Bioresource Technology Reports 23 (2023) 101539

Contents lists available at ScienceDirect

Bioresource Technology Reports

journal homepage: www.sciencedirect.com/journal/bioresource-technology-reports

Principal component analysis revealed the key influencing factors of kombucha bacterial cellulose yield and productivity

Paramasivan Balasubramanian *, Prachi Tanya Praharaj

Bio Energy and Environment (BEE) Laboratory, Department of Biotechnology & Medical Engineering, National Institute of Technology Rourkela, 769 008, India

ARTICLE INFO

Keywords:
Kombucha
Bacterial cellulose
Fermentation
SCOBY
Principal component analysis

ABSTRACT

A long-term approach for turning waste into useful products might aid in sustaining the commercialization of kombucha tea by-products. Using principal component analysis (PCA), this study sought to investigate how eleven variables affect the yield of kombucha bacterial cellulose. The input variables that were included types of tea leaves and sugar concentrations, SCOBY (symbiotic culture of bacteria and yeast) mass, kombucha starter tea, temperature, pH, and fermentation time. PCA found 75 % of data variability in the first four principal components (PCs). The correlations between the input variables were examined using the correlation centroid plot. Biplot combines the optimal loading PCA plot with the correlation centroid that revealed the high concentration of sugar contributes to high yield of bacterial cellulose. The findings provided insight into the condensed criteria as a vital hint for both academia and industry in the pursuit of a high cellulose yield in kombucha.

1. Introduction

A wide group of bacteria could synthesize bacterial cellulose (BC) to defend against ultraviolet (UV) radiation or drift at the air-liquid interface to collect oxygen, which attracts the central attention of researchers recently. Unlike plant cellulose purification that needs harsh chemicals, BC is pure, low-energy, and free of hemicelluloses, lignin, and pectin (Huang et al., 2014; Behera et al., 2022). Siró and Plackett (2010) obtained up to 85 % crystallinity indexes from spatial fibril organization. BC has a higher specific area (Sulaeva et al., 2015), stronger water holding potential and longer drying time than plant-cellulose fibres (Meftahi et al., 2010). BC exhibits great application potential in various sectors such as food, packaging, personal care, home chemicals, biomedicine, textiles, composite materials, and other sectors (Singhania et al., 2022). However, BC has been recognized as generally recognized as safe (GRAS) by Food and Drug Administration (FDA) (Shi et al., 2014). BC's indigestibility makes it ideal for dietetic cuisine (Mohite and Patil, 2014). BC's second important application is in personal care and household chemicals (Bianchet et al., 2020). Face masks comprised with BC contain more water and feel cool and smooth due to their nanoscale 3D reticulated network (Pacheco et al., 2018). BC has promising biological uses in wound dressing, dental implants, artificial skin, hemostatic materials, drug delivery, vascular grafts, tissue engineering scaffolds, and biosensors (Carvalho et al., 2019). BC is biocompatible and pure enough for biomedical uses. BC-based wound dressings, drug delivery, vascular grafts, tympanic membrane replacement, and contact lenses are being marketed (Coelho et al., 2019). BC is used to deliver anticancer drugs and regulate drug release in several biomedical applications (Dey et al., 2023).

Interest in maximizing kombucha SCOBY (symbiotic culture of bacteria and yeast) production has increased in response to rising demand for BC and the potential for increased profits. It will be advantageous to optimize the fermentation process in order to achieve a higher production of SCOBY despite the limited availability of substrates. It has been demonstrated that the SCOBY yield can be affected by a variety of factors, including the type and quantity of substrate used, the length of fermentation, the initial pH, the incubation temperature, and the dimensions of the culture vessel (Abd El-Salam, 2012; Treviño-Garza et al., 2020). Traditional statistical modelling for the purpose of optimization has, up to this point, included carrying out trial runs with a number of different manufacturing configurations, which consumes both time and resources. Few experimental investigations have been carried out in the field of kombucha fermentation in order to find the essential process parameters that must be satisfied in order to achieve the highest possible SCOBY yield. The results of previous initiatives that were comparable have not yet been sufficient to provide information regarding the primary elements that influence the yield of SCOBY. Therefore, it is necessary to devise a strategy for reducing the major variables that

* Corresponding author.
E-mail address: biobala@nitrkl.ac.in (P. Balasubramanian).

https://doi.org/10.1016/j.biteb.2023.101539
Received 24 May 2023; Received in revised form 2 July 2023; Accepted 3 July 2023
Available online 4 July 2023
2589-014X/

determine SCOBY yield into meaningful contributing factors based on the existing experimental datasets. A chemometric tool called principal component analysis (PCA) can thus give a visual representation of the correlations and variances present in the datasets under study. These tools can also provide vital insight into the primary characteristics of SCOBY yield.

PCA is a method that can reduce the number of dimensions of a dataset while maintaining its heterogeneity and identifying the relationships between the variables that were entered. These not only help in the identification of a pattern in the dataset that is the subject of the inquiry, but they also contribute to the generation of new latent variables that more accurately describe the dataset. It has already been done with chemometric analysis for evaluating changes in plant xylem following exposure to increasing CO_2 (Kim et al., 2015), biomass characterization and classification as a potential feedstock for production of biodiesel (Škrbić et al., 2015), distinguishing between wood pellet residue based on near infrared spectroscopy (Mancini et al., 2018), and catalyst-based total oxidation of propane (Škrbić et al., 2018).

It is important to highlight that no prior research has used PCA to produce the correlations and criteria that can be used to produce high SCOBY yields in laboratories and scale them up to meet industry demands. Therefore, the present study employed PCA for comprehending the kombucha bacterial cellulose production to delineate the key significant factors affecting its yield. PCA was performed with the XLSTAT® software in order to identify the parameters that had the greatest impact on SCOBY yield and, as a direct result of this, to determine the correlations that existed between each of these parameters. Therefore, scientists and industrialists can direct their attention to the correlations and key influencing components affecting SCOBY yield that have been mentioned above in order to achieve higher yields and improvised properties.

2. Materials and methods

2.1. Data collection on SCOBY BC yield and pre-processing

A whole of 323 datasets were retrieved from the graphs, tables, and supporting supplementary information from 30 articles of kombucha fermentation literature that had previously been published to conduct the PCA. Input factors for the process were taken into consideration including tea leaf mass, sugar concentration, SCOBY inoculum mass, and kombucha starting tea volume. In addition, the fermentation medium's pH, duration, and temperature were considered as operating parameters. Four different types of tea extracts (green tea, black tea, black + green tea, and waste tea) were employed as auxiliary variables in the dataset using one-hot encoding. The kombucha bacterial cellulose yield and productivity were the output metrics that were targeted. It is important to note that the cellulose yield and productivity for kombucha were determined for this study using a wet-weight basis because this is how most published literature data was available. However, in a few instances, when the yield was only reported in dry weight, it was converted to wet weight as outlined in Sharma and Bhardwaj (2019). In the initial exploratory data analysis (EDA), missing values were imputed using arithmetic mean. During this stage, a few outliers (extreme datasets) affecting model dependability and outcome were dropped from the investigation. Each data entry's yield, as a result, was divided into four categories: low (0–20 g/L), medium (20–40 g/L), high (40–80 g/L), and very high (>80 g/L). The EDA parameters additionally contained the count, mean, quartile values, standard deviation, and maximum and lowest values for each input and output parameter. For a result of n items (yield categories) selected by p variables (like tea leaf mass, sugar concentration, SCOBY inoculum mass, kombucha starter tea volume, pH, time, and temperature of the fermentation medium), the initial raw dataset was normalized (to correlate all input variables with different units on a single scale) using mean and standard deviation (Eq. (1)) to compare different variables on the same scale.

$$z = (value - mean)/standard\ deviation \tag{1}$$

where z is the standardized value of the data point under a given variable p.

2.2. PCA runs on the pre-processed datasets

Using the standardized variable values, a correlation matrix (Eq. (2)) was built as $p * p$ symmetric matrix with correlation values for all potential variable pairs.

$$R = \begin{bmatrix} 1 & r_{12} & r_{13} & \dots & r_{1p} \\ r_{21} & 1 & r_{23} & \dots & r_{2p} \\ r_{31} & r_{32} & 1 & \dots & r_{3p} \\ : & : & : & : & : \\ r_{p1} & r_{p2} & r_{p3} & \dots & 1 \end{bmatrix} \tag{2}$$

where R is the Pearson correlation coefficient between variable xj and xk as described in Eq. (3).

$$r_{jk} = \frac{S_{jk}}{S_j S_k} = \frac{\sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)\left(x_{ik} - \overline{x}_k\right)}{\sqrt{\sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)^2}\sqrt{\sum_{i=1}^{n}\left(x_{ik} - \overline{x}_k\right)^2}} \tag{3}$$

where S is the covariance between j th and kth variables out of total p variables.

This correlation matrix was used to determine the strength (eigen values) and direction (eigen vectors) of each parent variable in the dataset. Because each variable had a value and an impact on the newly created dimensions, they individually contributed to the result (yield). The top k values of the eigen vector corresponding to the largest of the k eigen values were picked in descending order. Principal components (PCs) are a single or mixed mixture of the parent variables and comprised all of the condensed and compressed information and variability of the full dataset, resulting in new dimensions. PCA emphasized the most important interdependencies and correlations between the variables when applied to a dataset, in addition to revealing the PCs. A total of 7 quantitative variables were analyzed for kombucha bacterial cellulose yield, which was further categorized into low, medium, high, and very high ranges. In addition, waste tea, mixed green and black tea, green tea and black tea types were employed as supplemental variables to better highlight the dataset profiles. These extra variables were given the numbers 0 and 1 so that they could be distinguished for data analysis. The standardized dataset was used to execute correlation type PCA utilizing five filter factors on the XLSTAT® 167 2021.3.1.1149 program. Since the study variables had many levels and dimensions, the correlation type PCA was performed.

2.3. Interpretation of PCA results

After pre-processing and normalizing the dataset, correlation type PCA was performed with five filter factors to remove noise. Standard statistical methods were examined and PCA results were interpreted as outlined in (Kapoor et al., 2023).

2.3.1. Selection of principal components (PCs)

When interpreting the results of PCA, the resultant PCs were evaluated using Kaiser's rule (Kumar, 2017). This rule stipulates that in order to preserve the most information from the parent p variables, only PCs with eigenvalues that are greater than one should be used in the PCA analysis. In general, PCs that had a high variability score were taken into consideration because it is anticipated that these PCs will effectively reflect the great majority of the variance present throughout the entire dataset.

2.3.2. Scree plot

The scree plot has long been used to graphically depict the cumulative variability and eigenvalues scores of PCs. In circumstances with acceptable dataset fit for PCA, the scree plot tends to exhibit a very apparent elbow-like turn from higher to lower PCs; this phenomenon is known as the "elbow effect" (Zhu, 2004). A more continuous scree plot suggests that the dataset contains a big number of PCs that are all equally important.

2.3.3. Eigenvectors, factor scores and contribution (%) of variables

The eigenvectors illustrate how much a PC's orientation is influenced by the parent input variable. The coordinates of the observed data points on the PCA dimensions under consideration correspond to the factor scores given by PCA findings. They are the hidden constructs that make the development of PCs possible through the use of variables. The significance of a component in determining a particular PC can be simply inferred from the percentage of that element's contribution to various PCs. This has been graphically depicted in order to highlight the graphical impact that each variable has on the various PCs.

2.3.4. Squared cosines of variables

Squared cosine values (SCV) are used to illustrate the graphical strength of given variables on PCA axes. This is done to assist in preventing interpretational errors that may occur during the process of result analysis. Working with the square of cosine and 1-cosine creates a result of linked variance as distinguished to unconnected variance, which enables for a better comprehension of the variance of the factors when dealing with a set of independent variables, such as in the PCA procedure. For the best fit of the data, values ranging from 0 to 1 for the squared cosines, also known as R^2 or the correlation coefficient, were considered as these values give the proportion of the total variation that can be attributed to independent variables.

2.3.5. Correlations between variables and PCs, loading plot, biplot

To illustrate where each variable fell along the PC axes according to their degree of correlation, the correlation centroid plot was employed. The horizontal axis represents the first PC, and the vertical axis represents the second PC, while the length of the line from the variable data point to the origin indicates how accurately the variable is depicted on the factor map. The distribution of data points for a given variable over the four quadrants of the centroids reveals patterns in the relationships between the various variables. As a rule of thumb, variables that are positively associated will be placed in the same quadrant, whereas variables that are negatively correlated will be placed in opposite quadrants. The loading plot has long been a standard in PCA data visualization, where the selected PCs are plotted against the observations. When points on a loading plot are evenly distributed across the graph, it could mean that PCA was ineffective at reducing the dimensionality of the dataset, or that there are better proportions of information under other PCs with substantial unique clusters. A biplot combines the loading plot with the correlation centroid to show the user how their variables and observations are correlated. In this study, a distance biplot was used to analyze the impact of vector variables in relation to the PC axes by measuring their length.

3. Results and discussion

3.1. Exploratory data analysis of input variables

A total of 323 datasets collected from the thirty relevant publications made up the final database. The substrate being employed (sucrose and tea), operational conditions (temperature, pH, and fermentation duration), and inoculum (kombucha starter tea and SCOBY mass) can all be categorized as input factors. Types of tea leaves such as green tea, black tea, waste tea, and a 2:1 mix of green and black tea for kombucha fermentation are also included in the database as the secondary input parameters. Additionally, operational factors like the depth and surface area of growth tanks affect the quantity and quality of cellulose. However, they were left out owing to a lack of available data in the published literature. EDA revealed the most fundamental statistical explanations of data distribution. Table 1 displays the data analysis of the input and output parameters used in PCA development. The count shows how many entries there are in the database for a specific parameter. The mean is used to represent each parameter's average value. The standard deviation and the four quartiles (25, 50, 75, and 100 %) shed light on the distribution of the data. The parameters' extreme values, the minimum and maximum values, are also displayed with the help of violin plot (Fig. 1). The main sources of nitrogen and carbon for yeast and bacteria to develop are tea and sugar (sucrose).

Sugared tea broth is the best fermentation medium for biofilm (bacterial cellulose) production, as the symbiotic yeast converts sucrose to fructose and glucose, which are further oxidized to ethanol. Glucose creates more gluconic acid, which hinders cellulose formation, hence sucrose is preferred (Greenwalt et al., 2000). Kombucha fermentation increases acetic acid bacteria growth and bacterial cellulose synthesis by producing ethanol (Soh and Lee, 2002). Caffeine in tea may boost bacterial cellulose production (Chakravorty et al., 2019). Sugar and tea concentrations range from 0 to 550 and 0–200 g/L in the database shows the varied study circumstances. However, 60–120 g/L sucrose and 5–10 g/L tea are optimal (Laavanya et al., 2021). Al-Kalifawi and Hassan (2014) found that increasing substrate concentrations decrease biofilm output (Goh et al., 2012a). Over 75 % of datasets were within the recommended tea and sugar content range. This study considers the operational parameters such as temperature, pH, and fermentation time. Kombucha ferments at room temperature in the dark. Though the database reports the range of 20–45 °C fermentation temperatures, the mean temperature is the preferred one. As per the dataset, pH of 2–4 is close to optimal levels as low pH reduces microbial growth and cellulose synthesis, while neutral pH promotes pollutant growth (Sharma and Bhardwaj, 2019).

Fermentation may continue between 7 and 56 days, according to experimental data gathered from the body of research on kombucha fermentation investigations. However, a prolonged fermentation period decreases cellulose yield for two main reasons: 1) The depletion of nutrients; and 2) the fermentation-related accumulation of carbon dioxide (CO_2) at the biofilm's media interface. Long-term CO_2 buildup in the interface, which has an impact on the bacteria, is what ultimately causes the anaerobic situation (Chen and Liu, 2000). The substrates, or inoculum, have an effect on how a biofilm forms. The tea-sugar broth is inoculated by adding fresh SCOBY cultures. A cellulose fibre matrix with microbial cells trapped in it makes up a SCOBY biofilm. Growing a single strain of cellulose-producing bacteria is demonstrated to be less effective than SCOBY culture in the production of cellulose (Sharma and Bhardwaj, 2019). To gently lower the pH of the culture broth and ensure the survival of the SCOBY microbial population, fermented kombucha tea is added as a starter. 5–25 % and 8-40 g/L of kombucha starter tea and SCOBY mass are utilized in combination as the inoculation for kombucha fermentation studies. With a range of 0.38–465 g/L, the wet weight of biofilm output averages 55 g/L of bacterial cellulose. The statistical analysis indicates that the above 75 % yield ranges between 13 and 64 g/L, which is comparable to the BC yield seen under typical fermentation conditions. 75 % of the productivity has an average of 4.7 g/L/day and ranges from 1 to 5 g/L/day. Fig. 2 depicts the Pearson correlation matrix on relationship between the input and output features of Kombucha fermentation for achieving higher SCOBY yield.

3.2. Principal components and correlations derived through PCA

3.2.1. Eigenvalue and scree plot

To begin interpreting PCA results, the resulting PCs were assessed using Kaiser's criterion (Kumar, 2017), which stipulates that only PCs with eigenvalues greater than one should be considered for PCA

Table 1
Statistical analysis of input and output parameters for SCOBY yield from kombucha fermentation.

	Tea (g/L)	Sugar (g/L)	SCOBY (g/L)	Kombucha tea (ml/L)	pH	Temperature (°C)	Duration (days)	Yield (g/L)	Productivity (g/L-day)
Count	317	317	309	276	227	296	323	323	323
Mean	18.21	105.77	17.92	88.77	3.80	27.36	13.50	55.05	4.79
Standard deviation	33.88	101.41	11.73	63.78	0.95	3.92	8.64	75.05	6.41
Minimum	0.00	0.00	0.00	0.00	2.00	20.00	1.00	0.36	0.02
25 %	5.00	50.00	8.00	50.00	3.00	25.00	7.00	13.19	1.12
50 %	8.10	80.00	20.00	100.00	3.60	28.00	14.00	25.00	2.61
75 %	12.00	100.00	30.00	100.00	4.00	30.00	15.00	64.79	5.19
Maximum	200.00	550.00	40.00	250.00	7.60	45.00	56.00	456.00	41.31

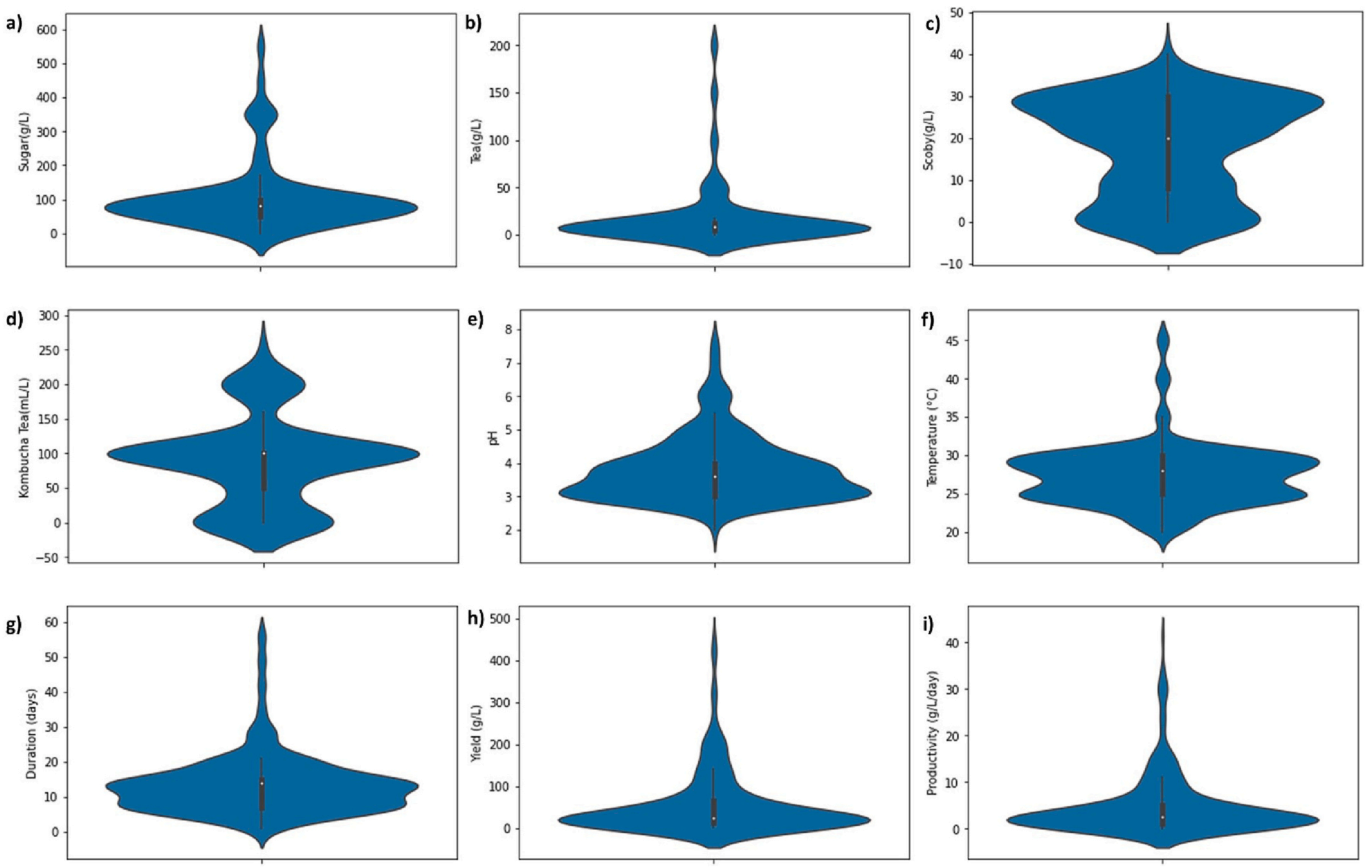

Fig. 1. Violin plot depicts the distribution and density of a) sugar, b) tea, c) SCOBY, d) kombucha tea inoculum e) pH of the fermentation medium, f) fermentation temperature, g) fermentation duration, h) SCOBY yield and i) SCOBY productivity.

analysis. Following PCA, the dataset returns the top three PCs with eigenvalues greater than one, accounting for approximately 61.39 % of the variation (as seen in Table 2). In order to solve the number of factors problem, a second criterion is used. If a component accounts for a certain proportion (or percentage) of variance in the data, it is kept. The number of dimensions that sum up to a suitably substantial fraction of the variance is often preferred (Géron, 2017). When researchers solve the number-of-components problem using the "cumulative percent of variation accounted for" criterion, they typically keep sufficient components so that the cumulative percent of variance addressed for is at least 70 % (Cangelosi and Goriely, 2007). PC4 has an eigenvalue of 0.95, which may be rounded to 1, and so can be considered, resulting in a total variance of 74.96 % (as seen in Table 2).

A third criterion for solving the number of factors problem is using a typical Scree plot as an example of selecting PCs. The eigenvalues and cumulative eigenvalues have historically been represented using the scree plot. The curve frequently has an elbow where the explained variance stops expanding as rapidly as the dataset's intrinsic dimensionality (Géron, 2017). The scree plot (shown in Fig. 3) identifies the first five PCs ($2.014 > 1.243 > 1.040 > 0.950 > 0.865$) as having higher values than other PCs of comparable size. The elbow occurs at PC5, as seen in Fig. 3 implying that PC1, PC2, PC3, and PC4 can be considered for further study.

3.2.2. Eigenvectors, factor loadings, and contribution of variables

Eigenvectors in PCA define the directions of the feature space of PCs. These vector values provide the user with information about the parent variables of the dataset's informational compression directions. The more a variable contributes to the PC axis, the closer its vector value is to $+1$. Values close to -1 are not important because they have no impact on the PC axis. The less a variable contributes to a PC, the lower its eigenvector value as seen in the PC score chart. A correlation chart between PCs and variables, which shows that positive values suggest a bigger contribution to PC and negative values indicate a lower

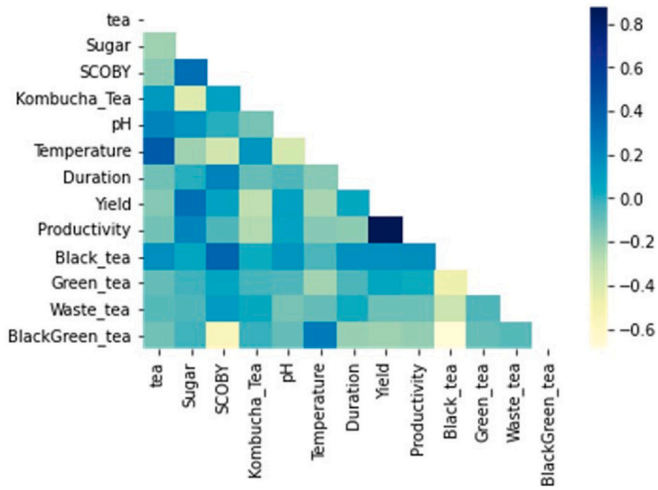

Fig. 2. Pearson correlation matrix depicting relationship between the input and output features of kombucha fermentation for achieving higher SCOBY yield.

contribution, also supports this. As shown in Table 3, PC1's direction is influenced by fermentation temperature, tea leaves, and kombucha starter tea (which each contribute 0.518, 0.427, and 0.295 to the PC1 axis), whereas PC2's direction is influenced by kombucha starter tea, duration of the fermentation process, and SCOBY inoculum mass (which each contribute 0.556, 0.437, and 0.408 to the PC2 axis, respectively). The highest recorded eigenvector for the entire dataset is for pH, which contributes 0.715 to the PC3 axis. Next, tea leaves contribute 0.620 to the PC4 axis, while duration of the fermentation process contributes 0.577 to the PC5 axis. Since the eigenvalue is negative in both PC1 and PC2, sugar and pH make up less of each PC. As shown in Table 3, pH contributes more to PC1 (−0.177) than PC2 (−0.412) whereas sugar contributes more to PC2 (−0.343) than PC1 (−0.462). Further, PC4 (0.216) and PC3 (0.715) respectively have the highest contributions from sugar concentration and pH of the fermentation medium.

Factor loading of PCs is used to calculate the correlation coefficients between the study variables and aided in generating PCs. They depict the components that contribute to making a specific component, with the magnitude of loading playing a role. PC1 had the highest loading values from the fermentation temperature, tea leaves, and kombucha starter tea, which were 0.735, 0.606, and 0.419 respectively. According to Table 4, PC2 found that the kombucha starter tea, fermentation time, and SCOBY inoculum mass had the highest loading values of 0.620, 0.487, and 0.115. The level of the importance of factor loadings generally followed the same pattern as that of eigenvectors, concretizing the dimensions of the new feature space.

3.2.3. *Correlation between variables and principal components*

Analysis of correlations between the variables revealed that mass of the tea leaves, volume of kombucha starter tea and the fermentation temperature affects PC1 in increasing order. Similarly, SCOBY mass, duration of the fermentation process and the volume of kombucha starter tea affects PC2 in increasing order. From Fig. 4, it is seen that pH is negatively associated with fermentation duration as bacteria and yeast present in the inoculum multiply and create organic metabolites like

Table 3
Eigenvector of seven input parameters on different PCs with pH showing the maximum influence.

	F1	F2	F3	F4	F5
Tea (g/L)	0.427	−0.189	0.182	0.620	−0.190
Sugar (g/L)	−0.462	−0.343	−0.196	0.216	−0.448
SCOBY (g/L)	−0.418	0.408	0.153	0.309	−0.500
Kombucha tea (ml/L)	0.295	0.556	0.482	−0.106	−0.238
pH	−0.177	−0.412	0.715	0.228	0.301
Temperature (°C)	0.518	−0.105	−0.326	0.295	−0.180
Duration (days)	−0.205	0.437	−0.235	0.569	0.577

Table 2
Eigen value, variability (%) and cumulative variability (%) of PCs with eigen values greater than one, accounting for approximately 61.39 % of the variation.

	F1	F2	F3	F4	F5	F6	F7
Eigen value	2.014	1.243	1.040	0.950	0.865	0.537	0.351
Variability (%)	28.768	17.761	14.862	13.566	12.363	7.669	5.011
Cumulative (%)	28.768	46.529	61.391	74.958	87.320	94.989	100.000

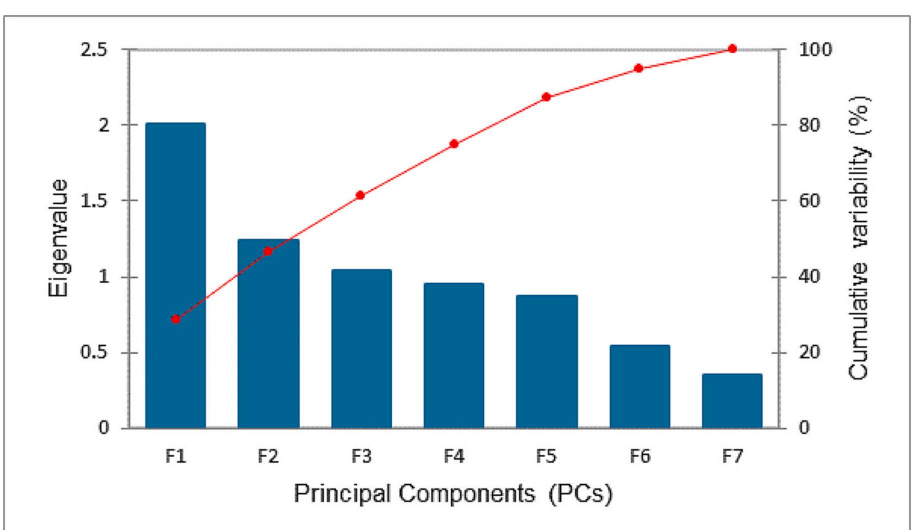

Fig. 3. Scree plot showing eigenvalue and cumulative variability (%) of PCs with the elbow occurring at PC5 of kombucha fermentation for achieving higher SCOBY yield.

Table 4
Factor loadings of seven input parameters on various PCs with the highest loading values depicted by temperature, tea leaves and kombucha starter tea in PC1.

	F1	F2	F3	F4	F5
Tea (g/L)	0.606	−0.211	0.186	0.604	−0.176
Sugar (g/L)	−0.656	−0.382	−0.200	0.210	−0.417
SCOBY (g/L)	−0.593	0.455	0.156	0.301	−0.465
Kombucha tea (ml/L)	0.419	0.620	0.492	−0.103	−0.221
pH	−0.251	−0.459	0.729	0.222	0.280
Temperature (°C)	0.735	−0.117	−0.332	0.288	−0.168
Duration (days)	−0.292	0.487	−0.240	0.554	0.537
Black tea	−0.116	0.115	0.139	0.347	−0.111
Green tea	−0.115	0.043	0.045	−0.137	0.015
Waste tea	−0.055	0.136	−0.011	−0.049	−0.026
Black + green tea	0.264	−0.243	−0.196	−0.301	0.122

acetic acid (Ahmed et al., 2020). Zhao et al. (2018) reported the drop in pH of the fermentation medium from 6.71 to 3.91 in three days and then dropped to the lowest value (about 2.9–3.0) on the tenth day. Similarly, Jayabalan et al. (2014) observed the change in pH from 5.44 to 3.0 after about five to seven days. Nummer (2013) postulated that kombucha shall be treated as contaminated if the pH does not drop to 4.6 or lower within 7 days, as higher pH levels can foster undesirable development of other microbial niches. The correlation centroid shows that pH is negatively correlated to both kombucha starter tea and SCOBY mass. The significant drop in pH within the first hour is a noteworthy observation in the majority of investigations. Hammel et al. (2016) reported similar observations in case of kombucha fermentation, where the pH dropped from 5.5 to 4.75 within the first hour. This considerable drop in pH is unlikely to be the consequence of fermentation alone, and it's more likely that acidic components from the SCOBY mass itself were discharged.

Every time a new batch of kombucha is fermented, a portion of SCOBY from a previously fermented batch is transferred to a fresh tea container. The last batch of kombucha would have reached a low pH level, and the SCOBY would have absorbed this low-pH solution.

When this SCOBY mass enters a fresh tea container, it gradually releases its acidic contents, leading in a decrease in pH (Hammel et al., 2016). The initial pH of kombucha tea reported in literature ranged from 3.75 to 5.0, indicating that the SCOBY mass may release acidic chemicals from a prior brew regardless of any ongoing fermentation. In some trials, a tiny quantity of previous kombucha tea is added to the current batch in order to reduce the initial pH and provide a more accommodating environment for the SCOBY (Sievers et al., 1995). Because the pH of kombucha starter tea is between 2.5 and 4, adding it to a new batch of kombucha will lower the pH of the fresh batch. Regardless of temperature effects, the pH dropped to 4.6 in less than 12 h at room and refrigerated temperatures, resulting in a safe product (Hammel et al., 2016). Because no food-borne illness-causing bacteria can develop or survive in such an environment, kombucha is no longer potentially harmful.

From Fig. 4, it is seen that sugar has a negative relationship with both SCOBY mass and fermentation duration, which aligns with the literature as sugar is used as a substrate during fermentation, and as many nutrients have been used for the microorganisms' growth and metabolism. Microbes in kombucha break sucrose into glucose, fructose, ethanol, acetic acid, lactic acid, and other metabolites (Malbaša et al., 2008). However, it is noteworthy to mention from Fig. 4 that sugar had a positive relationship with pH because the starter culture of kombucha uses various sugars, which resulted in the creation of many organic acids, lowering the tea's pH (Jayabalan et al., 2014). Additionally, there is a negative link between the SCOBY mass and a green-black tea blend, but a positive correlation between the SCOBY mass with waste tea, green tea, and black tea. Green tea exhibited a significantly higher positive impact on BC yield as far as tea type is concerned, followed by black tea. On the other hand, it was discovered that mixing black and green tea decreased the BC yield. Since the reason for this effect is unknown, a thorough analysis of the ingredients in tea is necessary (Priyadharshini et al., 2022).

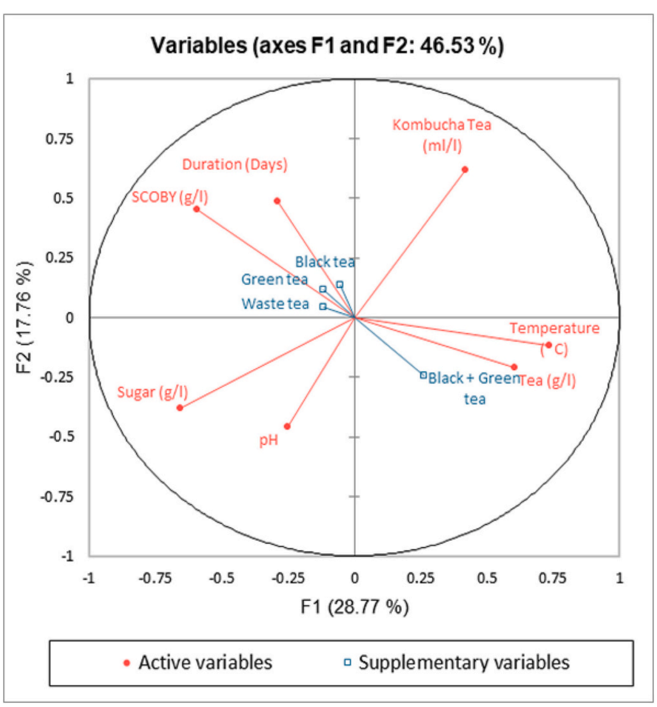

Fig. 4. Correlation centroid showing correlation between the different input variables of kombucha fermentation for achieving higher SCOBY yield.

3.2.4. PCA loading plot

The loading plot depicts the observations in the form of a plot against the PCs that were selected. Generally occurring as separate clusters of adjacent characteristic values, which can then be categorized according to their proximity to one another. From Fig. 5, it is noticed that high and very high BC yield points are clustered significantly in the third quadrant. Further, low, and medium BC yield points are clustered significantly in the fourth and second quadrant, respectively. However, medium yield points are clustered near the center in the second quadrant. Low BC yield datapoints are grouped close to the center of the first quadrant. Likewise, the first quadrant has the fewest high and very high yield points, followed by second, fourth and third quadrants. Due to the clustering of high and very high yield points in the first quadrant, where sugar and pH also happen to be located according to the correlation centroid, sugar and pH have the greatest impact on high BC yield.

3.2.5. PCA biplot

To show insights into variable and observation correlations, a biplot (Fig. 6) combines the optimal loading PCA plot with the correlation centroid. As the production of bacterial cellulose is dependent on the availability of a carbon source, it has been observed in the third quadrant that a high concentration of sugar contributes to a high to very high yield of bacterial cellulose. This finding is in agreement with the research that indicates a concentration of sucrose at 100 g/L produced the highest yield of pellicle (Al-Kalifawi and Hassan, 2014). The kombucha cellulose output increased as sugar content rose to a range of 70–100 g/L. This observation was consistent with the ideal circumstances described in the literature (Laavanya et al., 2021). The usage of sugar at a concentration lower than optimal can diminish BC yield since high sugar reserves are used up by bacteria, resulting in substrate exhaustion.

Malbaša et al. (2008) found that, in comparison to sucrose concentrations of 70 g/L, 50 g/L, and 35 g/L, kombucha fermented with a high

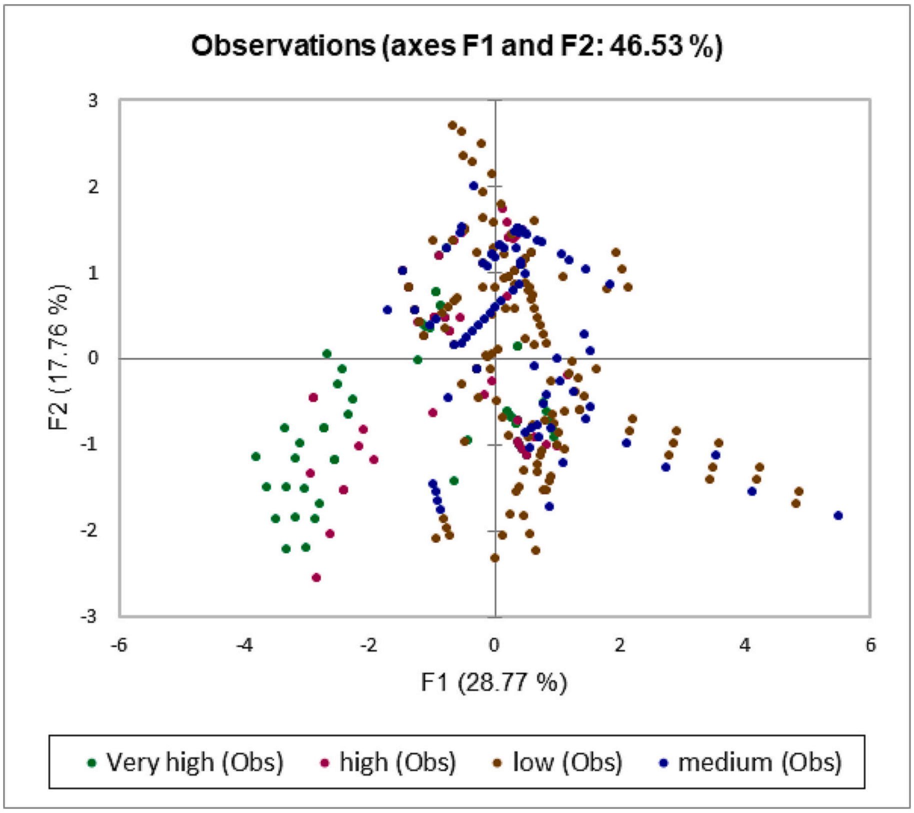

Fig. 5. Loading plot showing different categories of SCOBY yield points from kombucha fermentation.

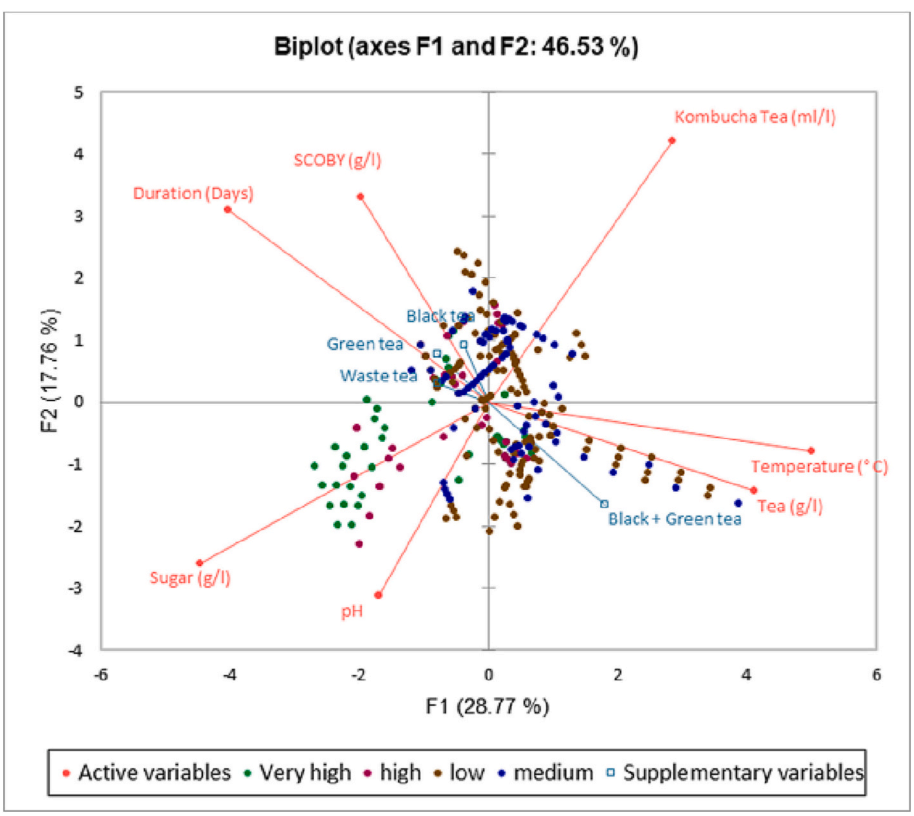

Fig. 6. Biplot highlighting the various SCOBY yields during the kombucha fermentation.

sucrose concentration (70 g/L) produced the most bacterial cellulose (260 g). The most significant factor was sugar content, with high and mid-range sugar concentrations enhancing BC yield. On the other hand, a sufficient amount of sugar may decrease the amount of bacterial cellulose produced due to product inhibitory interactions (Al-Kalifawi and Hassan, 2014). High sugar levels cause increased metabolic activity, which increases the production of gluconic acid that prevents bacteria from synthesizing cellulose (Sharma and Bhardwaj, 2019). Another explanation could be that chemicals build up inside the cell as a result of an imbalance in the rates of nutrient uptake and utilization (Goh et al., 2012b).

In quadrant 3, pH of 2–4 results in a medium cellulose output (Goh et al., 2012a), however pH over 4 results in a reduced yield (Sharma and Bhardwaj, 2019) because it encourages the growth of other non-desired microbial niches. In quadrant 2, the yield rises from low to medium and then falls to low, which aligns with the literature that BC yield increases gradually over the course of fermentation, peaks at 18 days, and then falls (Gargey et al., 2019). The decline in BC yield after two weeks could be due to SCOBY microorganisms reaching the stationary phase and decline in dissolved oxygen levels. As a result, only bacteria connected to SCOBY mass on the air-liquid interface can actively generate cellulose, whereas bacteria in the broth have a poor production capacity of BC (Al-Kalifawi and Hassan, 2014).

It is noteworthy to mention that all data collected from the publications for PCA employed static fermentation experiments only. This explains why extended fermentation times have a detrimental impact on kombucha bacterial cellulose yield. Fourth quadrant shows that as tea leaf concentration increases, yield decreases from medium to low, which is consistent with the literature that shows 10 g/L black tea had the highest bacterial pellicle value when compared to another set of seven tea concentrations (5 g/L, 15 g/L, 30 g/L, 60 g/L, 90 g/L, 120 g/L). Tea leaves are rich in polyphenols with antibacterial and antioxidant effects. Therefore, greater quantities of tea may inhibit bacterial growth, resulting in decreased kombucha biofilm production (Sharma and Bhardwaj, 2019). Several earlier investigations have reported the substitution of acetic acid for kombucha starter tea. Thus, a positive effect might be observed even in the absence of kombucha starting tea. An increase in kombucha tea concentration (>100 ml/L) may reduce biofilm yield because the pH will become unsuitable for yeast and bacterial growth (Soh and Lee, 2002). The concentration of SCOBY mass (30 g/L) is favorable for boosting the biofilm yield. SCOBY mass concentrations range from 0 to 30 g/L in 75 % of the data. As the initial SCOBY mass concentration rises to 30 g/L in the second quadrant, the BC yield increases from low to medium.

3.2.6. *Squared cosines of variables*

SCV describes the accuracy of the variable representation on the factor map. The fraction of variation among dependent variables is calculated using squared cosines. The variable is likely to be well represented on the PCs if the SCV is high. In this instance, the variable is close to the correlation circle's edge. A low value for squared cosine indicates that the PCs are not accurately capturing the variable. In this case, the variable is close to the circle's center. For this investigation, SCV between 0 and 1 were chosen because they represent the ideal data fit. Negative values for the squared cosine can be ignored because they aren't useful. The highest R^2 value among the five selected PCs is in PC1, followed by PC2 and PC3 (as shown in Table 5). This suggests that in the new feature space produced by the five PCs, PC1 has the most significant data distribution and fits the best in it. The finest data representation in PC1 can be found in the categories of tea concentration, sugar concentration, SCOBY mass, and fermentation temperature, with maximum SCV of 0.367, 0.431, 0.352, and 0.541, respectively. The optimum data representation in PC2, PC3, and PC4 may be found for kombucha starter tea, pH of the fermentation process, and fermentation time, with SCV of 0.384, 0.532, and 0.307, respectively.

Table 5
Squared cosines of the input variables on PCs with PC1 having the highest value of R^2.

	F1	F2	F3	F4	F5
Tea (g/L)	**0.367**	0.044	0.035	0.365	0.031
Sugar (g/L)	**0.431**	0.146	0.040	0.044	0.174
SCOBY (g/L)	**0.352**	0.207	0.024	0.091	0.217
Kombucha tea (ml/L)	0.175	**0.384**	0.242	0.011	0.049
pH	0.063	0.211	**0.532**	0.049	0.078
Temperature (°C)	**0.541**	0.014	0.110	0.083	0.028
Duration (days)	0.085	0.237	0.058	**0.307**	0.288
Black tea	0.013	0.013	0.019	**0.120**	0.012
Green tea	0.013	0.002	0.002	**0.019**	0.000
Waste tea	0.003	**0.018**	0.000	0.002	0.001
Black + green tea	0.069	0.059	0.038	**0.091**	0.015

Bold values represents maximum squared cosine of input variables w.r.t. different PCs.

4. Conclusion

PCA was employed to identify the key factors influencing the kombucha BC yield. PC1's direction is influenced by fermentation temperature, tea leaves, and kombucha starter tea. The correlation between seven primary input variables was studied using correlation centroid. Biplot shows that high sugar concentrations contribute to very high yield of BC. pH of 2–4 and increased fermentation time is favorable for kombucha BC yield. Other highly influential input parameters, such as the fermenter's surface area and SCOBY type, can be included in future analyses to make predictions more credible that could aid in scaling up.

CRediT authorship contribution statement

Paramasivan Balasubramanian: Conceptualization, Methodology, Investigation, Review & editing, Supervision, Final approval.
Prachi Tanya Praharaj: Investigation, Data curation, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors acknowledge the Department of Biotechnology (DBT) of Government of India for funding the research under Biotechnology Ignition Grant (BIG) of Biotechnology Industry Research Assistance Council (BIRAC) [BIRAC/KIIT0471/BIG-13/18]. The authors are thankful to the Department of Biotechnology and Medical Engineering, National Institute of Technology Rourkela, for providing the necessary research facilities. Further thanks to the lab members of Bio Energy and Environment (BEE) laboratory for their assistance throughout all aspects of this study and for their help in improvising the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biteb.2023.101539.

References

Abd El-Salam, S.S., 2012. Bacterial cellulose of kombucha mushroom tea. N. Y. Sci. J. 5 (4), 81–87.

Ahmed, R.F., Hikal, M.S., Abou-Taleb, K.A., 2020. Biological, chemical and antioxidant activities of different types kombucha. Ann. Agric. Sci. 65 (1), 35–41.

Al-Kalifawi, E.J., Hassan, I.A., 2014. Factors Influence on the yield of bacterial cellulose of Kombucha (Khubdat Humza). Baghdad Sci. J. 11 (3), 1420–1428.

Behera, B., Laavanya, D., Balasubramanian, P., 2022. Techno-economic feasibility assessment of bacterial cellulose biofilm production during the Kombucha fermentation process. Bioresour. Technol. 346, 126659.

Bianchet, R.T., Cubas, A.L.V., Machado, M.M., Moecke, E.H.S., 2020. Applicability of bacterial cellulose in cosmetics–bibliometric review. Biotechnol. Rep. 27, e00502.

Cangelosi, R., Goriely, A., 2007. Component retention in principal component analysis with application to cDNA microarray data. Biol. Direct 2 (1), 1–21.

Carvalho, T., Guedes, G., Sousa, F.L., Freire, C.S., Santos, H.A., 2019. Latest advances on bacterial cellulose-based materials for wound healing, delivery systems, and tissue engineering. Biotechnol. J. 14 (12), 1900059.

Chakravorty, S., Bhattacharya, S., Bhattacharya, D., Sarkar, S., Gachhui, R., 2019. Kombucha: a promising functional beverage prepared from tea. In: Non-alcoholic Beverages. Woodhead Publishing, pp. 285–327.

Chen, C., Liu, B.Y., 2000. Changes in major components of tea fungus metabolites during prolonged fermentation. J. Appl. Microbiol. 89 (5), 834–839.

Coelho, F., do Vale Braido, G.V., Cavicchioli, M., Mendes, L.S., Specian, S.S., Franchi, L. P., et al., 2019. Toxicity of therapeutic contact lenses based on bacterial cellulose with coatings to provide transparency. Cont. Lens Anterior Eye 42, 512–519.

Dey, B., Behera, B., Parvathy, K.K., Jayaraman, S., Paramasivan, B., 2023. Functionalized bacterial cellulose-based biopolymers for biomedical applications: current research trends and challenges. Biotic Resources 175–191.

Gargey, I.A., Indira, D., Jayabalan, R., Balasubramanian, P., 2019. Optimization of etherification reactions for recycling of tea fungal biomass waste into carboxymethylcellulose. In: Green Buildings and Sustainable Engineering. Springer, Singapore, pp. 337–346.

Géron, A., 2017. Hands-on machine Learning With Scikit-learn and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems.

Goh, W.N., Rosma, A., Kaur, B., Fazilah, A., Karim, A.A., Bhat, R., 2012a. Fermentation of black tea broth (Kombucha): I. Effects of sucrose concentration and fermentation time on the yield of microbial cellulose. Int. Food Res. J. 19 (1), 109–117.

Goh, W.N., Rosma, A., Kaur, B., Fazilah, A., Karim, A.A., Bhat, R., 2012b. Microstructure and physical properties of microbial cellulose produced during fermentation of black tea broth (Kombucha). II. Int. Food Res. J. 19 (1), 153–158.

Greenwalt, C.J., Steinkraus, K.H., Ledford, R.A., 2000. Kombucha, the fermented tea: microbiology, composition, and claimed health effects. J. Food Prot. 63 (7), 976–981.

Hammel, R., Karakilic, V., Shaw, F., 2016. The affect of temperature and pH on the food safety of kombucha tea. BCIT Environ. Public Health J.

Huang, Y., Zhu, C., Yang, J., Nie, Y., Chen, C., Sun, D., 2014. Recent advances in bacterial cellulose. Cellulose 21 (1), 1–30.

Jayabalan, R., Malbaša, R.V., Lončar, E.S., Vitas, J.S., Sathishkumar, M., 2014. A review on kombucha tea—microbiology, composition, fermentation, beneficial effects, toxicity, and tea fungus. Compr. Rev. Food Sci. Food Saf. 13 (4), 538–550.

Kapoor, A., Krishnamoorthy, N., Pathy, A., Balasubramanian, P., 2023. Chemometric analysis unravelling the effect of key influencing factors on algal biochar yield. Algal Res. 69, 102908.

Kim, K., Labbé, N., Warren, J.M., Elder, T., Rials, T.G., 2015. Chemical and anatomical changes in Liquidambar styraciflua L. xylem after long term exposure to elevated CO2. Environ. Pollut. 198, 179–185.

Kumar, K., 2017. Principal component analysis: most favourite tool in chemometrics. Resonance 22 (8), 747–759.

Laavanya, D., Shirkole, S., Balasubramanian, P., 2021. Current challenges, applications and future perspectives of SCOBY cellulose of Kombucha fermentation. J. Clean. Prod. 295, 126454.

Malbaša, R., Lončar, E., Djurić, M., Došenović, I., 2008. Effect of sucrose concentration on the products of Kombucha fermentation on molasses. Food Chem. 108 (3), 926–932.

Mancini, M., Rinnan, Å., Pizzi, A., Mengarelli, C., Rossini, G., Duca, D., Toscano, G., 2018. Near infrared spectroscopy for the discrimination between different residues of the wood processing industry in the pellet sector. Fuel 217, 650–655.

Meftahi, A., Khajavi, R., Rashidi, A., Sattari, M., Yazdanshenas, M.E., Torabi, M., 2010. The effects of cotton gauze coating with microbial cellulose. Cellulose 17, 199–204.

Mohite, B.V., Patil, S.V., 2014. A novel biomaterial: bacterial cellulose and its new era applications. Biotechnol. Appl. Biochem. 61 (2), 101–110.

Nummer, B.A., 2013. Special report: kombucha brewing under the Food and Drug Administration Model Food Code: risk analysis and processing guidance. J. Environ. Health 76 (4), 8–11.

Pacheco, G., de Mello, C.V., Chiari-Andreo, B.G., Isaac, V.L.B., Ribeiro, S.J.L., Pecoraro, E., et al., 2018. Bacterial cellulose skin masks-properties and sensory tests. J. Cosmet. Dermatol. 17, 840–847. https://doi.org/10.1111/jocd.12441.

Priyadharshini, T., Nageshwari, K., Vimaladhasan, S., Prakash, S.P., Balasubramanian, P., 2022. Machine learning prediction of SCOBY cellulose yield from Kombucha tea fermentation. Bioresource Technol. Rep. 18, 101027.

Sharma, C., Bhardwaj, N.K., 2019. Biotransformation of fermented black tea into bacterial nanocellulose via symbiotic interplay of microorganisms. Int. J. Biol. Macromol. 132, 166–177.

Shi, Z., Zhang, Y., Phillips, G.O., Yang, G., 2014. Utilization of bacterial cellulose in food. Food Hydrocoll. 35, 539–545.

Sievers, M., Lanini, C., Weber, A., Schuler-Schmid, U., Teuber, M., 1995. Microbiology and fermentation balance in a kombucha beverage obtained from a tea fungus fermentation. Syst. Appl. Microbiol. 18 (4), 590–594.

Singhania, R.R., Patel, A.K., Tseng, Y.S., Kumar, V., Chen, C.W., Haldar, D., Saini, J.K., Dong, C.D., 2022. Developments in bioprocess for bacterial cellulose production. Bioresour. Technol. 344, 126343.

Siró, I., Plackett, D., 2010. Microfibrillated cellulose and new nanocomposite materials: a review. Cellulose 17 (3), 459–494.

Škrbić, B., Cvejanov, J., Đurišić-Mladenović, N., 2015. Chemometric characterization of vegetable oils based on the fatty acid profiles for selection of potential feedstocks for biodiesel production. J. Biobased Mater. Bioenergy 9 (3), 358–371.

Škrbić, B.D., Đurišić-Mladenović, N., Cvejanov, J., 2018. Differentiation of syngases produced by steam gasification of mono-and mixed sources feedstock: a chemometric approach. Energy Convers. Manag. 171, 1193–1201.

Soh, H.S., Lee, S.P., 2002. Production of microbial cellulose and acids in Kombucha. Prev. Nutr. Food Sci. 7 (1), 37–42.

Sulaeva, I., Henniges, U., Rosenau, T., Potthast, A., 2015. Bacterial cellulose as a material for wound treatment: properties and modifications. A review. Biotechnol. Adv. 33 (8), 1547–1571.

Treviño-Garza, M.Z., Guerrero-Medina, A.S., González-Sánchez, R.A., García-Gómez, C., Guzmán-Velasco, A., Báez-González, J.G., Márquez-Reyes, J.M., 2020. Production of microbial cellulose films from green tea (Camellia sinensis) kombucha with various carbon sources. Coatings 10 (11), 1132.

Zhao, Z.J., Sui, Y.C., Wu, H.W., Zhou, C.B., Hu, X.C., Zhang, J., 2018. Flavour chemical dynamics during fermentation of kombucha tea. Emirates J. Food Agric. 732–741.

Zhu, M., 2004. A Simple Technique for Automatically Selecting the Number of Principal Components via the Use of Profile Likelihood. Working Paper. University of Waterloo.