JINAV: Journal of Information and Visualization *ISSN: 2746-1440 (online)*
Vol. 7 No. 2 (2026)
https://doi.org/10.35877/454RI.jinav4949

Machine Learning-Based Classification of Gnetum gnemon Fiber Tensile Behavior Using the K-Nearest Neighbor Algorithm

S Suryadi, Nicodemus Rahanra, Hermanus Jemy Suripatty, & Wardhana Wahyu Dharsono*

Faculty of Technology and Engineering, Satya Wiyata University, Mandala, Nabire, Central Papua, Indonesia

Abstract

Natural fibers have been widely developed as reinforcement materials for composite structures due to their favorable mechanical properties, low density, and environmental sustainability. One promising natural fiber is Gnetum gnemon fiber. The mechanical characteristics of fibers are commonly obtained through tensile testing, which generates data representing the relationship between displacement and force. Manual analysis of large volumes of tensile test data is time-consuming and may lead to inconsistencies in data interpretation. This study aims to classify the tensile behavior of Gnetum gnemon fiber using the K-Nearest Neighbor (KNN) machine learning algorithm based on displacement and force parameters. The dataset consisted of 1,561 observations distributed across nine sample classes. The nine classes correspond to three fiber batches (S1, S2 and S3), each represented by three replicate specimens (.1, .2 and .3), so that the label encodes both the batch and the replicate within that batch. After the wide-format laboratory export of 1,561 rows and 19 columns was restructured into tidy long format, 11,965 valid force–displacement records were available for modelling. Prior to classification, the data underwent preprocessing, including dataset restructuring, data cleaning, Min-Max normalization, and splitting into 80% training data and 20% testing data. The number of neighbours was set to k = 5, selected by stratified 10-fold cross-validation on the training partition, giving 9,572 training and 2,393 testing records. The KNN model was developed using the Euclidean distance metric to determine class membership based on the nearest Neighbor and was evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. The experimental results showed that the KNN model achieved an accuracy of 69.4% with an average F1-score of approximately 70%. When the nine labels were aggregated into their three parent fiber groups, accuracy rose to 89.3%, and 65.2% of the remaining errors occurred between replicate specimens of the same group rather than between different groups. The prediction visualization and confusion matrix indicated that most samples were successfully classified into their respective classes, although some misclassifications occurred among classes with similar mechanical characteristics. These findings demonstrate that the KNN algorithm is capable of classifying the tensile behavior of Gnetum gnemon fiber with satisfactory performance and has the potential to support more efficient analysis of the mechanical characteristics of natural fiber materials.

Keywords: gnetum gnemon, machine learning, k-nearest neighbor

Received: 15 June 2026 Revised: 30 June 2026 Published: 31 August 2026

1. Introduction

Mechanical characterization is an essential stage in materials research because it provides valuable information regarding the response of a material under applied loading (Saaidia, Belaadi, Boumaaza, & Alshahrani, 2023). One of the most employed characterization methods is the tensile test, which describes material behavior through the relationship between force and displacement during the loading process. Tensile test results are widely used to evaluate the mechanical characteristics of materials, including tensile strength, deformation behavior, and failure patterns (Kanthavel, 2022). Therefore, the analysis of tensile test data plays a crucial role in supporting material development and engineering decision-making across various applications.

Recent advances in data acquisition systems for mechanical testing equipment have significantly improved the capability to continuously record experimental data with high resolution. Each tensile test generates a large number of force–displacement data pairs that describe the material response from the initial loading stage until specimen failure (Jayakrishna et al., 2018). While the availability of large datasets enables more detailed material characterization, it

* Corresponding author.
E-mail address: wardhana.wd@gmail.com

also increases the complexity of data analysis. Manual processing of such datasets is time-consuming, particularly when numerous specimens are tested, and may lead to inconsistent interpretations due to the subjective identification of data patterns (Covallane et al., 2025).

One natural material that can be characterized through tensile testing is Gnetum gnemon fiber. Tensile testing of this fiber produces force–displacement curves that represent the mechanical behavior of individual specimens (Saaidia, Belaadi, Boumaaza, Alshahrani, et al., 2023). However, the resulting datasets generally exhibit considerable variation among samples, making mechanical behavior patterns difficult to identify using conventional analytical approaches. These variations may arise from differences in fiber structure as well as testing conditions. Consequently, a systematic and consistent analytical method capable of recognizing similarities in data patterns is required (Upadhyay et al., 2025).

The rapid development of computational technologies has encouraged the application of machine learning as an effective approach for analyzing experimental data(Conejero et al., 2021; Fawcett, 2006; Nakao et al., 2022; Tv & KN, 2019). Unlike conventional analysis, which relies heavily on researchers' interpretations, machine learning automatically learns the underlying characteristics of data by identifying patterns within datasets (Gardner et al., 2022; Sindhu & Indirani, 2022; Turyahabwe et al., 2026; Zheng, 2015). This approach has been widely adopted in various engineering fields because it improves analytical efficiency, reduces subjectivity, and provides reliable classification and prediction performance. In the field of material characterization, machine learning has increasingly been employed to facilitate the identification of mechanical properties from laboratory testing data (Beldar & Kadbhane, 2024; Chen & Gu, 2020; Schleder et al., 2019).

Among various classification algorithms, the K-Nearest Neighbor (KNN) algorithm is one of the most widely used because of its conceptual simplicity, ease of implementation, and robust performance across different types of datasets. KNN classifies an observation according to its proximity to neighboring samples with known class labels (Beldar & Kadbhane, 2024). This distance-based mechanism enables data with similar characteristics to be assigned to the same class. By utilizing force and displacement as the primary input features, the KNN algorithm has the potential to identify tensile behavior patterns more objectively than conventional approaches based solely on visual inspection of tensile curves (Mahesh et al., 2022).

Previous studies have demonstrated that machine learning techniques can improve the effectiveness of data analysis in materials engineering, particularly for predicting mechanical properties and classifying material characteristics (Swain & Biswas, 2018). However, most existing studies have focused on predicting mechanical property values or evaluating material performance using specific machine learning algorithms. Research specifically investigating the application of the K-Nearest Neighbor algorithm to classify tensile test behavior based on force and displacement parameters for

netum gnemon fiber remains limited (Ganesan et al., 2021). This research gap indicates that the application of KNN-based classification to tensile test data of Gnetum gnemon fiber offers significant opportunities for further investigation.

In this study, the classification of Gnetum gnemon fiber tensile behavior was performed using the K-Nearest Neighbor algorithm. Prior to the classification process, the dataset underwent a preprocessing stage consisting of data cleaning, feature normalization, separation of features and class labels, and partitioning into training and testing datasets. The KNN model was subsequently trained using the training dataset to learn the underlying data patterns and evaluated on the testing dataset using accuracy, precision, recall, and F1-score as performance metrics. This workflow is expected to provide a more effective and reliable classification process for tensile test data.

The objective of this study is to develop a classification model for the tensile behavior of Gnetum gnemon fiber using the K-Nearest Neighbor algorithm based on force and displacement parameters. The primary contribution of this research lies in the application of machine learning to automate the classification of tensile testing data, thereby reducing reliance on manual interpretation through a systematic and quantitative computational approach. The findings are expected to provide an alternative methodology for processing experimental material data and establish a foundation for the future development of artificial intelligence-based systems for mechanical characterization.

The practical value of this classification task follows directly from the way the labels are constructed. Each label identifies the fiber batch from which a specimen was drawn together with the replicate index of that specimen within the batch, so a model that assigns a force–displacement record to the correct batch demonstrates that the batch possesses a mechanically distinguishable tensile signature, whereas confusion between replicates of the same batch indicates acceptable within-batch reproducibility rather than model failure. Reading the classifier in this way supports two concrete engineering decisions: screening a newly extracted fiber batch for deviation from material that has already been characterised, and flagging individual specimens whose tensile response is atypical of their own batch before those specimens are committed to composite fabrication. The classifier is thus proposed as a consistency-checking instrument for experimental material data rather than as a substitute for the measurement of tensile strength itself.

2. Method

2.1. Research Design

This study employed a quantitative research approach using a machine learning-based classification method with the K-Nearest Neighbor (KNN) algorithm(Sidik & Sen, 2019; VijayanV & Ravikumar, 2014). The research workflow began with the collection of a tensile test dataset of Gnetum gnemon fiber, followed by data preprocessing, classification model development, and model performance evaluation using several classification metrics (Selvaraj et al., 2026).

2.2. Research Dataset

The dataset used in this study consisted of tensile test data of Gnetum gnemon fiber obtained from laboratory experiments. The original dataset was stored in Microsoft Excel format and exported from mechanical testing software (Zhang & He, 2022). It comprised 1,561 observations and 19 variables. The dataset contained two primary parameters, namely Displacement (mm) and Force (N), which were used as the input features for the classification model, while the specimen identity served as the class label. The dataset included nine sample classes: S1.1, S1.2, S1.3, S2.1, S2.2, S2.3, S3.1, S3.2, and S3.3. The sample codes are structured hierarchically. The first index denotes the fiber group, S1, S2 and S3, corresponding to three separate batches of Gnetum gnemon fiber that were extracted and conditioned under identical procedures, while the second index denotes the replicate specimen tested within that batch (.1, .2 and .3). Each of the nine codes therefore identifies one tensile-test specimen, and the nine codes together represent three fiber batches with three replicates each. The stated figure of 1,561 observations refers to the number of rows in the original wide-format export, in which every specimen occupies its own pair of Displacement and Force columns. After the file was restructured into tidy long format and rows containing no measurement were discarded, 11,965 valid force–displacement records remained, and it is these records that constitute the dataset used for model training and testing.

Table 1. Dataset Specification

Parameter	Description
Research object	Gnetum gnemon fiber
Number of observations	1,561 rows in the wide-format export; 11,965 records after restructuring into long format
Number of variables	19
Features	Displacement (mm), Force (N)
Label	Sample class (S1.1–S3.3) — the index before the point denotes the fiber group (S1–S3) and the index after the point denotes the replicate specimen within that group (1–3)
Number of classes	9 classes
Data format	Microsoft Excel (.xlsx)

2.3. Preprocessing Data

Prior to the classification process, the dataset underwent a preprocessing stage to ensure compatibility with the requirements of the K-Nearest Neighbor (KNN) algorithm. This stage included dataset restructuring, data quality assessment, column name standardization, numerical data type conversion, and feature normalization using the Min–Max normalization method (Sahare et al., 2024). The dataset was subsequently separated into input features, namely Displacement and Force, and the output variable (class label), representing the specimen class. Finally, the dataset was divided into 80% training data and 20% testing data before model training. Restructuring converted the wide-format export into a single table in which each row contains one Displacement value, one Force value and the corresponding sample label, yielding 11,965 valid records after rows containing no measurement were removed. The Min–Max normalization parameters were fitted on the training partition only and then applied to the testing partition, so that no information from the testing data influenced the scaling. The split was stratified by sample class in order to preserve the class proportions in both partitions and was performed with a fixed random seed for reproducibility, producing 9,572 training records and 2,393 testing records (Kohavi, 1995; Pedregosa et al., 2011).

2.4. K-Nearest Neighbor Model Development

The classification model was developed using the K-Nearest Neighbor (KNN) algorithm, which classifies data based on the proximity of observations within the feature space. The algorithm predicts the class of a new observation according to the majority class among its k nearest Neighbor, where proximity is determined using the Euclidean distance metric (Arunachalam et al., 2026). In this study, Displacement (mm) and Force (N) were used as the input features to represent the mechanical characteristics of Gnetum gnemon fiber. The number of neighbours, k, is the principal hyperparameter of the algorithm and directly controls the bias–variance balance of the decision boundary: small values produce boundaries that follow local structure closely and are sensitive to noise, whereas large values oversmooth the boundary and blur classes that differ only slightly (Cover & Hart, 1967; Hastie et al., 2009). In this study k was therefore not fixed arbitrarily. Candidate values from k = 1 to k = 25 were evaluated using stratified 10-fold cross-validation on the training partition only, with the Euclidean distance metric and uniform neighbour weighting held constant, and the value producing the highest mean cross-validated accuracy was retained. Values that are multiples of the number of classes were excluded in order to reduce the likelihood of ties in the majority vote. This procedure selected k = 5, which was used to generate all results reported in Section 3. The testing partition was not consulted at any stage of the selection of k, so the reported performance remains an unbiased estimate (Guo et al., 2003; Kohavi, 1995).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2.5. Model Evaluation

The developed model was evaluated using the testing dataset to assess the ability of the K-Nearest Neighbor (KNN) algorithm to classify the tensile behavior of Gnetum gnemon fiber. Model performance was measured using four widely adopted classification metrics: accuracy, precision, recall, and F1-score (Sundaravadivel & Satheeskumar, 2026). In addition, a confusion matrix was employed to analyze the number of correct predictions and misclassifications for each sample class. The classification results were further visualized using a two-dimensional scatter plot to illustrate the distribution of samples based on the Displacement and Force variables. Furthermore, the confusion matrix provided a comprehensive representation of the model's classification performance by highlighting the prediction outcomes for all sample classes. The four metrics are computed from the counts of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) obtained for each class. Accuracy is the proportion of correctly classified records over all records, as expressed by the equation given in Section D. Precision is defined as TP / (TP + FP) and expresses the proportion of records assigned to a class that genuinely belong to that class; recall is defined as TP / (TP + FN) and expresses the proportion of records of a class that the model successfully retrieves; and the F1-score is their harmonic mean, 2 × (Precision × Recall) / (Precision + Recall). Because the nine classes are not equally represented in the testing partition, macro-averaged and weighted-averaged values are reported alongside the per-class figures (Sokolova & Lapalme, 2009).

3. Result and Discussion

3.1. Manual Prediction Test

In addition to the evaluation performed using the testing dataset, the K-Nearest Neighbors (KNN) model was further validated through a manual prediction test. This procedure involved providing the trained model with new feature values that had not been previously included in either the training or testing datasets (Sri Sabarinathan et al., 2026). Based on these input values, the model determined the predicted class by identifying the nearest neighbors within the training dataset using the KNN classification mechanism. For example, consider a new tensile test observation with the following feature values:

Table 2. Manual Prediction Test

Displacement	Force
0.030	1.10

Applying the trained model to this observation, the two feature values were first scaled using the Min–Max parameters fitted on the training partition, the Euclidean distances to all training records were computed, and the five nearest neighbours were retrieved. The majority label among those neighbours was S1.3, and the observation was therefore assigned to class S1.3. This outcome is consistent with the restructured data shown in Table 5, in which the S1.3 specimen records the displacement and force pair (0.0317 mm, 1.097 N) that is closest to the queried values, whereas the remaining specimens are separated by a larger distance in the normalized feature space. The test confirms that the trained model reproduces the expected nearest-neighbour behaviour on an input that was not part of either partition, rather than merely reproducing memorised training records.

3.2. Prediction Results Visualization

The prediction results were visualized to illustrate the capability of the K-Nearest Neighbors (KNN) algorithm in classifying the tensile test data of Gnetum gnemon fiber. Presenting the classification results in a visual format facilitates the interpretation of data distribution patterns and the relationship between the Displacement and Force variables across different sample classes. Furthermore, visualization enables the observation of class separability produced by the classification model, providing a more intuitive understanding of its overall performance.

The visualization also serves to evaluate the model's ability to recognize patterns based on the similarity of sample characteristics. When samples belonging to the same class form well-defined clusters with minimal overlap with other classes, the model demonstrates good classification performance. Conversely, the presence of samples from different classes within the same region of the feature space indicates similarities in their mechanical characteristics, which may lead to classification errors (Murugan et al., 2025).In addition to illustrating data distribution, the visualization provides an initial qualitative assessment of the model before quantitative evaluation is performed using the confusion matrix, accuracy, precision, recall, and F1-score. Therefore, the prediction visualization not only serves as an effective means of presenting the classification results but also provides insight into how the KNN algorithm constructs classification patterns based on tensile test data.

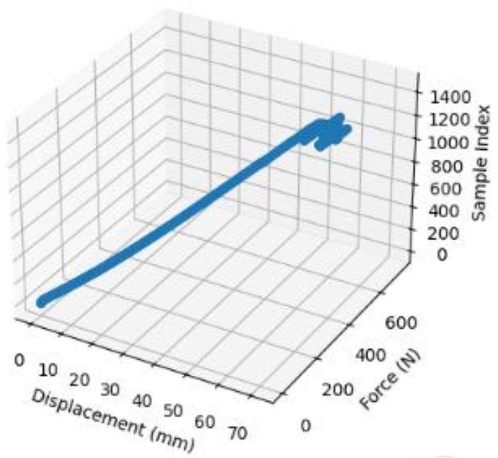

Fig 1. Three-Dimensional (3D) Visualization of Gnetum gnemon Tensile Test Data (exploratory visualization of the raw restructured dataset; the sample-index axis is an acquisition-order index and is not a model feature)

The three-dimensional (3D) visualization is based on three primary variables:

a. X-axis – Displacement (mm): Represents the elongation of the material under tensile loading.
b. Y-axis – Force (N): Represents the magnitude of the tensile force applied to the material.
c. Z-axis – Sample Index: Represents the sequential order of the tested specimens. This axis records the position of each measurement within the acquisition sequence and is used only to spread the records visually so that the continuity of the recorded curves can be inspected. It is not one of the two input features supplied to the classifier, which operates exclusively on Displacement and Force; consequently the classification itself remains a two-feature problem.

Figure 1 is an exploratory plot of the complete restructured dataset generated before classification, and it therefore displays the recorded force–displacement trajectory rather than predicted class labels. As shown in Figure 1, the records follow a single continuous and monotonically increasing trajectory across the full acquisition sequence, which confirms that the restructuring procedure preserved the shape of the tensile curves and introduced no ordering artefacts. This observation indicates that the Displacement and Force parameters provide an adequate representation of the mechanical behavior of Gnetum gnemon fiber. The class-level observations described in the remainder of this paragraph refer to the classified visualizations in Figures 2 and 3 and to the confusion matrix in Figure 4, and not to Figure 1. Nevertheless, several regions exhibit class overlap, resulting in some samples being assigned to incorrect classes. This overlap suggests that certain classes possess relatively similar mechanical characteristics, thereby reducing the classification accuracy of the model. Therefore, the qualitative findings obtained from the 3D visualization should be further validated through quantitative evaluation using the confusion matrix and standard performance metrics, including accuracy, precision, recall, and F1-score. These evaluation measures provide a more comprehensive assessment of the capability of the K-Nearest Neighbors (KNN) algorithm to classify Gnetum gnemon tensile test data.

3.3. Interpretation of the Graph

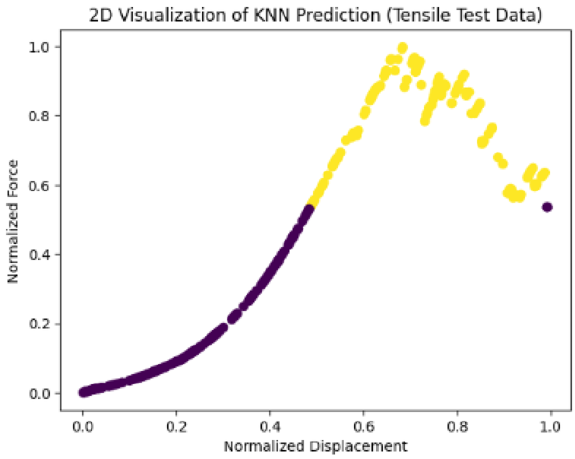

Fig 2. Two-Dimensional (2D) Visualization of Gnetum gnemon Tensile Test Data

3.4. Interpretation

The two-dimensional (2D) visualization reveals several important observations:

a. The data exhibit an increasing curve pattern, indicating a positive relationship between Displacement and Force.
b. The K-Nearest Neighbors (KNN) model classifies the data into two prediction groups based on the magnitude of the force values:
   a) Low-force class
   b) High-force class
c. The color of each data point represents the predicted class assigned by the KNN model.

From a mechanical perspective, the scatter plot can be interpreted as follows:

a. Lower-left region: Initial stage of material deformation.
b. Central region: Progressive elastic deformation stage.
c. Upper-right region: Region approaching the maximum tensile strength of the material.

This type of visualization is commonly employed in machine learning-based materials research because it facilitates the interpretation of classification patterns within the feature space. The prediction results of the KNN model are presented as a two-dimensional scatter plot, where Displacement is plotted on the x-axis and Force on the y-axis. Each point represents a tensile test sample classified by the KNN model, while different colors indicate the predicted class assigned according to the similarity of observations within the feature space. It should be emphasised that Figure 2 is a preliminary two-class demonstration produced on the normalized features, in which the records were grouped into a low-force and a high-force regime solely to illustrate how the distance-based decision rule partitions the feature space. Figure 2 is not the nine-class model reported in this study; the nine-class prediction is presented in Figure 3 and quantified in Table 6 and Figure 4.

3.5. Dataset Exploration and Restructuring

The initial stage of this study involved exploring the Gnetum gnemon tensile test dataset. The objective of this exploration was to understand the dataset structure, identify its storage format, and determine the preprocessing procedures required before applying the K-Nearest Neighbors (KNN) classification algorithm. The exploration results showed that the dataset consisted of 1,561 observations and 19 variables. The dataset was directly exported from mechanical testing software in Microsoft Excel format. Consequently, its original structure was not immediately suitable for machine learning analysis. The data were organized as an experimental table with multiple header levels, requiring a restructuring process to ensure that each observation was represented as a single row in the dataset before further preprocessing and classification As stated in Section 2.B, the figures of 1,561 observations and 19 variables describe the wide-format export rather than the modelling dataset. The 19 columns comprise one index column together with a Displacement column and a Force column for each of the nine specimens, and the restructuring procedure converts these columns into 11,965 single-record rows, each carrying one Displacement value, one Force value and the corresponding sample label.

Table 3. Characteristics of the Original Dataset

Parameter	Value
Number of observations	1,561
Number of variables	19
Data format	Microsoft Excel (.xlsx)
Data type	Tensile test data
Dataset structure	Multi-level header
Readiness for machine learning	Not ready; requires dataset restructuring and preprocessing

Based on the dataset exploration, it was found that the first row contained the sample identification information, while the second row contained the names of the measured variables. The numerical data began in the subsequent rows; therefore, the dataset could not be directly imported using a standard tabular structure. Consequently, a dataset restructuring process was required before further preprocessing and machine learning analysis.

Table 4. Structure of the Exported Dataset

Row	Data Content	Description
Row 0	S1.1, S1.2, S1.3, S2.1, S2.2, S2.3, S3.1, S3.2, S3.3	Sample identifiers
Row 1	Displacement (mm), Force (N)	Variable names
Row 2 onward	Numerical data	Measured displacement and force values obtained from the tensile tests

Each sample consisted of two primary variables, namely Displacement (mm) and Force (N). Accordingly, each pair of columns represented a single tested specimen. This data organization indicated that the dataset did not conform to the tidy data format, as the sample identifiers and measurement variables were arranged horizontally rather than organized into a tabular structure suitable for machine learning analysis.

Table 5. Dataset Format After Restructuring

Sample	Displacement (mm)	Force (N)
S1.1	0.0023	1.036
S1.2	0.0041	1.036
S1.3	0.0317	1.097
S2.1	0.0024	1.041
S2.2	0.0026	1.076
S2.3	0.0032	1.096
S3.1	0.0343	1.087
S3.2	0.0028	1.031
S3.3	0.0027	1.046

After the restructuring process was completed, a data quality assessment was performed as part of the preprocessing stage. This assessment aimed to identify potential issues such as missing values, duplicate records, and data type inconsistencies that could affect the model training process. The results indicated that several columns contained incomplete names due to the export process from the mechanical testing software. Therefore, the column names were standardized, and all numerical values were converted into appropriate numeric data types before feature normalization and K-Nearest Neighbors (KNN) model development were performed. Consequently, the final dataset satisfied the requirements for classification analysis and was ready for both model training and testing.

3.6. Model Evaluation Results

The classification model was evaluated using accuracy, precision, recall, and F1-score to assess the performance of the K-Nearest Neighbors (KNN) algorithm in classifying the tensile test data of Gnetum gnemon fiber. The experimental results showed that the model achieved an overall classification accuracy of 69.4% across nine sample classes.

a. Overall accuracy: 69.4%
b. Number of classes: 9 (S1.1–S3.3)
c. Number of neighbours: k = 5, selected by stratified 10-fold cross-validation (Euclidean distance, uniform weighting)
d. Training records: 9,572; testing records: 2,393 (80:20 stratified split of 11,965 restructured records)
e. Macro-averaged precision, recall and F1-score: 0.704, 0.697 and 0.699; weighted averages: 0.699, 0.694 and 0.694
f. Group-level accuracy after aggregating the nine labels into the three fiber groups S1, S2 and S3: 89.3%

Table 6. Evaluation Results

Sample	Precision	Recall	F1-score
S1.1	0.70	0.80	0.75
S1.2	0.57	0.68	0.62
S1.3	0.76	0.70	0.73
S2.1	0.69	0.71	0.70
S2.2	0.82	0.86	0.84
S2.3	0.54	0.51	0.52
S3.1	0.66	0.66	0.66
S3.2	0.81	0.68	0.74
S3.3	0.78	0.69	0.73

Averaging the per-class results in Table 6 gives a macro-averaged precision of 0.704, a macro-averaged recall of 0.697 and a macro-averaged F1-score of 0.699, with corresponding weighted averages of 0.699, 0.694 and 0.694. The close agreement between the macro-averaged and weighted-averaged values indicates that the overall accuracy is not driven by the larger classes alone and that performance is broadly comparable across the nine labels (Sokolova & Lapalme, 2009).

3.7. Prediction Visualization

The prediction results generated by the K-Nearest Neighbors (KNN) algorithm are presented in Figure 3. The scatter plot illustrates the relationship between Displacement (mm) and Force (N) for the nine sample classes of Gnetum gnemon fiber. The color of each data point represents the class predicted by the KNN model, while the marker shape indicates the corresponding ground-truth class (true label).

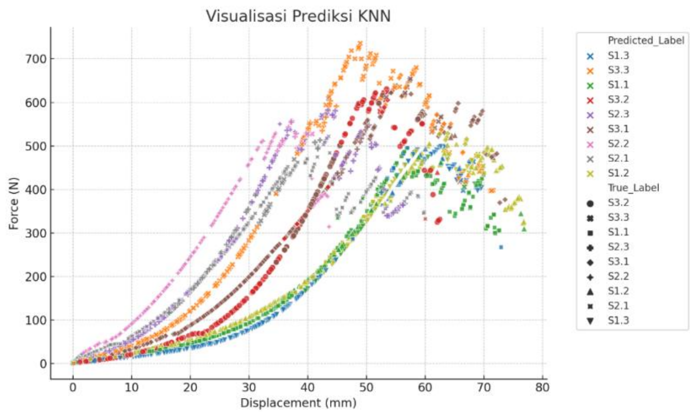

Fig 3. Prediction Visualization KNN

As shown in Figure 3, most data points with similar characteristics form relatively well-defined clusters, indicating that the K-Nearest Neighbors (KNN) algorithm is capable of capturing the relationship between Displacement and Force for each sample class. However, class overlap is still observed within the displacement range of approximately 45–65 mm and the force range of approximately 350–550 N. This overlap suggests that several samples possess similar mechanical characteristics, leading to misclassification of some observations.

Overall, the visualization supports the quantitative evaluation results, in which the KNN model achieved an accuracy of 69.4%. These findings indicate that the KNN algorithm is capable of classifying the tensile behavior of Gnetum gnemon fiber with satisfactory performance, although some sample classes remain difficult to distinguish because of their similar mechanical characteristics.

3.8. Confusion Matrix

The confusion matrix was employed to evaluate the performance of the K-Nearest Neighbors (KNN) algorithm in classifying the tensile test data of Gnetum gnemon fiber across the nine sample classes (Rahman et al., 2026). This matrix illustrates the relationship between the ground-truth labels (true labels) and the predicted labels, thereby providing detailed information on the number of correctly classified observations as well as the misclassifications for each sample class.

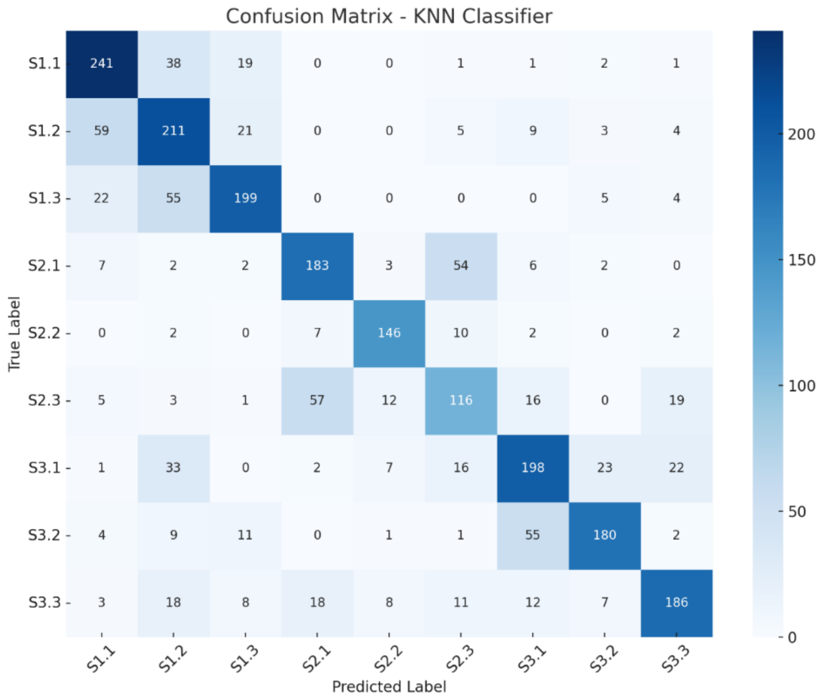

Fig 4. Confusion Matrix of the K-Nearest Neighbors Classifier for the Nine Gnetum gnemon Sample Classes (testing set, n = 2,393)

As shown in Figure 4, the majority of observations were correctly classified, as indicated by the dominant values along the main diagonal of the confusion matrix. Nevertheless, several misclassifications were observed for specific sample classes, particularly between S2.1 and S2.3, suggesting that these classes exhibit similar tensile test characteristics. Overall, the confusion matrix supports the performance of the K-Nearest Neighbors (KNN) model, which achieved an overall classification accuracy of 69.4%. These results indicate that the proposed model can classify the tensile behavior of Gnetum gnemon fiber with satisfactory performance. In numerical terms, 1,660 of the 2,393 testing records lie on the main diagonal. The largest off-diagonal entries are the mutual confusions between S2.1 and S2.3 (54 and 57 records), between S1.1 and S1.2 (38 and 59 records) and between S3.1 and S3.2 (23 and 55 records); in each case the confusion occurs between replicates belonging to the same fiber group.

Because each label encodes a fiber group together with a replicate index, the results in Figure 4 can also be read at the group level. Aggregating the predictions into the three parent groups S1, S2 and S3 raises accuracy from 69.4% to 89.3%, corresponding to 2,138 of the 2,393 testing records assigned to the correct fiber group, with group-level recall of 0.961 for S1, 0.895 for S2 and 0.819 for S3. Of the 733 misclassified records, 478 (65.2%) are confusions between replicates of the same group and only 255 (34.8%) cross a group boundary. The nine-class accuracy of 69.4% therefore understates the practically relevant capability of the model: the tensile signature of a fiber group is recovered reliably, while most residual error arises from the model being unable to distinguish three replicates drawn from the same batch, which are by design expected to behave similarly. For batch-level screening, which is the intended application, the operative error rate is therefore approximately one in ten rather than approximately three in ten.

The results of this study demonstrate that the K-Nearest Neighbors (KNN) algorithm is capable of classifying the tensile behavior of Gnetum gnemon fiber using two primary parameters, namely Displacement and Force. These parameters were successfully utilized as input features to distinguish the mechanical characteristics of the nine sample classes examined in this study. This finding indicates that the relationship between tensile force and displacement during the loading process contains sufficient information to represent the mechanical behavior of each specimen, making it suitable for automated classification (Beldar & Kadbhane, 2024).

The evaluation results showed that the KNN model achieved an overall classification accuracy of 69.4%, indicating that the majority of the testing samples were correctly classified. This level of accuracy suggests that the algorithm effectively identified similarities among observations based on the proximity of their displacement and force values (Swain & Biswas, 2018). Consequently, the proposed machine learning-based classification approach provides a practical alternative to conventional tensile test analysis, which has traditionally relied on visual interpretation of force–displacement curves.

The class-wise evaluation revealed that the model performance was not uniform across all sample classes. Class S2.2 achieved the highest precision, recall, and F1-score, whereas S2.3 exhibited the lowest performance. These differences indicate that some classes possess more distinguishable relationships between displacement and force, while others exhibit mechanical characteristics that are more similar to neighboring classes, making them more difficult for the KNN algorithm to discriminate accurately.

The prediction visualization further demonstrated that most observations formed relatively well-defined clusters according to their respective classes. This clustering indicates that the KNN algorithm successfully captured the similarity among samples within the feature space defined by the displacement and force parameters. Nevertheless, class overlap remained evident in several regions, suggesting that some samples shared similar mechanical characteristics, which consequently led to misclassification.

The confusion matrix further supports these findings. Most observations were correctly classified, as reflected by the dominant values along the main diagonal of the matrix. However, noticeable misclassifications were observed, particularly between classes S2.1 and S2.3, which were frequently predicted as one another. This result suggests that these two classes exhibit relatively similar tensile behavior based on the selected input parameters, thereby limiting the algorithm's ability to distinguish them perfectly.

The findings of this study indicate that the two primary parameters, Displacement and Force, are sufficient to develop a classification model with satisfactory performance. These variables play a significant role in representing the mechanical behavior of Gnetum gnemon fiber (Zhang & He, 2022). By utilizing these two features, the identification of tensile behavior can be performed in a more systematic and objective manner compared with conventional approaches that rely primarily on manual interpretation of tensile test curves.

Overall, the proposed KNN-based classification model successfully achieved the objective of this study by automatically identifying patterns in tensile test data through a distance-based learning mechanism. The developed approach provides a faster, more consistent, and quantitatively measurable alternative for processing experimental material data, thereby demonstrating the potential of machine learning techniques in mechanical characterization.

Despite these promising results, the achieved classification accuracy of 69.4% indicates that there is still room for improving model performance. The observed misclassifications suggest that the mechanical characteristics of some sample classes cannot be completely distinguished using only the two input features employed in this study. Therefore, the findings provide a valuable foundation for future research on the classification of Gnetum gnemon fiber tensile behavior using machine learning approaches, particularly through the incorporation of additional discriminative features or the exploration of more advanced classification algorithms to improve predictive performance. Three specific limitations follow from the design of the experiment. First, the model uses only the instantaneous Displacement and Force values of each record and has no access to curve-level descriptors such as the initial slope, the strain at maximum load or the area under the force–displacement curve; this is the most probable reason why replicates within the same fiber group cannot be fully separated. Second, the labels identify individual specimens rather than distinct material treatments, so the nine-class task is intrinsically harder than a treatment-classification task and its attainable ceiling is bounded by the natural variability between replicates. Third, only the number of neighbours was tuned; the distance metric and the neighbour weighting scheme were held fixed, and distance-weighted voting or alternative metrics may yield further improvement (Guo et al., 2003). Future work should therefore incorporate curve-level features and evaluate ensemble classifiers on the same dataset.

4. Conclusion

This study successfully applied the K-Nearest Neighbors (KNN) algorithm to classify the tensile behavior of Gnetum gnemon fiber based on the Displacement and Force parameters. After dataset restructuring and preprocessing, the developed model achieved an overall classification accuracy of 69.4% across nine sample classes. The prediction visualization and confusion matrix indicated that most samples were correctly classified, although misclassifications occurred between classes with similar mechanical characteristics, particularly S2.1 and S2.3. Overall, the results demonstrate that the KNN algorithm is a promising approach for automating the classification of tensile test data and supporting more efficient and systematic mechanical characterization. When the nine labels are aggregated into their three parent fiber groups, accuracy rises from 69.4% to 89.3%, and 65.2% of the remaining errors are confusions between replicate specimens of the same group rather than between different groups. This indicates that the model recovers the tensile signature of a fiber batch reliably and that most residual confusion reflects the mechanical similarity that replicates of one batch are expected to share. It is this group-level performance, rather than the nine-class figure alone, that is relevant to the practical use of the model for screening fiber batches prior to composite fabrication.

References

Arunachalam, S. J., Saravanan, R., Sathish, T., & Vinuja, G. (2026). Effect of water absorption on the behaviour of jute/kenaf fiber at various lengths by ANN and RSM modelling. *Next Materials*, *11*, 101726.

Beldar, P., & Kadbhane, S. (2024). Utilizing machine learning to forecast mechanical characteristics of NaOH-treated jute fiber reinforced composite materials. *Materials Letters*, *377*, 137411.

Chen, C. T., & Gu, G. X. (2020). Generative Deep Neural Networks for Inverse Materials Design Using Backpropagation and Active Learning. *Advanced Science*, *7*(5), 1–10. https://doi.org/10.1002/advs.201902607

Conejero, J. M., Preciado, J. C., Fernández-García, A. J., Prieto, A. E., & Rodríguez-Echeverría, R. (2021). Towards the use of Data Engineering, Advanced Visualization techniques and Association Rules to support knowledge discovery for public policies. *Expert Systems with Applications*, *170*. https://doi.org/10.1016/j.eswa.2020.114509

Covallane, K., Johan, D., Singh, R. K., Sinha, R., Boodala, D., Jaiswal, K. K., & Kumar, K. S. (2025). Experimental and Numerical Investigation of Jute Fibre-Reinforced Composite, a Sustainable Material for Green Energy. *Engineering Proceedings*, *95*(1), 17.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Ganesan, V., Shanmugam, V., Kaliyamoorthy, B., Sanjeevi, S., Shanmugam, S. K., Alagumalai, V., Krishnamoorthy, Y., Försth, M., Sas, G., & Javad Razavi, S. M. (2021). Optimisation of mechanical properties in saw-dust/woven-jute fibre/polyester structural composites under liquid nitrogen environment using response surface methodology. *Polymers*, *13*(15), 2471.

Gardner, P., Bull, L. A., Dervilis, N., & Worden, K. (2022). Domain-adapted Gaussian mixture models for population-based structural health monitoring. *Journal of Civil Structural Health Monitoring*. https://doi.org/10.1007/s13349-022-00565-5

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN model-based approach in classification. In R. Meersman, Z. Tari, & D. C. Schmidt (Eds.), On the Move to Meaningful Internet Systems 2003 (Lecture Notes in Computer Science, Vol. 2888, pp. 986–996). Springer. https://doi.org/10.1007/978-3-540-39964-3_62

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Jayakrishna, K., Rajesh, M., Soundhar, A., Surendhar, K., & Sultan, M. T. H. (2018). Modeling and predicting the mechanical properties of nano clay dispersed natural fiber hybrid intra-ply woven fabric composite by using response surface methodology. In *Advanced Manufacturing and Materials Science: Selected Extended Papers of ICAMMS 2018* (pp. 221–233). Springer.

Kanthavel, K. (2022). *Optimizing various parameters influencing mechanical properties of banana / coir natural fiber composites using grey relational analysis and artificial neural network models*. https://doi.org/10.1177/1528083720930304

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI), 2, 1137–1143.

Mahesh, V., Mahesh, V., & Harursampath, D. (2022). Physio-mechanical characterization of jute/kevlar hybrid composites coupled with MADM approach for selection of composites. *Journal of Natural Fibers*, *19*(15), 11105–11113.

Murugan, A., Barik, D., Faisal, R. S., & Mani, M. (2025). *Process parameter optimization for enhanced mechanical and thermal properties of kenaf / jute hybrid composites using grey fuzzy logic*. 1–25.

Nakao, Y., Stumpf, S., Ahmed, S., Naseer, A., & Strappelli, L. (2022). Toward Involving End-users in Interactive Human-in-the-loop AI Fairness. *ACM Transactions on Interactive Intelligent Systems*, *12*(3), 1–30. https://doi.org/10.1145/3514258

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Rahman, M., Ismail, A. E., Ramli, M. F., Hani, A., & Rashid, A. (2026). *Prediction of tensile and flexural properties of fiber-reinforced epoxy composites using machine learning models Prediction of tensile and flexural properties of fiber-reinforced epoxy composites using machine learning models*.

Saaidia, A., Belaadi, A., Boumaaza, M., & Alshahrani, H. (2023). Effect of Water Absorption on the Behavior of Jute and Sisal Fiber Biocomposites at Different Lengths : ANN and RSM Modeling Effect of Water Absorption on the Behavior of Jute and Sisal Fiber Biocomposites at Different Lengths : ANN and RSM Modeling. *Journal of Natural Fibers*, *20*(1). https://doi.org/10.1080/15440478.2022.2140326

Saaidia, A., Belaadi, A., Boumaaza, M., Alshahrani, H., & Bourchak, M. (2023). Effect of water absorption on the behavior of jute and sisal fiber biocomposites at different lengths: ANN and RSM modeling. *Journal of Natural Fibers*, *20*(1), 2140326.

Sahare, S. B., Jadhav, V. D., Satishkumar, P., Vijayan, D. S., & Saminathan, R. (2024). Performance analysis of Jute/Kenaf fiber polyester hybrid composites for non-structural applications. *Interactions*, *245*(1), 164.

Schleder, G. R., Padilha, A. C. M., Acosta, C. M., Costa, M., & Fazzio, A. (2019). From DFT to machine learning: recent approaches to materials science–a review. *Journal of Physics: Materials*, *2*(3), 032001. https://doi.org/10.1088/2515-7639/ab084b

Selvaraj, N., Velusamy, S., Tharmalingam, R., & Sivakumar, V. (2026). Experimental Characterization and AI-Based Prediction of Physical and Thermal Properties of Hybrid Fiber-Reinforced Polymer Matrix Composites Using Deep Learning Framework. *Fibers and Polymers*, 1–29.

Sidik, D. D., & Sen, T. W. (2019). Penggunaan Stacking Classifier Untuk Prediksi Curah Hujan. *IT for Society*, *4*(1), 21–27. https://doi.org/10.33021/itfs.v4i1.1180

Sindhu, P., & Indirani, G. (2022). Equilibrium Optimizer with Deep Convolutional Neural Network-based SqueezeNet Model for Grape Leaf Disease Classification in IoT Environment. *International Journal of Engineering Trends and Technology*, *70*(5), 94–102. https://doi.org/10.14445/22315381/IJETT-V70I5P212

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Sri Sabarinathan, R., Yazerarabath, J., Thendralvan, S., Srimathi, R., Srileka, A., & Suvathika, C. (2026). Prediction and optimization of mechanical and moisture properties of jute–coir hybrid composites using a DBN–KNN–GWO model. *Textile Research Journal*, 00405175261458660.

Sundaravadivel, P., & Satheeskumar, S. (2026). Advanced Machine Learning Assisted Prediction of Multi-Fiber Hybrid Natural Fiber/Epoxy Composites Based on Jute, Coconut Coir, Banana, and Pineapple Leaf Fibers. *Journal of Materials Research and Technology*.

Swain, P. T. R., & Biswas, S. (2018). A comparative analysis of physico-mechanical, water absorption, and morphological behaviour of surface modified woven jute fiber composites. *Polymer Composites*, *39*(8), 2952–2960.

Turyahabwe, A., Dennison, M. S., & George, O. S. (2026). Investigation of multi-performance optimization of banana/bark cloth reinforced epoxy composites using grey relational analysis for automotive interior applications. *Scientific Reports*, *16*(1), 14615.

Tv, B., & KN, N. (2019). A Comparative Study of Spectral Indices for Surface Water Delineation Using Landsat 8 Images. *International Conference on Data Science and Communication*, (September). https://doi.org/10.1109/IconDSC.2019.8816929

Upadhyay, V., Kumar, R., & Sharma, C. (2025). Mechanical Properties and Water Absorption Behavior of Jute/Nettle Hybrid Epoxy Composite. In *Advanced Functional and Composite Materials* (pp. 167–177). CRC Press.

VijayanV, V., & Ravikumar, A. (2014). Study of Data Mining Algorithms for Prediction and Diagnosis of Diabetes Mellitus. *International Journal of Computer Applications*, *95*(17), 12–16. https://doi.org/10.5120/16685-6801

Zhang, W., & He, C. (2022). Investigation and prediction of mechanical properties of hot water treated jute/poly (lactic acid) composite laminates using response surface methodology and genetic algorithm. *Polymer Composites*, *43*(10), 7170–7186.

Zheng, Y. U. (2015). Trajectory Data Mining : An Overview. *ACM Trans. Intell. Syst. Technol.*, *6*(3), 1–41. https://doi.org/10.1145/2743025