrix} 95.0489 \\ 100 \\ 108.884 \end{bmatrix} \quad (4)$$

$$f(t) = \begin{cases} \sqrt[3]{t}, & t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}t\sqrt{\frac{6}{29}} + \frac{4}{29}, & t \le \left(\frac{6}{29}\right)^3 \end{cases} \quad (5)$$

$$\begin{bmatrix} L^* \\ a^* \\ b^* \end{bmatrix} = \begin{bmatrix} 116f\left(\frac{Y}{Y_n}\right) - 16 \\ 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \\ 200\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \end{bmatrix} \quad (6)$$

where

X_n, Y_n, Z_n reference white point value in CIEXYZ

$f(t)$ function for quotient t used in Eq. (6)

t quotient between the X, Y, and Z values with their

respective reference white points
L^*, a^*, b^* CIELAB values

The CIELAB distribution of all images is shown in Fig. 2. L^*, which represents the lightness component of the image, exhibited a large spread (Fig. 2a), while the chromatic components a^*(green-red component, negative to positive) and b^*(blue-yellow component, negative to positive) appeared to be very concentrated (Fig. 2b). L^* had an intensity span of 53.8 intensity values (mean = 70.0, s.d. = 6.53) while a^* and b^* spanned 8.7 (mean = 7.2, s.d. = 0.69) and 14.5 (mean = 0.6, s.d. = 0.98) chromatic positional values, respectively. Since CIELAB colourspace is defined by the International Commission on Illumination (CIE) as an almost perceptually uniform colour space, the large variance in L^* values showed that it is the most significant contributor to the colour difference in the dataset. In this study, the upper colour stratum of the image is used to mitigate the influence of dark coloured defects (such as live knots, pinholes, hairline cracks, and mineral streaks) present in the wood images on the results.

The data were then cleaned up by winsorising the top and bottom 0.1%, thereby removing extreme outliers. Only samples with upper L^*(L^*_u) values between 54.2 and 90.1 were used in the analysis (N = 178,327 samples), with a trimmed mean value of 70.0 (s.d. = 6.48, median = 69.7). In practical terms, any wood with upper L^* values outside this range can either be included in the edge bins or rejected outright. Both mean and median were centrally located between the 5th (59.7) and 95th (81.2) percentiles with a skewness coefficient of 0.205. However, using the Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) normality tests, and by observing the normal quantile-quantile (QQ) plot (Fig. 2c), the distribution of the data is skewed from a normal distribution (p-values < 0.001 for both KS and AD).

2.3 Distribution fitting function

Maximum likelihood estimation (MLE) is a popular machine learning technique for obtaining values from discrete observations to fit a known or presumed function and returning the function's parameters. MLE's general likelihood function is shown in Eq. (7), and the maximisation equation in Eq. (8). MLE aims to obtain the distribution function's parameters where the observations have the highest joint probability (dot product of each observation's probability function). For this study, two distribution functions were tested, which are the Gaussian and beta distributions. Since the superpopulation appeared to be skewed, both of these methods were evaluated to determine which method predicts the centrality of each batch distribution best when selecting the ideal bins for colour sorting.

$$\mathcal{L}(\theta|x) = f(x|\theta) = \prod_{i=1}^{n} f_i(x_i|\theta) \tag{7}$$

$$\hat{\theta} = \arg\max_{\theta=1}^{n} \mathcal{L}(\theta|x) \tag{8}$$

where
$\mathcal{L}(\theta|x)$ objective likelihood function of the parameter θ given observation x
f the product of univariate density functions
n number of observations
$\hat{\theta}$ maximum likelihood estimator for parameter θ

The Gaussian distribution has a likelihood function as shown in Eq. (9). The estimated mean and variance are calculated using Eqs. (10) and (11) respectively.

$$\mathcal{L}_G(\mu, \sigma^2|x) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \tag{9}$$

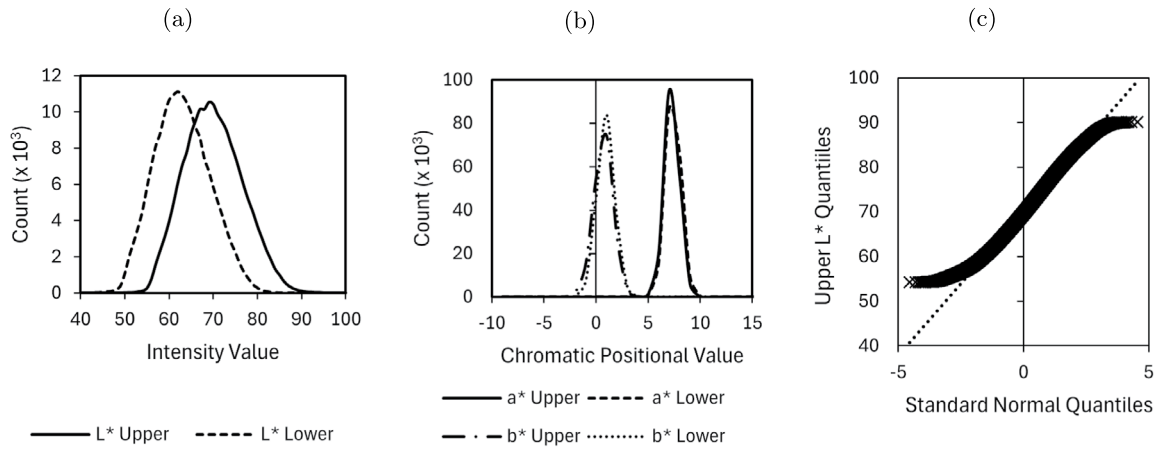

Fig. 2 Distribution of CIELAB **a** L^*, **b** a^* and b^* values of the upper and lower strata for all the wood images, and **c** normal quantile (Q-Q) plot for CIELAB L^* upper stratum values for the trimmed dataset

$$\hat{\mu}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{10}$$

$$\hat{\sigma}^2(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{11}$$

where
$\mathcal{L}_{\mathrm{G}}$ likelihood function for a Gaussian function with input x
μ, σ mean and variance of observation's distribution
n number of observations
$\hat{\mu}, \hat{\sigma}^2$ estimated μ and σ^2 of the Gaussian distribution

For the beta distribution, the likelihood function is shown in Eq. (12), given the gamma function in Eq. (13). The estimated shape parameters ($\hat{\alpha}$ and $\hat{\beta}$) must first be obtained using iterative methods (such as the Gauss-Newton algorithm) (Gnanadesikan et al. 1967), approximation techniques (Beckman and Tietjen 1978), or the Bayesian approach (Talib Othman 2022). With the estimated shape parameters, its estimated mean can be calculated using Eq. (14) (Johnson et al. 1995) while the median can be approximated using Eq. (15) (Kerman 2011). In this study, the MLE-approximated Gaussian method shall be referred to as MLE-G, whereas the MLE-approximated beta distribution shall be referred to as MLE-B.

$$\mathcal{L}_{\mathrm{B}}(\alpha, \beta | x) = \prod_{i=1}^{n} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \tag{12}$$

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt = (x-1)\Gamma(x-1) \tag{13}$$

$$\hat{\mu}_{\mathrm{B}} = \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}} \tag{14}$$

$$\hat{m}_B \approx \frac{\hat{\alpha} - \frac{1}{3}}{\hat{\alpha}+\hat{\beta}-\frac{2}{3}} \tag{15}$$

where
$\mathcal{L}_{\mathrm{B}}$ likelihood function for beta function with input x, where $0 \leq x \leq 1$
Γ gamma function
α, β shape parameters for beta distribution
n number of observations
$\hat{\alpha}, \hat{\beta}$ estimated shape parameters for beta distribution
$\hat{\mu_{\mathrm{B}}}, \hat{m}_{\mathrm{B}}$ estimated mean and median of the beta distribution

2.4 Performance assessment

Previous colour sorting research typically used humans to measure the performance of their proposed models. However, this is not practical on the scale of the data used in this study. Furthermore, KayuSort had been shown to perform better than humans with much smaller variances in colour in each cluster (Tan et al. 2025).

ΔE^*_{00} is a well established method to measure the difference between two colours and had been used in previous wood research such as to perform colour grading using a spectrophotometer (Defoirdt et al. 2012) and to characterise colour change before and after wetting the timber (Meints et al. 2017). Therefore, to assess the performance of each method in this study, inspiration was taken from the study conducted by Defoirdt et al. (2012). ΔE^*_{00} was originally conceived in 1976 as an attempt to establish a quantitative measure of colour difference in a parametrically equal manner, as shown in Eq. (16).

$$\Delta E^*_{76} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \tag{16}$$

where ΔE^*_{76} colour difference value defined in CIE76 standard
$\Delta L^*, \Delta a^*, \Delta b^*$ difference in L^*, a^* and b^* values between two colours

However, this equation was found to be imprecise and was improved in 1994 with compensation factors, and again in 2000 (CIEDE2000) culminating in Eq. (17) (Luo et al. 2001; Sharma et al. 2005; Mokrzycki and Tatol 2011). Details on each of the terms in Eq. (17) with respect to the values of L^*, a^* and b^* can be found in the study by Sharma et al. (2005). All ΔE^*_{00} values used in this paper refer to that of CIEDE2000.

$$\begin{aligned}
\Delta E^*_{00} &= \sqrt{\mathcal{L}^2 + \mathcal{C}^2 + \mathcal{H}^2 + c} \\
\mathcal{L} &= \frac{\Delta L'}{k_L S_L} \\
\mathcal{C} &= \frac{\Delta C'}{k_C S_C} \\
\mathcal{H} &= \frac{\Delta H'}{k_H S_H} \\
c &= R_T \frac{\Delta C'}{k_C S_C} \frac{\Delta H'}{k_H S_H}
\end{aligned} \tag{17}$$

where
ΔE^*_{00} colour difference value as defined in CIEDE2000 standard
$\Delta L'$ difference in L^* values between two colours
$\Delta C'$ difference in chroma values
$\Delta H'$ difference in hue values in circular coordinates

k_L, k_C, k_H parametric weights
S_L, S_C, S_H compensation factor for lightness, chroma, and hue respectively
R_T hue rotation

The ideal measurement is to average all the ΔE^*_{00} values of every pair of samples in the same colour bin, which gives an indication of how similar all constituents are to each other. However, this method is computationally expensive as the number of samples increases. Instead, we used a simpler approach by obtaining the centroid RGB colour of the samples in each bin and calculating the ΔE^*_{00} of every constituent of the bin to their respective centroid. In doing this, we only needed to perform $N\Delta E^*_{00}$ calculations, rather than its combination ${}^N C_2$ number of calculations (where N is the number of samples in the bin).

The CIE latest guideline on the interpretation of ΔE^*_{00} values is shown in Table 2 (Karma 2020). ΔE^*_{00} values are bounded between 0 and 100, where colour difference values below one unit are imperceptible to humans. Many applications involving the need for accurate colour reproduction, such as professional displays, use ΔE^*_{00} values of 2.0 (ViewSonicCorporation 2021), 2.3 (Light Illusion Ltd 2025), or 3.0 (BenQ Corporation 2014) as thresholds under which to define high colour accuracy. Determining the acceptable ΔE^*_{00} value is industry- and application-specific, such as 2.0 in the printing industry (Huda 2017).

Furthermore, according to the CIE guideline in Table 2 and observing the standard deviation values in Table 1, the values of a^* and b^* in the data did not appear to contribute much to the perceptible colour difference, while the variance of L^* disproportionately influenced the value of ΔE^*_{00}. Furthermore, each batch was only going to be sorted into three or four bins, and hence the chromatic differences (a^* and b^*) were likely overwhelmed by intensity differences (L^*). Since the images were stratified into lighter wood colour and darker grain colour (and possibly minor defects), the L^*_{u} values representing lighter wood colour were the main focus in this study.

The centroid colour value for each cluster was found and converted to CIELAB, as per Eq. (18). Subsequently, to score each clustering approach, the overall average ΔE^*_{00} value ($\Delta \bar{E}^*_{00}$) was calculated, which is the average ΔE^*_{00}($\Delta \bar{E}^*_{00,c}$) of all constituents of the cluster c to the centroid colour of their respective cluster, as shown in Eq. (19). The clustering method that scored the lowest overall average $\Delta \bar{E}^*_{00}$ was deemed the ideal method, as it meant that it had the smallest variance in their colour differences with the centroid colour. Therefore, the smaller the value, the greater the perceived colour consistency. This is similar to the CIELAB tolerancing approach (Huda 2017) when setting an acceptable value of ΔE^*_{00} with the centroid colour being analogous to the tolerancing reference colour. Also, because the values of ΔE^*_{00} are always positive by definition, taking an average of the ΔE^*_{00} values is almost equivalent to determining the mean absolute error of the constituents to their centroid colour value.

$$\begin{aligned}\begin{bmatrix}\bar{L}^*_c\\ \bar{a}^*_c\\ \bar{b}^*_c\end{bmatrix} &= f_{\mathrm{Lab}}\left(\begin{bmatrix}\bar{R}_c\\ \bar{G}_c\\ \bar{B}_c\end{bmatrix}\right)\\ &= f_{\mathrm{Lab}}\left(\begin{bmatrix}\frac{1}{N_c}\sum_{i=1}^{N_c} R_{c,i}\\ \frac{1}{N_c}\sum_{i=1}^{N_c} G_{c,i}\\ \frac{1}{N_c}\sum_{i=1}^{N_c} B_{c,i}\end{bmatrix}\right)\end{aligned} \tag{18}$$

$$\begin{aligned}\Delta \bar{E}^*_{00} &= \frac{1}{N}\sum_{c=1}^{N} \Delta \bar{E}^*_{00,c}\\ &= \frac{1}{N}\sum_{c=1}^{N}\left[\frac{1}{N_c}\sum_{i=1}^{N_c} f_{\Delta \mathrm{E}}\left(\begin{bmatrix}\bar{L}^*_c\\ \bar{a}^*_c\\ \bar{b}^*_c\end{bmatrix}, \begin{bmatrix}L^*_{c,i}\\ a^*_{c,i}\\ b^*_{c,i}\end{bmatrix}\right)\right]\end{aligned} \tag{19}$$

where
$\bar{L}^*_c, \bar{a}^*_c, \bar{b}^*_c$ centroid L^*, a^* and b^* values of cluster c
$\bar{L}^*_{c,i}, \bar{a}^*_{c,i}, \bar{b}^*_{c,i}$ L^*, a^* and b^* values of each element i in cluster c

$\bar{R}_c, \bar{G}_c, \bar{B}_c$ average red, green and blue values of the cluster c

$\bar{R}_{c,i}, \bar{G}_{c,i}, \bar{B}_{c,i}$ red, green and blue values of each element in cluster c
$f_{\mathrm{Lab}}(C)$ function that converts colour C from RGB to CIELAB, Eqs. (1)–(6)
$f_{\Delta \mathrm{E}}(C_1, C_2)$ function that calculates ΔE^*_{00} between C_1 and C_2, Eq. (17)
$\Delta \bar{E}^*_{00}$ overall average colour difference to their respective centroids

$\Delta \bar{E}^*_{00,c}$ average colour difference of cluster i to its centroid
N number of clusters
N_c number of elements in cluster c

Table 2 International commision on illumination (CIE) guidelines on interpreting ΔE^*_{00} values as defined by CIEDE2000 (Karma 2020)

ΔE^*_{00}	Perception
< 1	Imperceptible to human eyes
1–2	Perceptible through close observation
2–10	Perceptible at a glance
10–49	Colours are more similar than opposite
> 49	Colours are dissimilar
100	Exact opposite

The use of this approach has a similar consequence to the tolerancing method deployed by colour professionals. Because the judgement of ΔE^*_{00} is purely against the reference colour (in the case of this study, the centroid colour), two assessed objects that are within the acceptable range of ΔE^*_{00} to the reference may possibly have up to almost double the ΔE^*_{00} when compared to each other if they are on diametrically opposing sides of the reference colour within the CIELAB tolerancing bounds. However, since this is an acceptable norm in the interpretation of ΔE^*_{00} this consequence was assumed to be acceptable in this study.

2.5 Colour sorting setup optimization

2.5.1 Batch size selection

In this study, a batch is defined as the quantity of lumber that will be segregated into the desired number of bins. In practice, the bin configuration are to be re-evaluated for every batch of timber. However, since the dataset for each period (periods one to five, as per Table 1) was obtained sequentially, batch analysis was performed convolutionally for each period as separate evaluation sets.

The first step taken was to assess the most suitable parameter to locate the central position of the batch, and this is done by assessing the distribution that is most representative of the batch data. Two distributions were tested, Gaussian and beta. The normalities of the distributions for batch sizes 25, 50, 100, 200, 250, 500, 750, and 1000 were evaluated using KS, AD, and SW tests (at 5% significance level). To test their goodness of fit to a beta distribution, since the two-parameter beta distribution is defined for values from 0 to 1, L^*_u values were first normalised using Eq. (20) with the assumption that the L^*_u extremities are between 44 and 100 (approximately 10 levels of intensity below 54.2 and above 90.1, which are trimmed extremities in the dataset). MLE was then used to estimate the fitted probability density function to a beta distribution for each batch size, and subsequently equal numbers of samples to the batch size were randomly generated using this function (Monte Carlo method). These samples were then evaluated against the actual batch using the KS and AD tests (at 5% significance level) to determine their goodness of fit to a beta distribution.

$$\tilde{L}^*_u = \frac{L^*_u - 44}{56} \tag{20}$$

where
$\tilde{L}^*_u$ normalised L^*_u value

Another useful metric is range coverage, Q, or the proportion of which the average interquartile (IQR), mid 90th percentile (M90), mid 98th percentile (M98) and extremum (EX) ranges of each batch size straddle the L^*_u intensity range of the superpopulation (54.2 to 90.1), as expressed in Eq. (21). This provides a useful indication of the band in which the central majority of the samples in each batch are occupying.

$$Q = \frac{100}{(L^*_{u,max} - L^*_{u,min})} \cdot \frac{1}{n} \sum_{i=1}^{n} R_i\% \tag{21}$$

where
Q Range coverage, percentage of superpopulation's L^* intensity range
n total number of batches evaluated
$L^*_{u,max}$ the maximum L^*_u of the superpopulation, 90.1
$L^*_{u,min}$ the minimum L^*_u of the superpopulation, 54.2
R_i IQR, M90, M98 or EX range of batch at iteration i

Once these results were obtained, the batch size was then selected based on their distribution characteristics and practical considerations on the production floor.

2.5.2 Overclustering and bin configuration

Several different overclustering methods were tested to try to provide greater resolution to the colour range. By having more colour clusters, each cluster will invariably have a smaller average value of ΔE^*_{00}. Five methods of clustering the values of L^*_u were tested: (1) percentile, (2) fixed width, (3) Otsu multilevel image thresholding (Otsu multithresholding), (4) clustering using k-means on L^*_u values, and (5) clustering using k-means on all three L^*_u, upper a^*(a^*_u) and upper b^*(b^*_u) values.

Percentile methods were widely tested and used in previous research (Bianconi et al. 2013; Rožman et al. 2006; Defoirdt et al. 2012), including past iterations of KayuSort, therefore included in the analysis. The fixed width clustering was tested to see whether the splitting into equal clusters in the L^*_u range worked better. A different clustering method, Otsu multithresholding, was introduced for colour sorting to see if separation by minimising intercluster variance works better than the other methods. Figure 3 shows Otsu-clustered groups and selected wood sample images whose L^*_u values correspond close to the mid point of each cluster. Lastly, k-means clustering is another popular machine learning tool for unsupervised clustering. K-means used in this study used the squared Euclidean distance of points to their respective centroids as the performance optimisation parameter with the maximum number of iterations capped

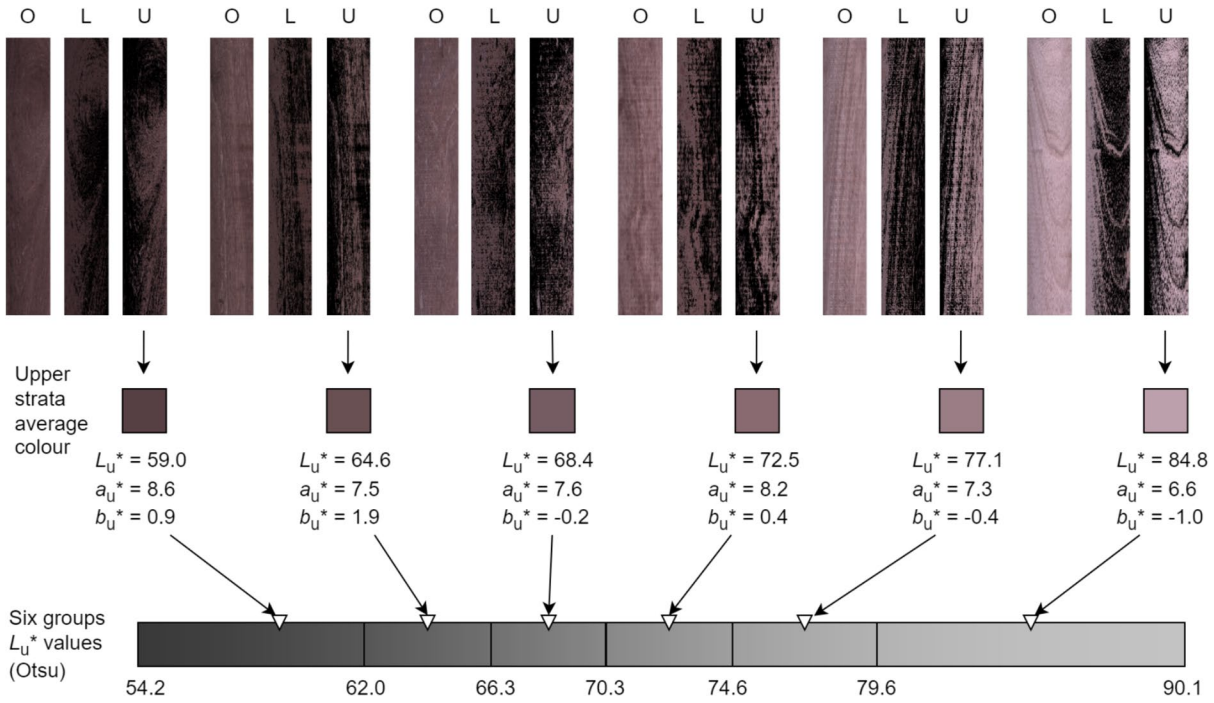

Fig. 3 Six selected wood sample images (labelled 'O') close to the midpoint of each cluster (clustered using Otsu multithresholding), their respective upper ('U') and lower ('L') stratum images after KayuSort's Otsu bisection, and their average upper L^*(L_u^*), a^*(a_u^*) and b^*(b_u^*) values; note that wood images were cropped to similar lengths for illustration purposes

at 1000. The numbers of overclusters tested ranged from 3 to 10 (KayuSort's current bin count limit) to gauge the efficacy of each clustering method. To evaluate these different clustering methods, the entire trimmed dataset was sorted into their respective clusters for each method using the L_u^* values, and their average ΔE_{00}^* values were calculated. The clustering method that scored the lowest overall ΔE_{00}^* was considered the winning method.

The average ΔE_{00}^* performance of the original KayuSort percentile-based clustering was also evaluated using the original dataset and the superpopulation dataset. This value was used as a performance baseline to compare the results of the different clustering methods used in this study.

Having determined the batch size and the winning overclustering method, the selected batch size was used to evaluate the average ΔE_{00}^* values for all combinations of overclustered colour clusters from three to nine for three colour sorting bins (CSB) and from four to ten for four-CSB. The target CSB was chosen by calculating (i) the statistical median L_u^* value of the batch, (ii) the MLE-G mean of the batch, (iii) the MLE-B mean, and (iv) the MLE-B median of the batch, and matching them to the closest midpoint of each CSB (measured by Euclidean distance). The midpoints of each CSB were determined by taking the average of the two extreme boundary values that spanned three or four bins. The batches were then sorted into their respective bins according to which set of CSB they fall into, and their overall average ΔE_{00}^* was calculated and tabulated. As in the previous steps, the configuration that scored the lowest overall average value of ΔE_{00}^* was considered the ideal bin configuration for the batch size used.

It should be noted that the average ΔE_{00}^* performance of the recently upgraded KayuSort algorithm is identical to the percentile overclustering of six clusters with four-CSB.

2.5.3 'Burn-in rate' recommendation

The final step in this study was to determine the minimum amount of observations or samples necessary to provide an acceptable prediction of the batch's colour distribution, which is also known as the 'burn-in rate'. By convention, sampling ratios of 30% or greater are required to adequately predict the characteristics of the population. Realistically, however, a 'burn-in rate' of 30% for a batch size of 250 pieces (which is 75 pieces) in the industry is highly impractical. Even going by the United States Defence Standard MIL-STD-105 (adopted in ASTM E2234), which recommends a sample size of 32 for batch sizes between 151 and 280, the 'burn-in rate' is still rather high. The purpose of this study was to recommend a lower acceptable 'burn-in rate', and the 'burn-in rates' of 3 to 30 samples (hereafter these samples are referred to as kernels for brevity) shall be tested for the three- and four-CSB configurations.

To predict the batch distribution, four methods were used to predict the CSBs to use, which are the statistical median, MLE-G mean, MLE-B mean, and MLE-B median L^*_u values of the kernels. These values were then used to select the CSB as per previous section for the entire batch, except now using values derived using the kernels. Then, these selections were compared with the actual selection for the entire batch to determine their accuracy rate, AR, as shown in Eq. (22).

$$\mathrm{AR} = \frac{\text{iterations with correct CSBs}}{\text{total iterations}} \times 100\% \tag{22}$$

Lastly, the average ΔE^*_{00} results for each kernel size were calculated for the method with the highest accuracy rate. With these average ΔE^*_{00} values, the percentage of instances when they exceed 2.0 was determined (2.0 benchmark was chosen in line with the standard acceptable value of ΔE^*_{00}). As described earlier, the total possible range of ΔE^*_{00} is 4.0 (2.0 around the centroid colour), which puts it in the lower half of the perception range of 2–10 (perceptible at a glance, Table 2). However, since the wood industry colour matching requirements are generally not as stringent as those in the automotive or printing industry, this average ΔE^*_{00} of 2.0 will serve as a useful benchmark for colour sorting performance.

3 Results and discussion

3.1 Batch size selection

The results of the normality test and the batch L^*_u span as a percentage of the superpopulation L^*_u span, Q, for batch sizes of 25–1000 are shown in Fig. 4. For the two normality tests (KS and AD tests), the acceptance rates of the null hypothesis H_0(shown in Fig. 4a with the null hypothesis H_0 being that the distribution is normal) showed a decrease in the acceptance rate of the null hypothesis as the batch size increased. However, the goodness of fit to the beta distribution remained relatively high for both the KS and AD tests (above 80%). This meant that a beta distribution was a good approximation for the distribution of the batch population at the various batch sizes tested. Also, according to Fig. 4b, the distribution appeared quite similar even as batch sizes increased, with the interquartile, the middle 90^{th} percentile and the middle 98^{th} percentile ranges slowly but steadily increasing.

The selection of the batch size is critical, as it determines how well the timber is separated into similar colours within a restricted number of colour bins. Too large a batch size would result in the batch's colour distribution approaching the superpopulation, which negates the efficacy of higher-resolution sorting, while too small a batch size would result in insufficient timber to produce the product with similar

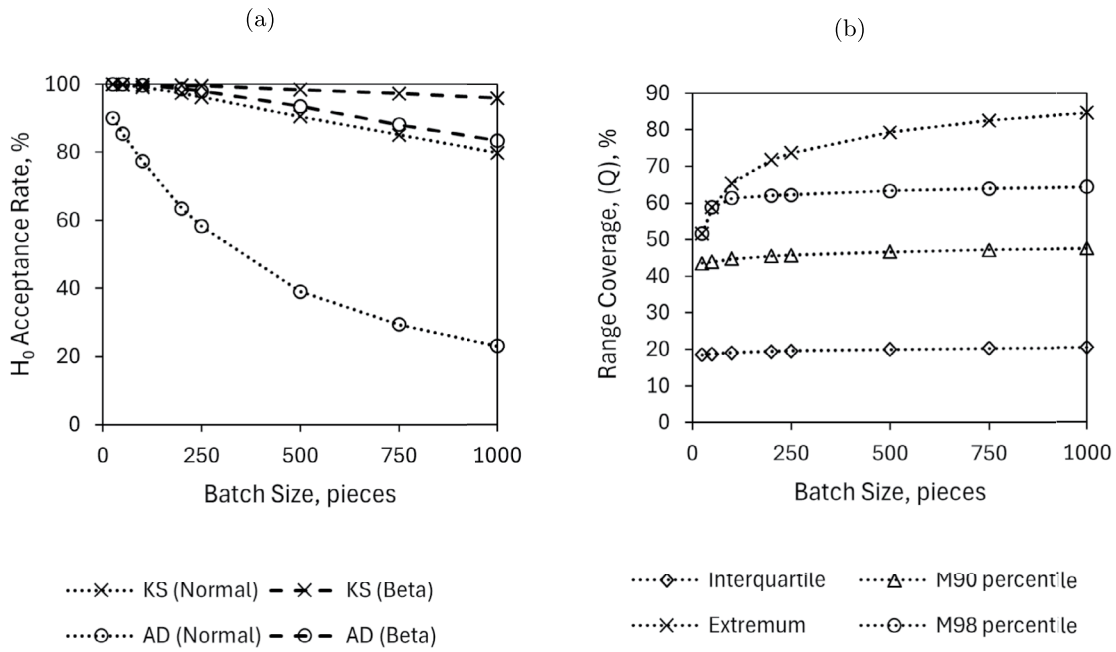

Fig. 4 **a** Rates of which normality test results for Kolmogorov-Smirnov (KS) and Anderson-Darling (AD), and the goodness to fit test results for KS and AD to a beta distribution showed that the batch L^*_u distribution was similar to a normal and beta distribution, tested for various batch sizes, convolved across superpopulation dataset; **b** Range Coverage, Q, per Eq. (21), which is the percentage of which the averages of all the batches' interquartile, mid 90^{th} percentile (M90), mid 98^{th} percentile (M98), and extremum ranges spans the L^*_u intensity range of the superpopulation

colour before the colour grouping shifts. As can be seen in Fig. 4, selection of batch sizes between 250 and 1000 appeared to be feasible. A drop in performance can be expected as the increased span for outliers increases, but shall be minimal. For this study, the batch size of 250 was selected because it represented an equivalent amount of wood to produce approximately four pieces of laminated scantling. The batch size of 1000 was also tested to see the difference in performance between the two batch sizes and assess the possibility of using the same setting for both batch sizes. For actual implementation, this number can be varied depending on the product requirement (such as required product size and input timber size), and the analysis presented here may be adjusted accordingly.

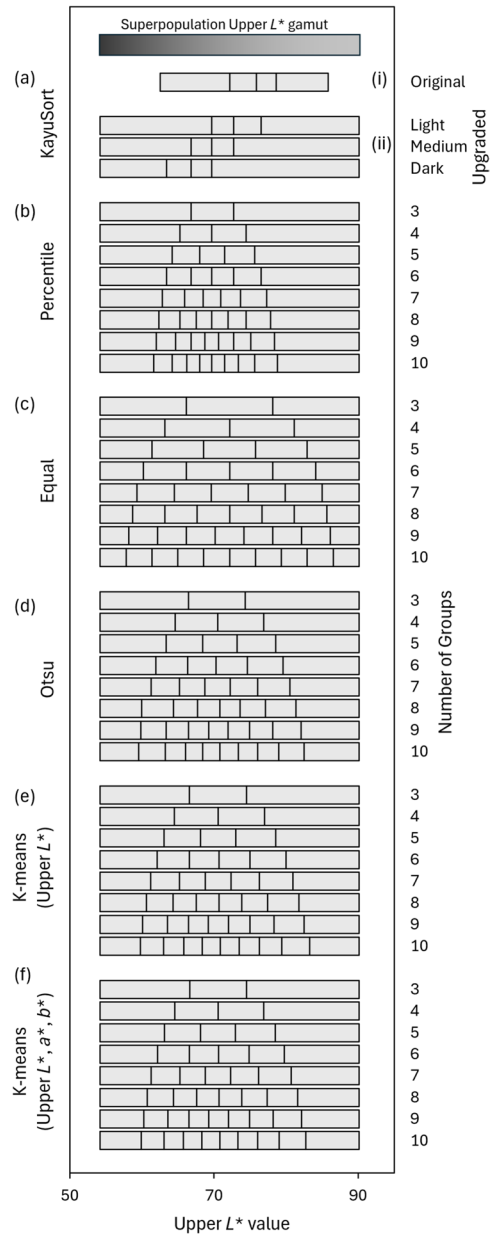

Fig. 5 Segmentation chart for **a** KayuSort's (i) original (percentile, original dataset) and (ii) the recently upgraded (percentile, superpopulation) upper L^*(L^*_u) clustering, the four methods used in this study to cluster the L^*_u of the superpopulation which are **b** percentile based, **c** equal widths, **d** Otsu multithresholding, and **e** k-means, and **f** k-means clustering of the superpopulation using all three upper L^*, a^* and b^* values (only the L^*_u values is shown)

3.2 Overclustering and bin configuration

The results of the clustering tests are shown in both Figs. 5 and 6a. As can be seen in Fig. 5, the original KayuSort colour gamut (trained using 392 samples) covered a substantial range of the central percentile of the superpopulation's colours, with an overall average ΔE^*_{00} of 1.288 (s.d. = 0.808). However, due to the limited sample size used, the samples were skewed toward lighter coloured woods, and therefore the darker coloured woods were not sufficiently resolved, which was evident when the superpopulation data were used, resulting in a new overall average ΔE^*_{00} of 2.296 (s.d. = 1.684). Furthermore, having only a rigid four colour cluster meant that increasing the range of each cluster would inevitably result in larger ΔE^*_{00} values, hence poorer colour similarities within each sorted cluster. In addition, it can be seen that the recently upgraded over-clustering method (albeit manual selection of bins) resulted in much better clustering performance, with an overall average ΔE^*_{00} of 1.318 (s.d. = 0.775) over the original.

The Otsu multithresholding method was the best method to perform clustering, scoring lower overall average ΔE^*_{00} values although its performance was very similar to that of k-means (both L^*_u and all three L^*_u, a^*_u and b^*_u methods). The performance of the percentile method was marginally lower, whereas the equal spacing method performed the worst. This meant that Otsu multithresholding managed to cluster the L^*_u colour efficiently, and the a^*_u and b^*_u components of the image had inconsequential effects on the results (since they were used in ΔE^*_{00} calculations, although the clustering was performed only on the L^*_u values alone), as anticipated. Furthermore, given the large data set used in this study with the assumption of a good representation of the colour gamut of LRM wood, clustering based on inter-cluster variance in the L^*_u values gives a truer representation of how these L^*_u values are clustered rather than the traditional variance methods.

Interestingly, k-means had a very similar clustering characteristic as Otsu multithresholding, with very similar threshold and ΔE^*_{00} values. Otsu's original paper (Otsu 1979) indicated that, in the context of images, as the number of classes increased, the selected thresholds will become less credible. However, in the context of this study, they behaved very similarly to k-means and provided good clustering performance, despite the fact that the data were not

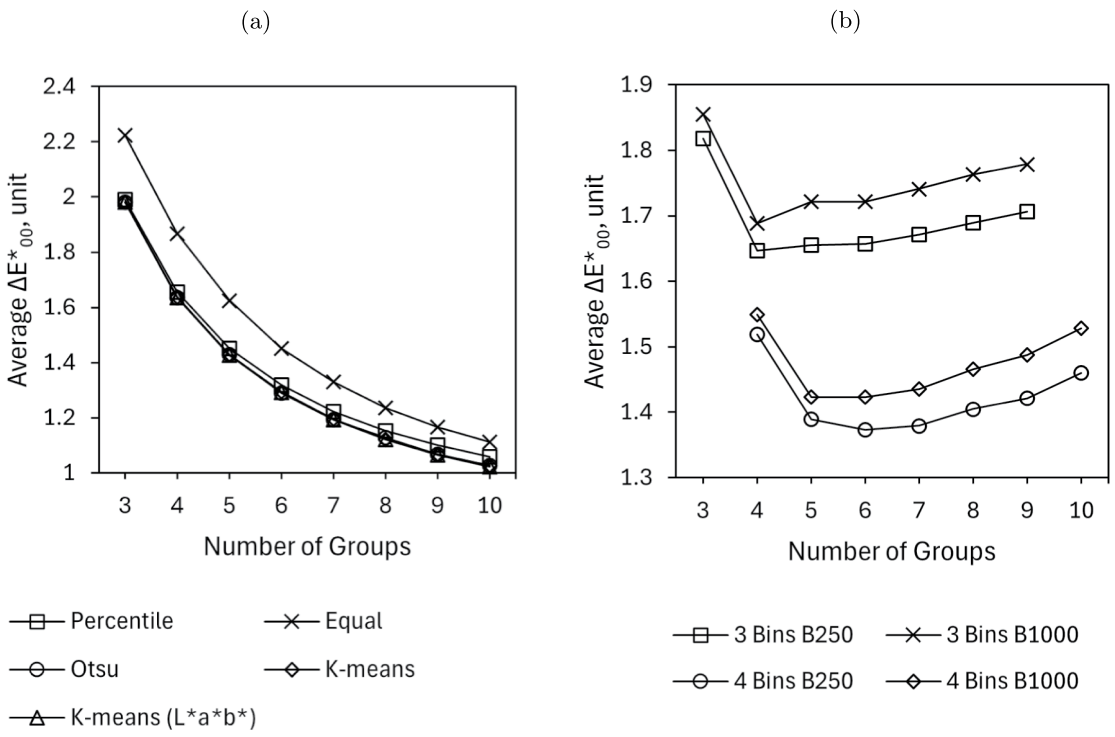

Fig. 6 **a** Average colour difference ($\Delta\bar{E}^*_{00}$) of the four methods (percentile, equal spacing, Otsu multithresholding, and k-means) of clustering the superpopulation dataset's upper L^* values into 3 to 10 clusters, as well as k-means clustering using all three upper L^*, a^* and b^* values, and **b** $\Delta\bar{E}^*_{00}$ for different three colour sorting bins (CSB) and four-CSB configurations for batch sizes of 250 (B250) and 1000 (B1000) samples, using maximum likelihood estimation beta distribution approximated median

visibly multimodal. In fact, Otsu multithresholding matched or slightly outperformed k-means using the three values L^*_u, a^*_u and b^*_u for seven clusters and less (except for five clusters). The performance of k-means was the best for eight clusters and beyond. This means that when there are more clusters available for sorting, the chromatic components can be used to improve sorting performance, but when the number of clusters is restricted, segmentation with the inclusion of chroma values actually degrades the performance slightly causing confusion at the thresholds. Although k-means in this study used the Euclidean distance as the optimisation parameter, which is identical to the older ΔE^*_{76}, Eq. (16), and the presence of adjustment factors and minor discontinuities in the newer ΔE^*_{00} computations (Sharma et al. 2005) may have contributed to this phenomenon. This results shows that comparable performance can be achieved by using only L^*_u values and Otsu multithresholding, compared to the use of all three L^*, a^* and b^* values and a slightly more sophisticated machine learning approach. Although k-means was found to be more computationally efficient for multithresholding applications compared to Otsu (Liu and Yu 2009), in this method, clustering is performed on the superpopulation only during the setup phase, therefore, inconsequential to the colour sorting speed.

This is in agreement with previous studies whereby k-means performed better for larger numbers of CSG (Lin et al. 2020a; Wang et al. 2021). Likewise, when the number of CSG is small (K being between 3 and 7), direct colour sorting without oversampling resulted in visibly poorly matched timber. The optimal number of bins for Wang et al. (2021) was 8 for a dataset of 1800 images from 200 samples of beech panels, while it was 16 bins for Lin et al. (2020a) based on 15,000 images. Furthermore, the CIELAB distribution of the superpopulation in this study (Fig. 2) corresponded closely to the observations for *Pometia pinnata* (Zhuang et al. 2022) where the L^* components of the wood colour had a larger spread (about 8 times) compared to the chroma components. This indicates that this approach to clustering may also work for other species. It must be noted that HSV colour space is not ideal as features for k-means analysis since it is a cylindrical coordinate system with the hue (H) and saturation (S) components being polar coordinates (with the assumption that k-means used Euclidean distance in calculating their centroids). Also, Zhuang et al. (2022) asserted that chromatic components cannot be ignored during colour sorting; this is only true when there are adequate bins available for colour sorting. In fact, their plot actually hinted at very good separability characteristics

using the luminance component alone at low bin counts, consistent with our findings.

Having determined that the Otsu multithresholding performed well, particularly for fewer clusters, each configuration was tested for performance when three- and four-CSBs were selected from up to 10 total clusters. All approximation methods tested performed very similarly (one-way ANOVA test p-values > 0.05) with the MLE-B median scoring the lowest ΔE^*_{00} values overall. This is consistent with the results in Fig. 4a, which showed the highest goodness of fit when tested using the KS and AD tests. Therefore, MLE-B median's results were reported here in Fig. 6b and used for the 'burn-in rate' tests. For the batch size of 250 pieces, it is clear from Fig. 6b that the overall average ΔE^*_{00} value reached their minimum ($\Delta E^*_{00} = 1.373$) at six colour overclusters for four-CSB (referred to as 6C4), four colour overclusters for three-CSB (referred to as 4C3) with ΔE^*_{00} of 1.646. For the batch size of 1000 pieces, five colour overclusters for four-CSB (referred to as 5C4) performed only marginally better than 6C4 for four-CSB ($\Delta E^*_{00} = 1.423$), while for three-CSB, 4C3 performed the best ($\Delta E^*_{00} = 1.688$).

This similar method can be used to obtain the ideal number of colour overclusters for any desired batch size and CSBs. Batch sizes between 250 and 1000 can use the 6C4 configuration when four sorting bins are desired, while the 4C3 configuration may be used when three sorting bins are to be used. There is a small increase in the average ΔE^*_{00} values as the batch size increased from 250 to 1000 (3.97% for four-CSB and 2.85% for three-CSB). The results also affirmed the choice made during the KayuSort upgrade to use six overclusters for four-CSB, although the rationale for that choice was purely to provide three choices of wood brightness for the user to choose from.

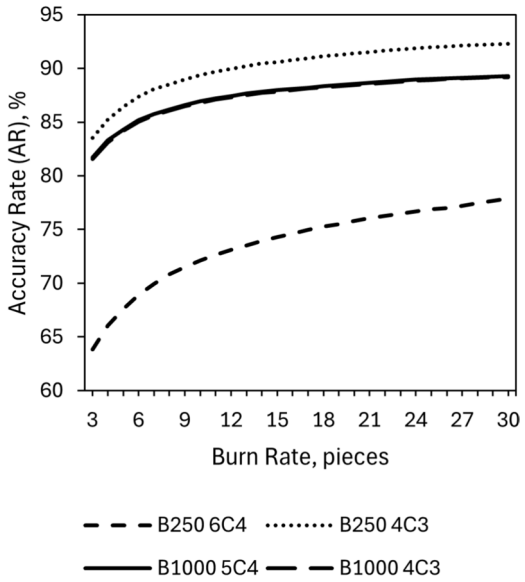

Fig. 7 Accuracy rates, AR, per Eq. (22), for 'burn-in rates' from 3 to 30 for four-bin (6C4 for batch size of 250 pieces, B250, and 5C4 for 1000 pieces, B1000) and three-bin configurations (4C3 for both batch sizes B250 and B1000) using maximum likelihood estimation beta distribution approximated median

This stage is analogous to searching for the optimal bin breadths that minimise the colour difference of all of the constituents in the bin. As can be seen from the results, as the number of desired bins increases, the optimal number of overclusters increased, with overcluster to CSB ratios of around 4:3 and 6:4. This conclusion depends on the batch size, and as the batch size increases, there is a greater possibility of obtaining a larger variety of wood of different species or provenance and therefore a wider colour distribution—this can be clearly seen in Fig. 6b where the values ΔE^*_{00} for the batch size of 1000 were much higher than those for the batch size of 250. Furthermore, the performance of 6C4 and 5C4 for the batch size of 1000 was very similar, with 5C4 slightly outperforming 6C4. It can be projected that this trend of reducing overclusters will continue as the batch size increases further, whereas conversely, as the batch size decreases, the optimal number of overclusters increases.

3.3 'Burn-in rate' recommendation

The MLE-B median approximation slightly outperformed the others in most of the results. However, it should be noted that a one-way ANOVA revealed that there were no statistically significant differences between the performances of MLE-G mean, MLE-B mean, and MLE-B median (post-hoc tests p-values > 0.95). The statistical median method performed the worst among the methods tested, with a statistically significant difference from the others (post-hoc tests p-values < 0.01, Bonferroni-adjusted significance level of 1.25%). Figure 7 shows the accuracy of each of the bin configurations using MLE-B median. For four-CSB with a 'burn-in rate' of 3 and batch size of 250 pieces, 6C4 only managed an accuracy rate of 63.9%, which improved to 77.9% with a 'burn-in rate' of 30. For a batch size of 1000, 5C4 managed an accuracy rate of 81.7% and improved to 89.3%. For three CSB, batch sizes of 250 and 1000 (both of them were 4C3 configuration), began at the 'burn-in rate' of 3 with accuracy rates of 83.5% and 81.6%, respectively, and reached 92.3% and 89.2% at the 'burn-in rate' of 30 pieces.

However, Fig. 8 gives a much clearer context to the performance results in Fig. 7, showing the actual average ΔE^*_{00} performance of each 'burn-in rate'. From Fig. 8b, the 'burn-in rate' of around 12 pieces appeared to be the ideal recommended 'burn-in rate' for four bins and batch size 250 (using the 6C4 configuration) as the percentage of batches with their average $\Delta E^*_{00} > 2.0$ units dipped below 1%. For a batch size of 1000 pieces using the 5C4 configuration, the 1% value was crossed at the 'burn-in rate' of 10 pieces. For

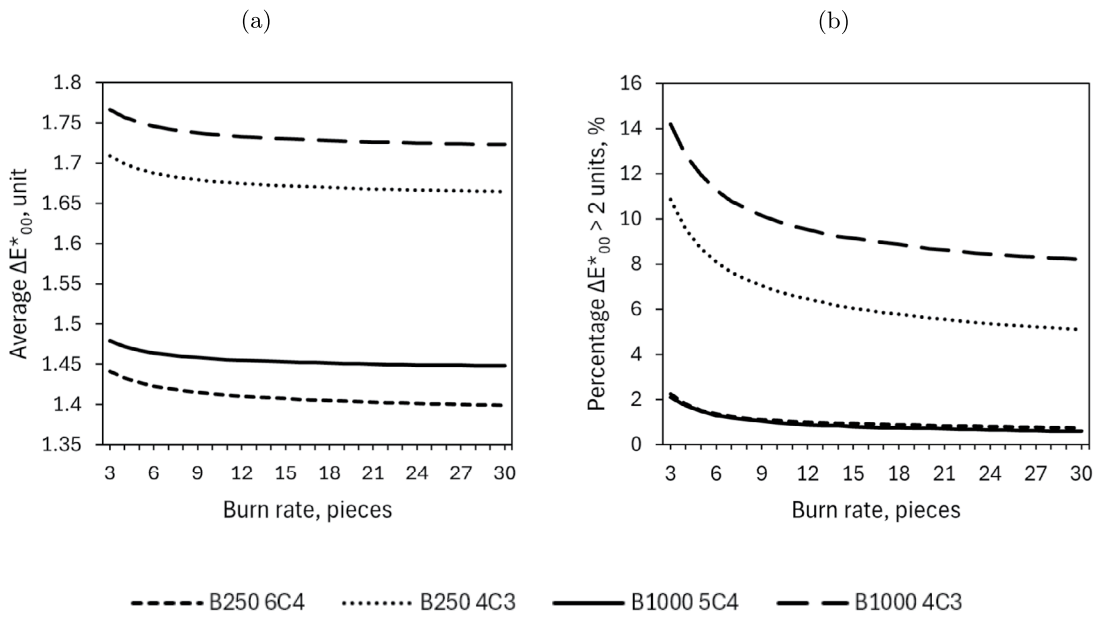

Fig. 8 **a** Average ΔE^*_{00} performance for different 'burn-in rates' for batch sizes 250 (B250) and 1000 (B1000) pieces for 6C4, 5C4 and 4C3 optimal configurations, and **b** the percentage of batches that had their ΔE^*_{00} values larger than 2 units, using maximum likelihood estimation beta distribution approximated median

three bins, while both options had an overall average $\Delta E^*_{00} < 2.0$, the percentages of batches that had $\Delta E^*_{00} > 2.0$ were substantially higher, which was expected due to the larger band of L^*_u values per cluster. For batch sizes of 250 and 1000 pieces (both using the 4C3 configuration), the 'burn-in rate' of 10 pieces saw percentages of $\Delta E^*_{00} > 2.0$ crossing below the levels of 7% and 10%, respectively. The drop in percentages thereafter appeared to be too gradual, making the increasing 'burn-in rate' impractical for industrial applications.

Furthermore, for four bins, if slightly higher mismatched colours can be tolerated (batch average $\Delta E^*_{00} > 2.0$ occurring around 2% of the time), a sampling size of three is sufficient to obtain an overall average ΔE^*_{00} of 1.45. For batch sizes between 250 and 1000, we expect the average ΔE^*_{00} to be confined between the two lines in Fig. 8a, with the 1% intersect (probability of $\Delta E^*_{00} > 2.0$) between the 'burn-in rates' of 10 and 12 pieces.

From these results, it is clear that there is substantial improvement over the original KayuSort setup when exposed to the superpopulation dataset (the original average ΔE^*_{00} was 2.296). Having improved the average ΔE^*_{00} to less than 1.8 in all tested configurations, this method had achieved ΔE^*_{00} under the benchmark level of 2.0. This highlights the advantage of knowing the breadth of wood colour and its distribution, especially for species with a wide array of colours. This is especially pertinent if the number of bins desired is restricted. Not all manufacturing facilities can accommodate colour sorting lines with 8, 16 or more bins; therefore, this study provides a pathway to improve colour similarity while maintaining a small number of bins. However, the methodology used in this study can be extended to a larger number of desired bins. Chromatic features can also be included by using k-means instead of Otsu multithresholding. The texture of the wood was not considered in this study due to the limited number of bins, but it can also be parameterised and included in the clustering and performance scoring processes. Deep learning techniques can be used to characterise and classify the grain pattern, such as the approach used by Liu et al. (2020) for rubberwood (*Hevea brasiliensis*). However, Liu et al. (2020) focused on mainly classifying grain patterns while separating their data into only two colour groups (dark and light); other industry experts had previously determined that 10 colour bins were ideal (Kurdthongmee 2008). This disparity on what constitutes good colour clustering illustrates the benefit of the flexible approach proposed by this research, which can account for different mill preferences for the number of bins.

The KayuSort trained SOM map for a six-overcluster setup using Otsu multithresholding is shown in Fig. 9. The results and insights gained from this study shall be incorporated into the next major software upgrade cycle of KayuSort. It should be noted that although KayuSort is currently used with industrial cameras, a study carried out (Liew et al. 2023) had shown that a well-calibrated high-quality webcam can achieve comparable performance, which will allow affordable implementations of this colour sorting solution in the future. Lastly, to address the issue of requiring long-term data on the total gamut of wood colour when training for a new species, an algorithm that retraces the methods used

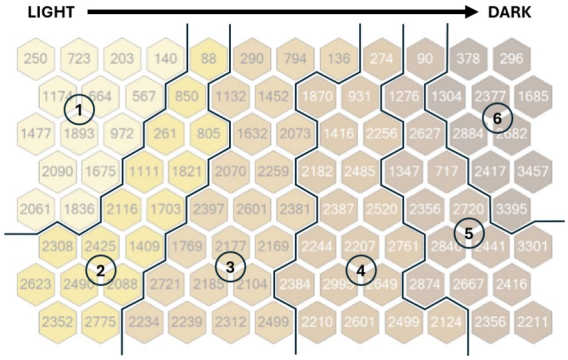

Fig. 9 KayuSort's self-organising map for six overclusters trained using the superpopulation images labelled using results from this study; cluster 1 being the lightest and cluster 6, the darkest

in this study shall be created to automatically update the training of the colour sorting map periodically as the system encounters new batches of timber with a different hue.

In this study, several shortcomings are recognised. Firstly, the erroneous selection of CSB was likely due to changes in the actual batch of timber mid-way of each period where there was a significant difference in their colour, hence resulting in erroneous predictions. This can be partially overcome by allowing the worker to reset the system's prediction when there is an actual known change in the batch, so that the predicted CSB of the previous batch does not get carried over to the new batch. Secondly, the two edge bins will invariably contain all the outliers, potentially causing larger colour variations. One possible solution would be to eject these outliers to a separate bin to be sorted again with a different batch. However, this means having to scan the wood multiple times, affecting production efficiency.

Future studies shall focus on the use of continuously sliding bins to better capture the nuances in the shift of wood colour rather than fixed clusters. In addition, other local species with wide colour ranges, such as kembang semangkuk (*Scaphium* spp.) and merpauh (*Swintonia* spp.) shall be researched. This is largely due to their undifferentiated sapwood, where there is a gradual transition in colour between the sapwood and heartwood regions. In this case, clustering has to be performed in both the intensity and chromatic domains to improve performance, and there may be a need for more bins than three or four. Another area of focus shall be the inclusion of the texture of wood into the sorting algorithm, which will be important in the sorting of species with prominent grain features such as oak (*Quercus* spp.).

4 Conclusion

This study showed the roadmap towards improving the performance of timber colour sorting when there is a large variance in the colour and the binning requirements are restricted. Clustering upper L^* values using Otsu multi-thresholding performed well, with performances very similar to that of k-means clustering using all upper L^*, a^* and b^* values. For the batch size of 250 pieces, six and four overclusters produced the lowest average ΔE^*_{00} for four (6C4) and three (4C3) bins, respectively, while for the batch size of 1000 pieces, five (5C4) and four (4C3) overclusters performed the best. Maximum likelihood estimation using beta approximation's median appeared to match the distribution of the entire dataset and each batch the best, and was used to predict the centrality of the batch when selecting the ideal bin configurations. The recommended 'burn-in rate' for 6C4 with the batch size of 250 pieces is 12 pieces, while all other configurations worked best with 10 pieces. Further research shall be conducted for other species with wide ranges of colours, especially those with undifferentiated sapwood.

Acknowledgements The authors extend their thanks to Chris Tan and Danny Tan of Sim Seng Huat Timber Industries Sdn. Bhd. (SSH) for graciously providing unfettered access to their facility to conduct this research and their unwavering support. Many thanks also to the Malaysian Timber Industry Board (MTIB) for the use of the wood colour sorting prototype located at SSH.

Author contributions All authors participated in the funding and design of this study. Tan, Ng, Fauthan, and Liew collected the data. Tan, Liew, and Ng processed the data and prepared the main manuscript text. All authors reviewed the manuscript.

Funding Open Access funding provided by The University of Osaka. This research was funded by the Japan Society for the Promotion of Science KAKENHI (research grant: JP23K20037), and internal funds from SIRIM Machinery Technology Centre and Hexagon World Sdn. Bhd. The wood samples used in this study were provided by SSH.

Data availability Data is available on request.

Declarations

Conflict of interest The authors declare no competing interests.

Generative AI No generative AI tools were used in any part of the production of this research paper.

References

Ashton PS, Heckenhauer J (2022) Tribe Shoreae (dipterocarpaceae subfamily Dipterocarpoideae) finally dissected. Kew Bull 77(4):885–903. https://doi.org/10.1007/s12225-022-10057-w

Beckman RJ, Tietjen GL (1978) Maximum likelihood estimation for the beta distribution. J Stat Comput Simul 7(3–4):253–258. https://doi.org/10.1080/00949657808810232

BenQ Corporation (2014) Definition of 'accurate colors' on a photographer monitor. Accessed 14 Mar 2025. https://www.benq.com/en-us/knowledge-center/knowledge/definition-of-accurate-colors-on-a-photographer-monitor.html

Bianconi F, Fernández A, González E, Saetta SA (2013) Performance analysis of colour descriptors for parquet sorting. Exp Syst Appl 40(5):1636–1644. https://doi.org/10.1016/j.eswa.2012.09.007

Cha JK, Park H, Kwon Y, Lee SM, Kim J, Kim WJ, Lee JH (2025) Enhancing wheat quality through color sorting: a novel approach for classifying kernels based on vitreousness. Front Plant Sci 16:1. https://doi.org/10.3389/fpls.2025.1534621

Defoirdt N, Wuijtens I, Boever LD, Coppens H, Bulcke JVD, Acker JV (2012) A colour assessment methodology for oak wood. Ann For Sci 69:939–946. https://doi.org/10.1007/s.3595-0.2-02.4-3

Gan KS, Lim SC (2004) Common commercial timbers of Peninsular Malaysia. Forest Research Institute Malaysia, Kuala Lumpur

Gnanadesikan R, Pinkham RS, Hughes LP (1967) Maximum likelihood estimation of the parameters of the beta distribution from smallest order statistics. Technometrics, 9(4), 607–620, Accessed 12 Feb 2025. https://www.stat.cmu.edu/technometrics/59-69/VOL-09-04/v0904607.pdf

Høibø O, Nyrud AQ (2010) Consumer perception of wood surfaces: the relationship between stated preferences and visual homogeneity. J Wood Sci 56(4):276–283. https://doi.org/10.1007/s10086-009-1104-7

Huda M (2017) *Tips for defining a realistic pass/fail tolerance*. https://www.xrite.com/blog/tips-to-define-tolerances

International Color Consortium (2004) Image technology colour management-Architecture, profile format, and data structure (Tech Rep.). Accessed 31 Oct 2024. http://www.color.org

International Organization for Standardization (2017) ISO 13655:2017-Graphic technology-Spectral measurement and colorimetric computation for graphic arts images (Tech Rep.). Geneva:International Organization for Standardization

Johnson NL, Kotz S, Balakrishnan N (1995) Beta distributions. Continuous univariate distributions (2nd ed, vol 2, chap 210). John Wiley & Sons, Inc

Karma IGM (2020) Determination and measurement of color dissimilarity. Int J Eng Emerg Technol 5(1):67–71. https://doi.org/10.24843/IJEET.2020.v05.i01.p13

Kerman J (2011) A closed-form approximation for the median of the beta distribution. arXiv:1111.0433

Kew Gardens (2024) Dipterocarpaceae-Rubroshorea (Meijer) P. S. Ashton & J. Heck. Royal Botanic Gardens. https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:77298305-1#publications. 26 Mar 2025

Kurdthongmee W (2008) Colour classification of rubberwood boards for fingerjoint manufacturing using a SOM neural network and image processing. Comput Electr Agric 64(2):85–92. https://doi.org/10.1016/j.compag.2008.04.002

Liew SJ (2024) Development of computer wood veneer colour sorting system for wood industry (Doctoral dissertation, Universiti Malaya, Kuala Lumpur). Accessed 11 Dec 2024. Retrieved from https://www.proquest.com/dissertations-theses/development-computerized-wood-veneer-colour/docview/3143980066/se-2?accountid=16714

Liew SJ, Ng SC, Mustapa MZ, Usop Z, Fauthan MA, Mahalil K, Tan CO (2023) Colour sorting of red oak, yellow poplar and maple veneers using self-organizing map: comparisons between different camera genres. Eur J Wood Wood Prod 81:777–789. https://doi.org/10.1007/s00107-022-01900-9

Light Illusion Ltd (2025) Chasing Delta-E, grey scale and primary colours for perfect calibration. https://lightillusion.com/chasing_delta-e.html

Lim SC, Nordahlia AS, Abd Latif M, Gan KS, Rahim S (2016) Identification and properties of Malaysian timbers. Forest Research Institute Malaysia

Lin Y, Chen D, Liang S, Qiu Y, Xu Z, Zhang J, Liu X (2020a) Wood color classification based on color spatial features and k-means algorithm. In: IECON Proceedings (Industrial Electronics Conference) 3847–3851. https://doi.org/10.1109/IECON43393.2020.9255104

Lin Y, Chen D, Liang S, Xu Z, Qiu Y, Zhang J, Liu X (2020b) Color classification of wooden boards based on machine vision and the clustering algorithm. Appl Sci 10(19):6816. https://doi.org/10.3390/app10196816

Liu D, Yu J (2009) Otsu method and K-means. In: 9th International conference on hybrid intelligent systems 1:344–349. https://doi.org/10.1109/HIS.2009.74

Liu S, Jiang W, Wu L, Wen H, Liu M, Wang Y (2020) Real-time classification of rubber wood boards using an SSR-based CNN. IEEE Trans Instrum Meas 69(11):8725–8734. https://doi.org/10.1109/TIM.2020.3001370

Lu Q, Conners RW, Kline DE, Araman PA (1997) Real-time algorithm for color sorting edge-glued panel parts. IEEE Int Conf Image Process 1:822–825. https://doi.org/10.1109/icip.1997.648090

Luo MR, Cui G, Rigg B (2001) The development of the CIE 2000 colour-difference formula: CIEDE2000. Color Res Appl 26(5):340–350. https://doi.org/10.1002/col.1049

MathWorks Incorporated (2024) MATLAB R2024b Update 1. MathWorks

Meenu M, Choudhary P, Kaushal P, Verma N, Yadav YK, Kalra S, Xu B (2024) New insights into recent trends and emerging technologies in apple sorting. ACS Food Sci Technol 4(2):290–303. https://doi.org/10.1021/acsfoodscitech.3c00552

Meints T, Teischinger A, Stingl R, Hansmann C (2017) Wood colour of central European wood species: CIELAB characterisation and colour intensification. Eur J Wood Wood Prod 75(4):499–509. https://doi.org/10.1007/s00107-016-1108-0

Microsoft Corporation (2022) Visual Studio Community 2022. Microsoft Corporation

Mokrzycki W, Tatol M (2011) Color difference Delta E: a survey. Mach Graph Vis 4(20):383–411 (). Accessed 6 Jan 2025 https://www.researchgate.net/publication/236023905

Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 20(1):62–66. https://doi.org/10.1109/TSMC.1979.4310076

Ranjan R, Shroff H, Sharrer K, Tsukuda S, Good C (2024) FilletCam AI: a handheld tool for precise fillet color profiling of Atlantic salmon and rainbow trout. J Agric Food Res 18:101461. https://doi.org/10.1016/j.jafr.2024.101461

Rožman D, Brezak M, Petrovic I (2006) Parquet sorting and grading based on color and texture analyses. IEEE Int Symp Ind Electr 1:655–660. https://doi.org/10.1109/ISIE.2006.295538

Savolainen J (2023) Matching method for mutated veneer sheet images using gray-level co-occurrence matrix features. Eur J

Wood Wood Prod 81(4):1021–1031. https://doi.org/10.1007/s00107-023-01946-3

Sharma G, Wu W, Dalal EN (2005) The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. Color Res Appl 30(1):21–30. https://doi.org/10.1002/col.20070

Talib Othman K (2022) Using the nonparametric methods to estimate parameters of the beta distribution. J Math Prob Equ Stat 3(2):78–84 (). Accessed 12 Feb 2025 http://www.mathematicaljournal.com

Tan CO, Ogata S, Ng SC, Yap HJ, Usop Z, Fauthan MA, Liew SJ (2025) Machine learning based density estimation of light red meranti (*Shorea* spp.): a segmented approach to multiple regression of self-organising maps colour clusters using custom made 'KayuSort' colour sorting software. Eur J Wood Wood Prod 83(1), 28, https://doi.org/10.1007/s00107-024-02188-7

Torniainen P, Jones D, Sandberg D (2021a) Colour as a quality indicator for industrially manufactured ThermoWood®. Wood Mater Sci Eng 16(4):287–289. https://doi.org/10.1080/17480272.2021.1958920

Torniainen P, Popescu CM, Jones D, Scharf A, Sandberg D (2021b) Correlation of studies between colour, structure and mechanical properties of commercially produced ThermoWood® treated Norway spruce and Scots pine. Forests 12:1165. https://doi.org/10.3390/f12091165

ViewSonic Corporation (2021) What is Delta E? And why is it important for color accuracy? Accessed 14 Mar 2025 https://www.viewsonic.com/library/creative-work/what-is-delta-e-and-why-is-it-important-for-color-accuracy/

Wang Z, Zhuang Z, Liu Y, Ding F, Tang M (2021) Color classification and texture recognition system of solid wood panels. Forests 12(9):1154. https://doi.org/10.3390/f12091154

Zhuang Z, Liu Y, Ding F, Wang Z (2021) Online color classification system of solid wood flooring based on characteristic features. Sensors (Switzerland) 21(2):1–13. https://doi.org/10.3390/s21020336

Zhuang Z, Liu Y, Yang Y, Shen Y, Gou B (2022) Color regression and sorting system of solid wood floor. Forests 13:1454. https://doi.org/10.3390/f13091454