Property-based biomass feedstock grading using *k*-Nearest Neighbour technique

[a*]Obafemi O. Olatunji, [b,c] Stephen Akinlabi, [a] Nkosinathi Madushele, [a] Paul A. Adedeji

[a,] *Department of Mechanical Engineering Science, University of Johannesburg, South Africa.*
[b] *Department of Mechanical and Industrial Engineering, University of Johannesburg, South Africa.*

[c] *Department of Mechanical Engineering, Covenant University, Ota, Nigeria.*

[*] *Corresponding author: tunjifemi@gmail.com*

Abstract

Energy generation from biomass requires a nexus of different sources irrespective of origin. A detailed and scientific understanding of the class to which a biomass resource belongs is therefore highly essential for energy generation. An intelligent classification of biomass resources based on properties offers a high prospect in analytical, operational and strategic decision-making. This study proposes the k-Nearest Neighbour (k-NN) classification model to classify biomass based on their properties. The study scientifically classified 214 biomass dataset obtained from several articles published in reputable journals. Four different values of k ($k = 1, 2, 3, 4$) were experimented for various self normalizing distance functions and their results compared for effectiveness and efficiency in order to determine the optimal model. The k-NN model based on Mahalanobis distance function revealed a great accuracy at $k = 3$ with Root Mean Squared Error (RMSE), Accuracy, Error, Sensitivity, Specificity, False positive rate, Kappa statistics and Computation time (in seconds) of 1.42, 0.703, 0.297, 0.580, 0.953, 0.047, 0.622, and 4.7 respectively. The authors concluded that k-NN based classification model is feasible and reliable for biomass classification. The implementation of this classification models shows that k-NN can serve as a handy tool for biomass resources classification irrespective of the sources and origins.

Keywords: Biomass classification; Energy; k-NN classifier; Mahalanobis distance

1 INTRODUCTION

Having come to the end of Millennium Development Goals (MDGs), the United Nations (UN) general assembly adopted the 2030 Sustainable Development Goals (SGDs) in September 2015. Focussing on inclusive development, the new agenda emphasize a holistic approach to achieving sustainable development for all [1]. Incidentally, many of the proposed SDGs are dependent on biomass [2]. Goal 7, 9, 12 and 13 of the SDGs substantially speak to the need for alternative energy [1]. By insight into the SDGs, biomass stands as the main avenue through which massive renewable energy can be sustainably achieved in order to ensure access to affordable, reliable, and modern energy for all. This would inevitably culminate in action to combat climate change and its impacts by elevating the sustainable use of biomass resources, since the power generation from fossil fuel continues to raise environmental concerns. Biomass has been identified as the only alternative naturally occurring fossil fuel substitute both regarding carbon content and the available quantity [3]. In 2013, 462 TWh of electricity was produced globally from biomass [4]. Ever since then, the global consumption of biofuel sourced from biomass has been on the increase, and this trend promises to continue as long as major issues surrounding the exploration of bioenergy is frontally addressed. In the future, biomass energy has the potential to provide cost-effective and sustainable energy supply to the population in the developing countries [5, 6]. There is an abundant supply of biomass feedstocks which can be converted to biofuels. These feedstocks include; agricultural crops, residues, forest products, energy crops and algae biomass. Municipal Solid Waste (MSW) is also gaining attention as a viable source of biofuel [7, 8].

There is a consensus that biomass fuel is a renewable energy resource, but lack of widely accepted terminology, classification metrics, and global standard lead to some serious misconception during the investigations. The main concerns surrounding biomass exploration are related to how to extend and improve the basic understanding of the composition and properties of biomass and also to profitably apply this knowledge in the interest of environmental safety [9]. The knowledge of the proximate properties of biomass and their classification are essential in their selection as an energy feedstock [10]. Among several classification criteria which have been adopted in the classification of biomass feedstock, the most prominent is based on origin and properties [11, 12]. There is no known scientific classification method based on artificial intelligence to the best of the authors' knowledge. Since energy generation from biomass often requires a nexus of different sources, a detailed and scientific understanding of the class to which a biomass resource belongs is highly essential. An intelligent

classification of biomass resources based on proximate properties of biomass feedstock offers a high prospect in analytical, operational and strategic decision-making.
Artificial intelligence has opened a new page in the field of data analysis. Machine learning has been applied to several processes, which include data mining and data analysis [13-18] related to supply chain management [19, 20], biomass elemental composition [21], municipal solid waste management [22-24], energy consumption prediction [25] and so on. One of the primary objectives of data mining is classification. It is a form of predictive modelling which defines groups within the entire population. The process of classification focusses on finding a model which describes a data class. The aim of classification is to use the derived function to predict the group to which a data point belongs using an unknown class label. By using the classification technique, one can learn the rules that form categories of data [26]. Data classification has been applied in several fields such as; medicine, credit ranking, customer behaviour, and strategic management [27]. Some of the artificial intelligent-based techniques which has been applied in the classification of various data are; Bayesian classifier, Ensemble classifier, Artificial Neural Network (ANN), *k*-Nearest Neighbour (*k*-NN), Support Vector Machine, SVM [14, 28]. The *k*-NN is a common instance-based learning algorithm, which is used to classify an unknown object by ranking the objects neighbour amidst the training data. The result of this then intelligently predicts the class of the new objects. The *k* in *k*-NN is a pointer to the number of proximate neighbours which are been evaluated in order to determine the class to which a dataset belongs [26, 29]. Apart from its simplicity, *k*-NN has proven to be an effective and efficient algorithm used in solving several real-life classification problems with good generalization and accuracy [30]. *k*-NN has been appraised as a transparent machine learning method with high predictive ability, even with little or no prior information about the data distribution [15, 31]. *k*-NN stands out as a data mining method of choice for several classification problems due to its notable advantages, which competitively edge-out other methods. Some of its advantages include [31];

i. Ability to handle training data that are too large to fit in-memory,
ii. It can measure the similarities between training tuples and test tuples without prior knowledge about data distribution,
iii. Reduction of error due to the inaccurate assumption.
iv. Lower computation time and high prediction accuracy.

The literature review has highlighted the success of *k*-NN in the classification of data for several applications [15, 17, 31, 32].

This study proposes a proof of concept for the classification of biomass based on the traditional classification of properties as proposed by Khan *et al.* [11] using *k*-NN classification algorithm. The research seeks to innovate a consistent, flexible, direct and easy to implement classification method for biomass properties. Identifying biomass classes is very vital to understanding the mechanism which lead to the varying biomass feedstock behaviour and for an improved prediction of biomass properties [33, 34]. This will engender an informed decision making about biomass resources management and feedstock production for industrial utilization especially as related to energy generation. Also, power plant developers will be enabled to ascertain the quality of the feedstock coming from various supply chain. Another objective of this study is to extend the boundary of knowledge in terms of biomass selection decision towards the hybridization of biomass sources for energy generation.

2 LITERATURE SURVEY

The application of *k*-NN as a machine learning approach spans more than 50 years [35]. Although *k*-NN was believed to have been introduced in 1951 in an unpublished medicine report, it did not gain much traction until 1960. Ever since then, it has become a renowned pattern recognition and

classification technique [26, 36]. k-NN has been widely applied to various classification and data recognition problems in several studies. Different k-values, distance metrics, and types of data have been applied.

Chomboon *et al.,* [37] evaluated the performance of k-NN classification algorithm using 11 distance functions which are; Euclidean, Chebyshev, Cosine, Hamming, Jaccard, Spearman, Standardized, Manhattan, Minkowski, Mahalanobis, and Correlation. Hypothetic data which were generated by MATLAB were divided in ratio 70 % training set and 30 % testing set. There was a similarity in the accuracy obtained from Euclidean, Standardized, Chebyshev, Manhattan, Minkowski, Mahalanobis. Also, the models based on these distances performed better than other evaluated distance measures. Using the Knowledge Discovery in Databases, KDD dataset obtained from Tavallaee *et al.,* [38] , Punam and Nitin [39] analysed the performance of k-NN algorithm. They considered Chebychev, Euclidean, Manhattan as the measures of distance. The numerical dataset contains 41 features and two classes. Accuracy, sensitivity, specificity was used for the performance assessment for each distance measure. The result obtained showed that Manhattan distance performed better than the other evaluated distance measures with an accuracy, sensitivity, specificity of 97.8 %, 96.8 %, and 98.35 % respectively.

In the medical field, three different kinds of medical dataset obtained by Hu *et al.,* [40] were analysed in order to explain the effect of distance metrics on the k-NN model. The dataset is made of categorical, numerical and mixed types of data which were collected from the University of California, Irvine UCI machine learning archive. Training and testing datasets were 90 % and 10 % respectively. The k values ranging from 1 to 15 were applied to four distance measures which are Euclidean, Chi-square, Cosine, Minkowski. The lowest accuracy was reported for Minkowski, Euclidean, and Cosine distance measures over the mixed type of datasets. Meanwhile, Chi-square distance metrics performed best for all types of dataset.

Todeschini *et al.,* [41, 42] evaluated the performance of k-NN classifier for eighteen different distance metrics using eight datasets. The distance measures evaluated are; Lagrange, Mahalanobis, Bhattacharyya, Canberra, Wave edge, Contracted Jaccard, Jaccard, Euclidean, Lance-William, Soergel, Clark, Cosine, and Correlation, and four locally centred Mahalanobis. The results showed that Manhattan, Euclidean, Soergel, Contracted Jaccard and Lance -Williams distance metrics give the highest accuracy.

Gjertsen [43] developed a multi-source forest inventory (MSFI) technique based on k-NN field plot from Norwegian National Forest Inventory (NFI), land cover maps, and satellite image data. The relationship between the spectral bands and the forest variables showed a low level of association.

Xingjie *et al.,* [44] applied k-NN to reconstruct the position of the control rod in form of an integer using Minkowski distance measure. The result of this simulation showed that the k-NN method can accurately reconstruct the position of control rod from the fixed in-core neutron detector. Table 1 shows some studies which have applied Mahalanobis distance function and the optimum accuracy that was obtained.

Table 1. Some studies that applied Mahalanobis Distance to K-NN

Ref	Author	K range	Mahalanobis Distance Accuracy
[45]	Prasath *et al.*	1	0.8113
[37]	Chomboon *et al.*	8	0.900
[43]	Gjertsen *et al.*	1-20	0.631
[44]	Xingjie *et al.*	1-40	0.43
[46]	Joshi A, Mehta	5-25	0.9529

2.1 Biomass classification

The energy conversion process selection has made the study of the classification of biomass material, which can fit as feedstock very important. Some of the properties of the biomass which have been discussed in the literature are thermal properties, elemental composition, proximate properties, structural properties, and chemical properties [47-51] These properties vary with the type of biomass, part of biomass plant and the species [52], growth process and growing conditions, geographical location, climate and seasons, soil types, water, nutrients and soil pH, age of plant, presence of pollutants and contaminants and the proximity to the pollutants, breeding efforts, harvesting time and technique [33, 34, 53], transport and storage condition, mixing of biomass from multiple sources, pre-treatment, post processing and so on [12, 53, 54]. In view of this diversity, a classification which eases the process of prediction of the properties of biomass is very vital. Khan *et al.,* [11] proposed that biomass should be classified into two classes based on; origin and properties. These have been discussed in [49, 50, 55] and it is outlined in Figure 1. Also, Tao *et al.* [33, 34] synthesizes biomass properties based on the energy properties [33] and ash elements [34] using principal component analysis, though there were no detailed discussion on the performance of this model based on the known standard classification metrics. Most recently, Mimi *et al.* [56] analysed several agro-waste samples which were derived from eastern china using principal component analysis (PCA). The present study presents an intelligent model for the classification of biomass which has been sorted into eight classes as shown in Table 2 and further discussed in section 3.1.

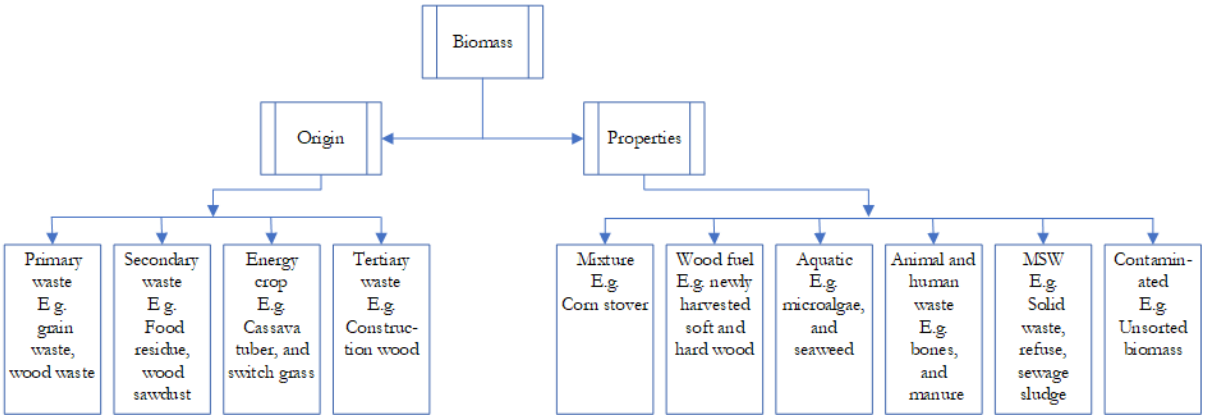

Figure 1. Biomass classification [11, 12]

3 METHODOLOGY

The essence of the classification is to apply the algorithm to information which comes from a known data source. This means that the class to which a data belong is evaluated as an input, from which it is transferred to higher dimensional space. In this new environment, the data is grouped with distance function based on the *k*-NN algorithm. Therefore, the data is classified into new groups to reflect the behaviour of the algorithm.

3.1 Data collection and classification

A total of 214 biomasses gathered from several trusted literature were first sorted based on properties and origin as proposed by Khan *et al.,* [11] the two categories were further sorted into eight classes. Class I consists of energy grasses and their parts (leaves and fibrous materials), class II consists of fruit residues and related plants, class III is the most diverse and it consist of wood, wood chips and pruning. Class IV is made of food crop residue, class V consist of biomass which were reported to have been subjected to pelletization process, since literature has shown that this process may substantially affect the properties of biomass due to the force of compaction [57, 58]. Class VI is

made of milling industry waste, class VII is an assortment of refuse and solid waste of municipal sources (MSW), this was sorted into a class due to their unpredictability and diverse composition [59]. Lastly, class VIII is made of biochar. The properties of biomass feedstocks were extracted from Jigisha *et al.,* [60, 61] and Nhuchhen *et al.,* [62]. To this end, 70 % of the whole dataset was used to train the model, while 30 % new dataset was used for model testing. The testing phase was meant to determine the capability of the *k*-NN model to generalize the classification of new biomass dataset. In order to address the detrimental consequence of overfitting, a stopping criterion was set. The network training stops automatically when the error reaches a stall value where significant reduction is not experienced.

Classification	Number of biomass samples	Classification based on Properties
I	43	Grasses/leaves/fibrous materials
II	38	Fruit residue—pit/shells/seeds
III	65	Wood and chips-barks/pruning/ wood/woodchips
IV	30	Food crop residue-Cob/hull/husk/ dust/ straws/stalks
V	5	Pelletized biomaterials
VI	7	Milling industry waste
VII	14	Refuse/MSW
VIII	12	Biomass chars

Table 2. Classification of biomass

3.2 Self-normalizing distance metrics

The performance of *k*-NN is substantially influenced by the selection of distance function [37, 39, 40]. However, different distance functions have different features. Several distance metrics have been applied as reported in section 2. There is a need to determine the distance function which produce the optimal result when applied to K-NN classifiers. Although several researches have shown that Euclidean distance function is most commonly used of distance metrics [63-65], Cover and Hart [29] posited that various distance functions can be tested to determine the optimal. Distance functions must be carefully selected in order to guarantee high accuracy of the classifier. For this intent, it is highly important to have distance metrics which capture the real behaviour of the biomass data. This study focuses on the self-normalizing distance metric, since it rules out the need for pre-normalisation which may be needed to eliminate data redundancy. There is no need to normalize the data, provided that the rounding errors used in inverting the covariance matrix are kept under control [66, 67]. This study evaluates the performance of Minkowski, Manhattan (city block), Euclidean and Mahalanobis distance functions applied to k-NN algorithms, therefore these distance functions are discussed as follows:

3.2.1 Minkowski distance function

Minkowski distance function is a metric which is in a vector space on which a norm is defined. It measures the geometric distance between two inputs using a variable scaling factor p. It can be calculated as follows;

$$MK(x,y) = \sqrt[p]{\sum_{j=1}^{n} \left|X_j - Y_j\right|^p} \tag{1}$$

Minkowski distance function can be manipulated to estimate the distance between two datapoints using different values of scaling factor p. For instance, when p=1, Manhattan distance is obtained, and when p=2 Euclidean distance is produced. These are discussed and formulated as follows:

3.2.2 Euclidean distance function

The Euclidean distance (EUD) is the real distance between two points that are connected by straight line. If $x = \{X_1, \ldots, X_n\}$ and $y = \{Y_1, \ldots, Y_n\}$ in an n-dimensional space, the Euclidean distance is given as;

$$EUD(x,y) = \sqrt{\sum_{j=1}^{n} \left|X_j - Y_j\right|^2} \tag{2}$$

3.2.3 Manhattan distance function

It is also called rectilinear distance, city block distance, or taxicab metrics. Manhattan distance function (MN) computes the distance between two data point following a grid-like path. It is also defined as the sum of the lengths of the projections of the line segment between the points onto the coordinate axes. If $x = \{X_1, \ldots, X_n\}$ and $y = \{Y_1, \ldots, Y_n\}$ in an n-dimensional space, the Manhattan distance is given as;

$$MN(x,y) = \sum_{j=1}^{n} \left|X_j - Y_j\right| \tag{3}$$

3.2.4 Mahalanobis Distance Function

The Mahalanobis metrics (DM) can be defined as the distance between an observation and the centre of each data group in an *n*-dimensional space defined by *n* variables and their covariances. For some classifications, Mahalanobis distance proved to be more efficient [68]. The problem of scale and correlation, which are associated with Euclidean distance function has been sufficiently addressed by Mahalanobis distance since it assumes an anisotropic gaussian distribution [69]. It does not depend on the unit of measurements of dataset [70]. It measures the separation of two groups of samples. It differs from Euclidean distance in that it takes into account the correlations of the data set and is scale-invariant [71]. Therefore, it is also called statistical distance [68], hence its relative advantage.

Suppose we have two groups with means $\bar{X}_j$ and $\bar{X}_k$, Mahalanobis distance function, DM, is given by:

$$DM_{(\bar{X}_j,\bar{X}_k)} = \sqrt{\left(\bar{X}_j - \bar{X}_k\right)^T} S^{-1}\left(\bar{X}_j - \bar{X}_k\right) \quad (4)$$

where S^{-1} is an inverse pooled covariance matrix S between $\bar{X}_j and\, \bar{X}_k$. This matrix is computed using weighted average of covariance matrices of both groups.

3.3 Modelling and Classification Algorithms

Based on the observation, a discrete class labels for an unclassified data can be achieved. Let $\{(x_1, y_1) \ldots . (x_{N,} y_N)\}$ represent the biomass dataset of q-dimension such that $X = \{x\}_{i=1}^{N} \subset \mathbb{R}$. and the corresponding traditional class label based on properties Y where $Y = \{y\}_{i=1}^{N} \subset \mathbb{R}$. The objective of the classification is to learn a model g such that a class label y' is predicted for an unknown pattern x'.
For a binary classifier, k-NN can be expressed as:

$$g_{KNN} = \begin{cases} 1 \;\; if \sum\limits_{i \in N_K(x')} y_i \geq 0 \\ -1 \;\; if \sum\limits_{i \in N_K(x')} y_i < 0 \end{cases} \quad (5)$$

For an instance of a multiclass classification, given an unknown pattern x', k-NN predicts the class label of the majority of k nearest pattern in data space such that:

$$g_{K-NN} = argmax_{y \in Y} \sum_{i \in N_K(x')} L(y_i = y) \quad (6)$$

It must be noted that the choice of k defines the locality of estimation in k-NN [40, 44, 72]. Therefore, the choice of k must be done carefully to avoid overfitting and underfitting of the dataset. Thus, several k values ranging from equations 1 to 4 were tested in this model. The distance weighted k-NN approach was experimented for different value of k in the biomass dataset. The selection of the optimum k was based on the minimum error rate.

3.4 The Model Algorithm

The realization of the algorithm which is used for biomass classification was achieved with MATLAB 2015 software installed on a desktop computer workstation with configuration 64 bits, 32GB RAM Intel (R) Core (TM) i7 5960X. The modelling steps involved in the formulation of the algorithm for the classification model for biomass properties based on k-NN is as highlighted:

Input parameter:

Biomass samples, q = 214 from various sources and with proximate properties.

Target output:

Biomass classification comprising eight classes (I, II, III, IV, V, VI, VII, VIII).

Start

Step 1: Initialization

Assigning of values and normalization of data matrix.

Step 2: Data Division

Division of dataset as 70% training and 30% testing data respectively.

Step 3: Training

Assigning of new label to an individual data point based on its k nearest neighbours.

Step 4: If any of the labels changed go to step 3 otherwise, proceed to step 5.

Step 5: Prediction
Assign labels to test data based on its k nearest neighbour.
Step 6: Forecast Comparison
Compare the predicted labels with actual labels and determine the error rate.
Step 7: Change the test set and go to step 3

Step 8. Determine the mean error rate after the stopping criterion is fulfilled.

End

3.5 Model Performance Assessment

The model evaluation was based on Root Mean Square Error (RMSE) and some classification based metrics which applies confusion matrix [73, 74]. These include Accuracy (Acc), Error (ERR), Specificity (TNR), Sensitivity (TPR), False positive rate (FPR) and *Kappa statistics.* The choice of RMSE is based on its use in numerous studies as an effective means of determining the eligibility of the model for prediction. Computation Time (CT) was also used to determine the rate of convergent of the iteration and the speed at which the computation is achieved. Given that y_k and $\widehat{y_k}$ for k=1,2,3…n is the predicted and observed biomass value respectively, the performance of the classification model was determined using the following measures:

Root Mean square Error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N}[y_k - \widehat{y_k}]}{N}} \tag{7}$$

Accuracy (Acc)

It is one the most commonly used measure of the performance of a classification model. It is the ratio between correctly classified data to the overall dataset. It is used to analyse and compare the quality of the classification. Under real life scenario, the cost of misclassification does vary, therefore, the level of accuracy required depend on the severity of the consequence or the impact of misclassification.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

Error rate (ERR)

It is a complement of accuracy which represent the misclassification rate or the amount of data that are not correctly classified.

$$ERR = 1 - Acc = \frac{FP + FN}{TP + TN + FP + FN} \tag{9}$$

Sensitivity (TPR)

It is also known as True positive rate (TPR), hit rate or recall. It is the ratio of number of positive and correctly classified samples to the overall number of positive samples. In this case, it the number of positive correctly predicted biomass samples in a class to the total number of biomass samples in that class. It is given as follows;

$$TPR = \frac{TP}{TP + FN} \tag{10}$$

Specificity (TNR)

It is also called True negative rate (TNR),or inverse recall and it is expressed as the ratio of correctly classified negative samples to the overall number of negative samples. In this case, it is the number of negative correctly predicted biomass samples in a class to the total number of biomass samples which are not supposed to be within that class. It is given as follows;

$$TNR = \frac{TN}{FP + TN} \tag{11}$$

False positive rate (FPR)

It is also known as fallout and it is defined as the ratio of incorrectly classified negative samples to the total number of negative samples. This performance metrics complements the specificity.

$$FPR = 1 - TNR = \frac{FP}{TN + FP} \tag{12}$$

Kappa statistic s(κ)

It is a measure of the true agreement that is achieved beyond that which is achieved by chance. This metric is applied to report the quantitative measure of agreement between different observations. Kappa statistics can be applied to binary or multiclass classification problems. It is given as;

$$\kappa = \frac{P_o - P_c}{1 - P_c} \tag{13}$$

Where P_o is the proportion of observed agreement and P_c is the proportion of agreement obtained by chance. The common referenced scale for the interpretation of kappa result is given by Landis and Koch [75] and it is shown in Table 3. Kappa value of 1 means there is a perfect agreement between the observations while a kappa value of 0 means there is no agreement. Although this scale may not be absolute as it was not supported by any empirical evidence, it has been applied in several studies and in different fields especially in medical science [76-79].

Table 3: Reference scale for kappa

Kappa	**Agreement**
$< \mathbf{0}$	Less than chance agreement or poor agreement
$\mathbf{0.01 - 0.2}$	Slight agreement
0.21-0.4	Fair agreement
0.41-0.6	Moderate agreement
0.61-0.8	Substantial agreement

0.81-0.99	Almost perfect agreement

Source [75]

4. RESULTS AND DISCUSSIONS

The proposed *k*-NN method employs the 70 % of the biomass dataset as the training data. The training data was fed into the model to determine its classification ability, for *k* ranging from 1 to 4 based on same set of data. This was done since different value of *k* produces different error rates and accuracy. All the performance assessment methods were computed from confusion matrix [73, 74] which were coded in MATLAB environment.

4.1 Evaluation of Self normalizing metrics

Figure 2 and Figure 3 show the overall classification accuracy for Minkowski, Manhattan, Euclidean and Mahalanobis distance functions based on *k* ranging from 1 to 4 for training and testing data respectively. Expectedly the error rate at *k*=1 is close to zero (ERR=0.067) for training data which means the accuracy of the model is approximately 100 %. This is because, for any given training data, its nearest neighbour will always be itself. This implies that there is almost around zero distance between a given data point and its neighbour. Had it been the accuracy of the validation curve follow similar pattern, the choice of *k* would have unarguably become 1. Another consequence of *k*=1 is an instance of boundary overfitting. That is what make the error rate to reach a minimum value and subsequently increase with increase in *k* value. Therefore, it would be wrong to assume that *k*=1 gives the optimal model.

It is observed that the accuracy of the Euclidean and Minkowski distance metrics were the same at both training and testing phase. This agrees with Chomboon *et al.* [37] regarding the similarity of accuracy obtained for Euclidean and Minkowski distance metrics based on their hypothetical data. The result obtained for this two were closely followed by Manhattan though with a slight margin. Based on the evaluated performance indices, optimum K value were obtained at k=3 for all distance metrics. Therefore, the RMSE, for all distance metric were presented in Figure 4. Mahalanobis function shown the most obvious difference with higher accuracy (73 % and 70.3 % for training and testing respectively) and the least RMSE (1.547 and 1.42 for training and testing respectively) compared to all other metrics which were investigated. In view of this, further discussion is based on Mahalanobis distance function since it produced the optimal performance of all the metrics.

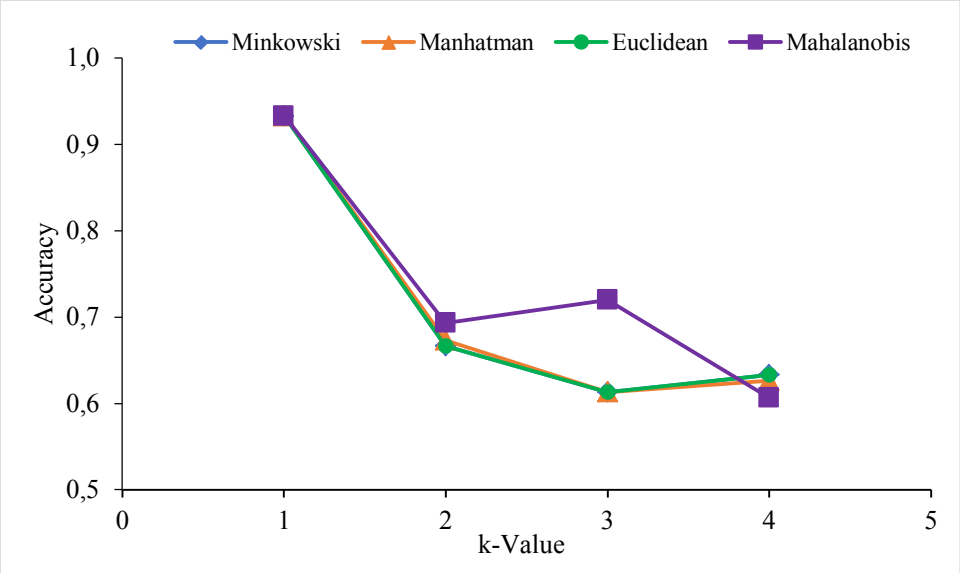

Figure 2:Comparison of classification accuracy for various distance metrics at training stage

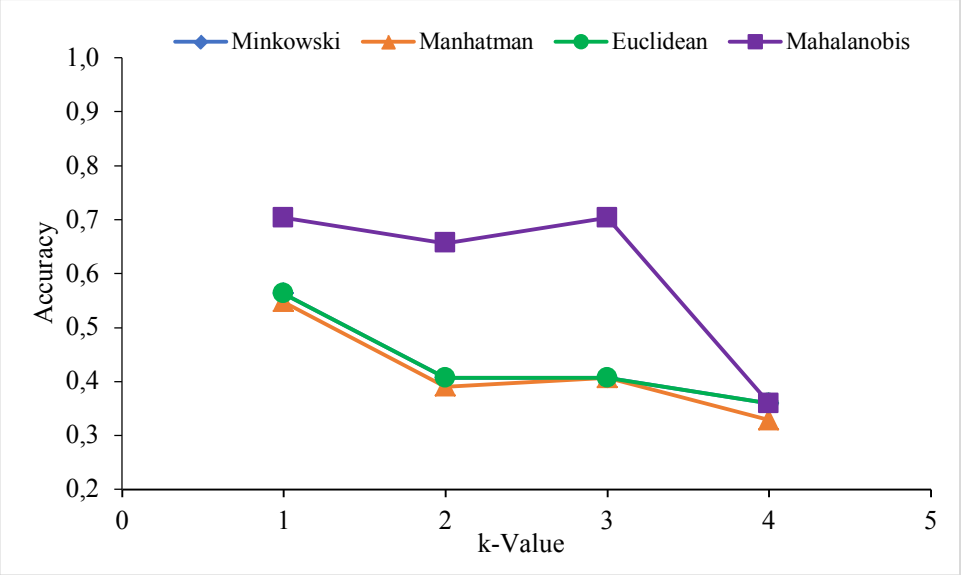

Figure 3: Comparison of classification accuracy for various distance metrics at testing stage

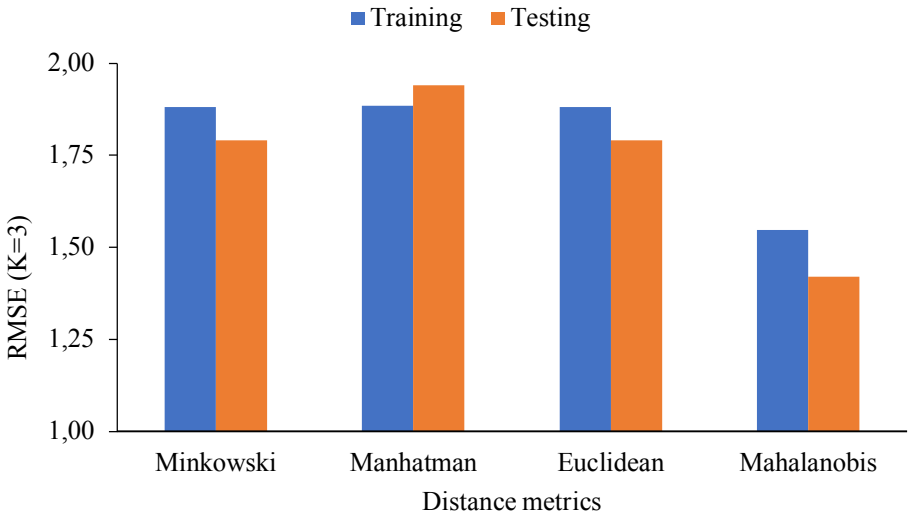

Figure 4:Optimim K value for different distance functions

4.2 Classification of Biomass based on optimum distance function

As stated in the previous section, Mahalanobis distance function produced the optimal performance based on various metrics investigated. In order to determine the optimal value of *k*, validation error and training error as defined by RMSE was plotted against *k* values as shown in Figure 5. It is noted that the optimum *k* value on the validation dataset coincides with the turning point on the training data curve which lies between the K values of 2 and 3. The *k* which gives the minimum RMSE on the validation curve is taken as the optimum *k* value which in this case is *k* =3, and it is considered for all predictions.

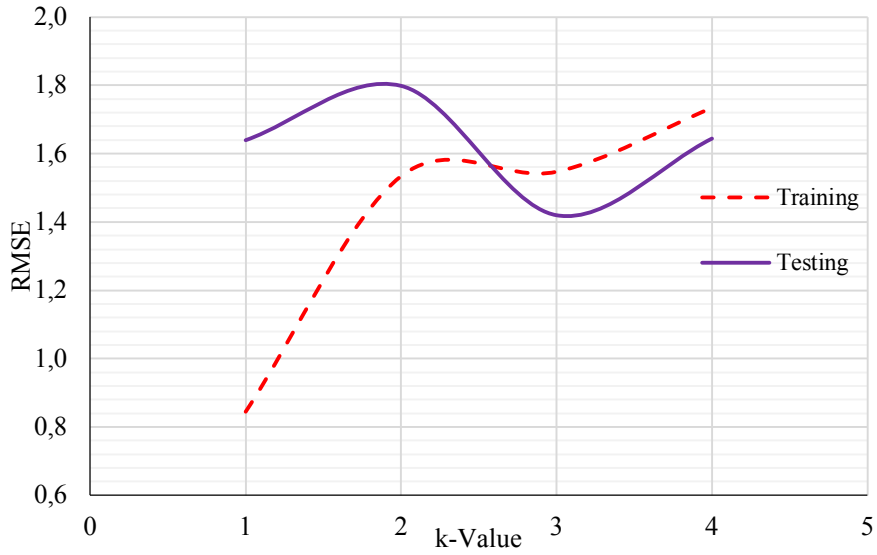

Figure 5:Optimum value of K for Mahalanobis function

With *k*=3 identified as the optimal *k* value, the trained model was first subjected to the same training data comprising 150 biomass datasets. As observed in Figure 6, the biomass samples in Class I, II, III, IV, V, VII, VIII yielded good prediction relative to the experimental biomass classes. In contrast, the biomass in class VI were poorly predicted. This may be due to sparse data which were in that class [80, 81]. For model evaluation with new datasets, a biomass dataset consisting of 64 new data that have not been previously applied during the training phase were utilized. As shown in Figure 7, the testing phase reflects the actual biomass classification. The 33 % of class VII were poorly classified while 18 % of the data in class VIII was wrongly predicted. The effectiveness of the classification process for these classes may be improved by introducing a larger dataset comprising nearly same number of datasets from all biomass classes. Interestingly, the agreement between the predicted value and actual biomass class which were correctly classified is very satisfactory. The biomass classes which were largely represented in the database produced better classification (classes I II and III). This means that the classification model may improve with larger dataset having more evenly distributed biomass feedstock in different classes [80].

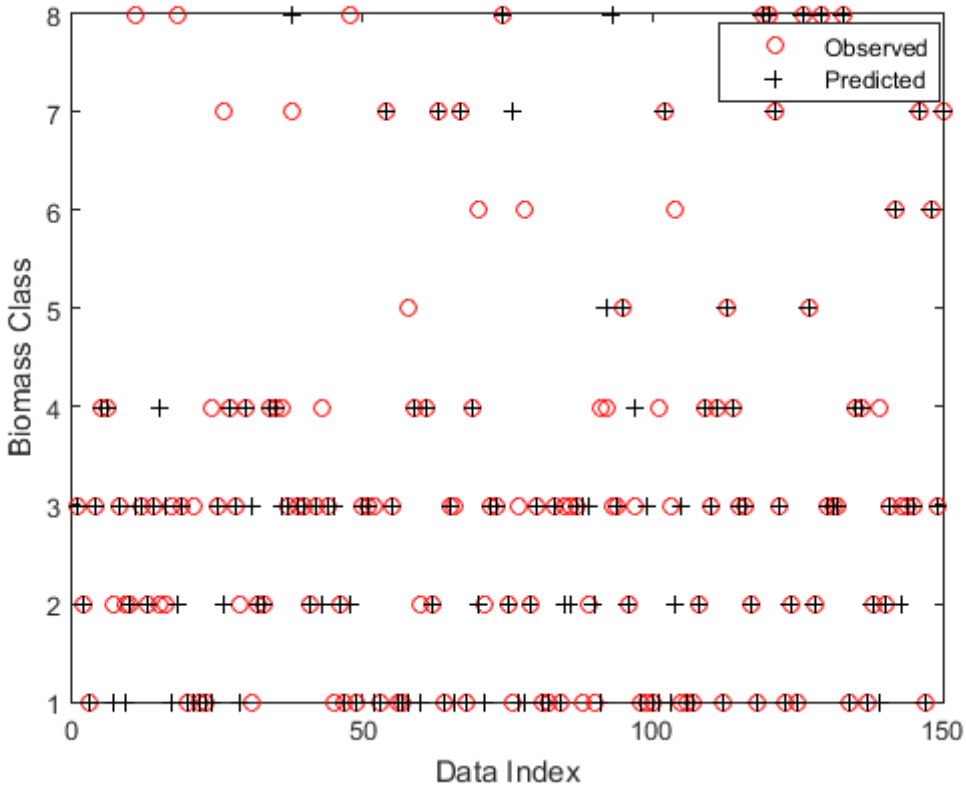

Figure 6: Prediction for Mahalanobis function based on the optimum training dataset at k=3

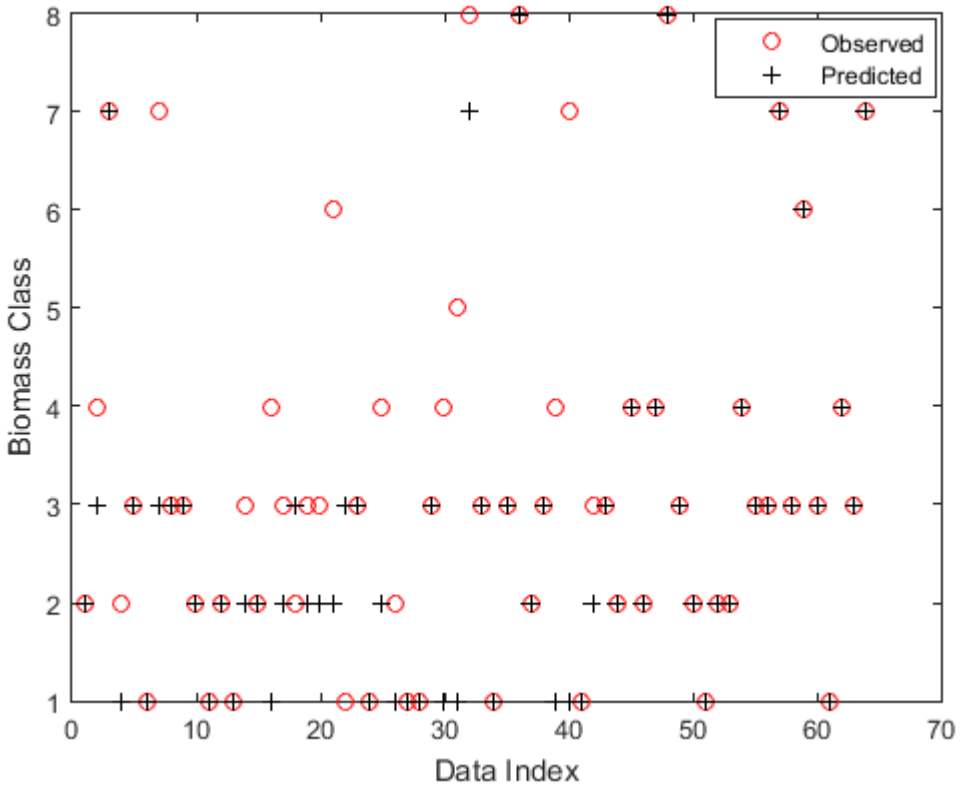

Figure 7:Prediction for Mahalanobis function based on the optimum test dataset at k=3

Figure 8 shows the percentage of biomass samples in each class based on the data applied at model validation stage. The result shows that the test sample was divided into seven distinct classes with I, II, III, IV, VI VII, VII been 26.6 %, 26.6 %, 29.7 %, 6.2 %, 1.6 %, 6.2 %, and 3.1 % respectively. Overall, Class III account for 29.7 % of biomass samples predicted in the classification while no sample was classified in class V. The poor prediction of class V maybe due to small data size associated with the class or the presence of outliers [82]. Shahshahani and Landgrebe [83] and Perez-Ortiz *et al.* [84] emphasized that when the dataset in a class is small, the classification result may not be satisfactory for that class.

The implication of these classification is that the model can aggregate the biomass feedstock based on the fuel properties and sources. The prominence of the feedstock is Class I to III suggests the relative importance of biomass residue in energy generation from biomass. Most agricultural processes produce some level of waste which can be applied in renewable energy generation through various conversion processes [53, 85-87].

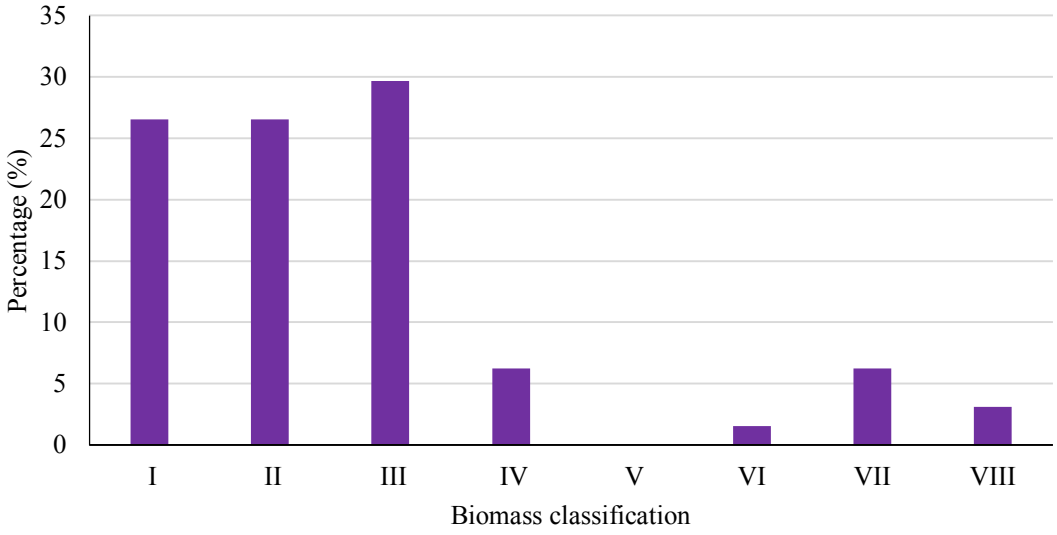

Figure 8:Percentage biomass classification

4.3 Measurement of model accuracy

Statistical evaluation and measure of computational time of the model training and testing phases gave the results presented in Table 2 and 3. The accuracy of the model was solely based on the statistical performance which are relevant to classification models since there is no known article in the literature which has utilized *k*-NN for the classification of biomass feedstock. The RMSE, Accuracy, misclassification, Sensitivity, Specificity, and FPR are reported in table 2 and 3. As previously stated, the optimum RMSE and accuracy were obtained at k =3. In overall, 70 % of the feedstock were correctly classified at the validation stage. FPR value of 0.047 at k =3 implies that there is 4.7 % chance that the model will misclassify the class, hence as the FPR increases, the accuracy of the model decreases. Also, Kappa statistics was used to measure the level of agreement between the observed number of biomass feedstock in a class and the number that were predicted for that class. Kappa value of 0.62 on the scale of 0 to 1 [75] suggest that there is substantial agreement between the actual classes and the predicted classes. This further lend a credence to the accuracy of 70 % which was obtained. The rate of convergence of the model was determined from the computation time. It was observed that the computational time decreases with increase in k -value.

Table 4:Training error assessment for different k values

	k=1	*k* =2	*k* =3	*k* =4
RMSE	0.845	1.534	1.547	1.734
Accuracy	0.933	0.693	0.720	0.607
Error	0.067	0.307	0.280	0.393
Sensitivity	0.910	0.600	0.685	0.496
Specificity	0.990	0.948	0.956	0.938
FPR	0.010	0.052	0.044	0.062
Kappa	0.695	0.357	0.414	0.444

Table 5:Testing error assessment for different k values

	k=1	*k* =2	*k* =3	*k* =4
RMSE	1.639	1.798	1.420	0.000
Accuracy	0.703	0.656	0.703	0.359
Error	0.438	0.594	0.297	0.641
Sensitivity	0.491	0.293	0.580	0.274
Specificity	0.931	0.909	0.953	0.898
FPR	0.069	0.091	0.047	0.102
Kappa	0.500	0.632	0.622	0.659
CT(s)	7.73	5.9	4.7	4.3

5. CONCLUSIONS

In this study, we have presented a novel approach to the classification of biomass feedstock. Our model is a proof of concept and a prototype of idea for using scientific methods such as *k*-NN in the classification of biomass data. The developed biomass classification model was assessed using some statistical parameters. The model developed was tested with new 64 data sets. Further studies should be conducted with larger biomass data which sufficiently cover all the biomass classes identified in this study. Also, other data mining techniques can be explored. This will enable more efficient classification, which could assist in the hybridization of biomass sources for energy generation. The model produced can effectively assist in the assortment of different biomass resources. This model will assist the power plant design engineers in deciding the biomass which can be combined for energy generation purpose.

References

[1] U. N. D. o. P. Information, "Sustainable Development Goals," *https://sustainabledevelopment.un.org/?menu=1300 acessed on 23rd July 2018,* 2015.

[2] A. Müller *et al.*, "The role of biomass in the Sustainable Development Goals: A reality check and governance implications," *Potsdam: Institute for Advanced Sustainability Studies (IASS),* 2015.
[3] H. Garg and G. Datta, "Global status on renewable energy," in *Solar energy heating and cooling methods in building, international workshop: Iran University of Science and Technology*, 1998, pp. 19-20.
[4] T. Kar and S. Keles, "Environmental impacts of biomass combustion for heating and electricity generation," *Journal of Engineering Research and Applied Science,* vol. 5, no. 2, pp. 458-465, 2016.
[5] M. Balat and G. Ayar, "Biomass energy in the world, use of biomass and potential trends," *Energy sources,* vol. 27, no. 10, pp. 931-940, 2005.
[6] M. F. Demirbas, M. Balat, and H. Balat, "Potential contribution of biomass to the sustainable energy development," *Energy Conversion and Management,* vol. 50, no. 7, pp. 1746-1760, 2009.
[7] R. D. Perlack, *Biomass as feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion-ton annual supply*. Oak Ridge National Laboratory, 2005.
[8] M. Guo, W. Song, and J. Buhain, "Bioenergy and biofuels: History, status, and perspective," *Renewable and Sustainable Energy Reviews,* vol. 42, pp. 712-725, 2015.
[9] J. B. Sluiter, R. O. Ruiz, C. J. Scarlata, A. D. Sluiter, and D. W. Templeton, "Compositional analysis of lignocellulosic feedstocks. 1. Review and description of methods," *J Agric Food Chem,* vol. 58, no. 16, pp. 9043-53, Aug 25 2010.
[10] S. Nanda, J. A. Kozinski, and A. Dalai, *Biomass-an overview on classification, composition and characterization*. 2013, pp. 1-35.
[11] A. Khan, W. De Jong, P. Jansens, and H. Spliethoff, "Biomass combustion in fluidized bed boilers: potential problems and remedies," *Fuel processing technology,* vol. 90, no. 1, pp. 21-50, 2009.
[12] S. V. Vassilev, D. Baxter, L. K. Andersen, and C. G. Vassileva, "An overview of the chemical composition of biomass," *Fuel,* vol. 89, no. 5, pp. 913-933, 2010.
[13] B. R. Hough, D. A. C. Beck, D. T. Schwartz, and J. Pfaendtner, "Application of machine learning to pyrolysis reaction networks: Reducing model solution time to enable process optimization," *Computers & Chemical Engineering,* vol. 104, pp. 56-63, 2017.
[14] A. Sahai, "Evaluation of Machine Learning Techniques for Green Energy Prediction," *arXiv preprint arXiv:1406.3726,* 2014.
[15] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Annals of translational medicine,* vol. 4, no. 11, 2016.
[16] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research,* vol. 5, no. Aug, pp. 975-1005, 2004.
[17] C. Crisci, B. Ghattas, and G. Perera, "A review of supervised machine learning algorithms and their applications to ecological data," *Ecological Modelling,* vol. 240, pp. 113-122, 2012.
[18] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, no. 3, pp. 273-297, 1995.
[19] O. O. Olatunji, O. O. Ayo, S. Akinlabi, F. Ishola, N. Madushele, and P. A. Adedeji, "Competitive advantage of carbon efficient supply chain in manufacturing industry," *Journal of Cleaner Production,* vol. 238, p. 117937, 2019/11/20/ 2019.
[20] M. Talaei, B. F. Moghaddam, M. S. Pishvaee, A. Bozorgi-Amiri, and S. Gholamnejad, "A robust fuzzy optimization model for carbon-efficient closed-loop supply chain network design problem: a numerical illustration in electronics industry," *Journal of Cleaner Production,* vol. 113, pp. 662-673, 2016.
[21] O. O. Olatunji, S. Akinlabi, N. Madushele, and P. A. Adedeji, "Estimation of the Elemental Composition of Biomass Using Hybrid Adaptive Neuro-Fuzzy Inference System," *BioEnergy Research,* pp. 1-11, 2019.
[22] O. Olatunji, S. Akinlabi, N. Madushele, and P. A. Adedeji, "Estimation of Municipal Solid Waste (MSW) combustion enthalpy for energy recovery," 2019.
[23] K. C. Drudi, R. Drudi, G. Martins, G. C. Antonio, and J. T. C. Leite, "Statistical model for heating value of municipal solid waste in Brazil based on gravimetric composition," *Waste Management,* vol. 87, pp. 782-790, 2019.
[24] I. Boumanchar *et al.*, "Municipal solid waste higher heating value prediction from ultimate analysis using multiple regression and genetic programming techniques," *Waste Management & Research,* p. 0734242X18816797, 2018.
[25] P. A. Adedeji, S. Akinlabi, O. Ajayi, and N. Madushele, "Non-linear autoregressive neural network (NARNET) with SSA filtering for a university energy consumption forecast," *Procedia Manufacturing,* vol. 33, pp. 176-183, 2019.
[26] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.
[27] A. Li and Y. Shi, "An integrated classification method: combination of LP and LDA," in *International Workshop on Internet and Network Economics*, 2005, pp. 758-767: Springer.
[28] H. Bhavsar and M. H. Panchal, "A review on support vector machine for data classification," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET),* vol. 1, no. 10, pp. pp: 185-189, 2012.
[29] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory,* vol. 13, no. 1, pp. 21-27, 1967.
[30] V. Surya Prasath, H. Arafat Abu Alfeilat, O. Lasassmeh, and A. Hassanat, "Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier-A Review," *arXiv preprint arXiv:1708.04321,* 2017.
[31] D. Adeniyi, Z. Wei, and Y. Yongquan, "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method," *Applied Computing and Informatics,* vol. 12, no. 1, pp. 90-108, 2016.
[32] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, 2003, pp. 986-996: Springer.

[33] G. Tao, T. A. Lestander, P. Geladi, and S. Xiong, "Biomass properties in association with plant species and assortments I: A synthesis based on literature data of energy properties," *Renewable and Sustainable Energy Reviews,* vol. 16, no. 5, pp. 3481-3506, 2012.
[34] G. Tao, P. Geladi, T. A. Lestander, and S. Xiong, "Biomass properties in association with plant species and assortments. II: A synthesis based on literature data for ash elements," *Renewable and Sustainable Energy Reviews,* vol. 16, no. 5, pp. 3507-3522, 2012.
[35] Z. Markov and D. T. Larose, "Data Mining the Web: Uncovering Patterns in Web Content," 2007.
[36] E. Leif and K.-N. Neighbor, "Scholarpedia 4 (2): 1883," ed, 2009.
[37] K. Chomboon, P. Chujai, P. Teerarassamee, K. Kerdprasop, and N. Kerdprasop, "An empirical study of distance metrics for k-nearest neighbor algorithm," in *Proceedings of the 3rd International Conference on Industrial Application Engineering*, 2015.
[38] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *Computational Intelligence for Security and Defense Applications, 2009. CISDA 2009. IEEE Symposium on*, 2009, pp. 1-6: IEEE.
[39] P. Mulak and N. Talhar, "Analysis of Distance Measures Using K-Nearest Neighbor Algorithm on KDD Dataset," *International Journal of Science and Research,* vol. 4, no. 7, pp. 2101-2104, 2015.
[40] L.-Y. Hu, M.-W. Huang, S.-W. Ke, and C.-F. Tsai, "The distance function effect on k-nearest neighbor classification for medical datasets," *SpringerPlus,* vol. 5, no. 1, p. 1304, 2016.
[41] R. Todeschini, D. Ballabio, V. Consonni, and F. Grisoni, "A new concept of higher-order similarity and the role of distance/similarity measures in local classification methods," *Chemometrics and Intelligent Laboratory Systems,* vol. 157, pp. 50-57, 2016.
[42] R. Todeschini, D. Ballabio, and V. Consonni, "Distances and other dissimilarity measures in chemometrics," *Encyclopedia of analytical chemistry,* pp. 1-34, 2015.
[43] A. K. Gjertsen, "Accuracy of forest mapping based on Landsat TM data and a kNN-based method," *Remote Sensing of Environment,* vol. 110, no. 4, pp. 420-430, 2007.
[44] X. Peng, Y. Cai, Q. Li, and K. Wang, "Control rod position reconstruction based on K-Nearest Neighbor Method," *Annals of Nuclear Energy,* vol. 102, pp. 231-235, 2017.
[45] V. Prasath, H. A. A. Alfeilat, O. Lasassmeh, and A. Hassanat, "Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier-A Review," *arXiv preprint arXiv:1708.04321,* 2017.
[46] A. Joshi and A. Mehta, "ANALYSIS OF K-NEAREST NEIGHBOR TECHNIQUE FOR BREAST CANCER DISEASE CLASSIFICATION," *Machine Learning,* vol. 98, p. 13, 2018.
[47] L. Nunes, J. Matias, and J. Catalão, "Biomass combustion systems: A review on the physical and chemical properties of the ashes," *Renewable and Sustainable Energy Reviews,* vol. 53, pp. 235-242, 2016.
[48] J. Cai *et al.*, "Review of physicochemical properties and analytical characterization of lignocellulosic biomass," *Renewable and Sustainable Energy Reviews,* vol. 76, pp. 309-322, 2017.
[49] J. M. Vargas-Moreno, A. J. Callejón-Ferre, J. Pérez-Alonso, and B. Velázquez-Martí, "A review of the mathematical models for predicting the heating value of biomass materials," *Renewable and Sustainable Energy Reviews,* vol. 16, no. 5, pp. 3065-3083, 2012.
[50] Y. D. Singh, P. Mahanta, and U. Bora, "Comprehensive characterization of lignocellulosic biomass through proximate, ultimate and compositional analysis for bioenergy production," *Renewable Energy,* vol. 103, pp. 490-500, 2017.
[51] W. de Jong, "Biomass Composition, Properties, and Characterization," *Biomass as a Sustainable Energy Source for the Future: Fundamentals of Conversion Processes,* pp. 36-68, 2014.
[52] D. L. Klass, *Biomass for renewable energy, fuels, and chemicals*. Elsevier, 1998.
[53] O. Obafemi, A. Stephen, O. Ajayi, P. Mashinini, and M. Nkosinathi, "Experimental investigation of thermal properties of Lignocellulosic biomass: A review," *IOP Conference Series: Materials Science and Engineering,* vol. 413, no. 1, p. 012054, 2018.
[54] J. S. Tumuluru, *Biomass Preprocessing and Pretreatments for Production of Biofuels: Mechanical, Chemical and Thermal Methods*. CRC Press, 2018.
[55] H. Singh and S. Mohapatra, "Study of Emission Characteristics and Noise of Dual Fuel Engine Run on Blends of Diesel and Producer Gas from Biomass Materials," THAPAR UNIVERSITY, PATIALA, 2016.
[56] M. Wang *et al.*, "To distinguish the primary characteristics of agro-waste biomass by the principal component analysis: An investigation in East China," *Waste Management,* vol. 90, pp. 100-120, 2019.
[57] W. Stelte *et al.*, "Pelletizing properties of torrefied spruce," *biomass and bioenergy,* vol. 35, no. 11, pp. 4690-4698, 2011.
[58] R. García, M. Gil, F. Rubiera, and C. Pevida, "Pelletization of wood and alternative residual biomass blends for producing industrial quality pellets," *Fuel,* vol. 251, pp. 739-753, 2019.
[59] S. Dutta, D. N. Priya, B. Chakradhar, and T. S. Jyothsna, "Value Added By-products Recovery from Municipal Solid Waste," in *Waste Valorisation and Recycling*: Springer, 2019, pp. 71-80.
[60] J. Parikh, S. Channiwala, and G. Ghosal, "A correlation for calculating HHV from proximate analysis of solid fuels," *Fuel,* vol. 84, no. 5, pp. 487-494, 2005.
[61] J. Parikh, S. A. Channiwala, and G. K. Ghosal, "A correlation for calculating elemental composition from proximate analysis of biomass materials," *Fuel,* vol. 86, no. 12-13, pp. 1710-1719, 2007.
[62] D. R. Nhuchhen and P. Abdul Salam, "Estimation of higher heating value of biomass from proximate analysis: A new approach," *Fuel,* vol. 99, pp. 55-63, 2012.
[63] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research,* vol. 10, no. Feb, pp. 207-244, 2009.

[64] M. Wölfel and H. K. Ekenel, "Feature weighted Mahalanobis distance: improved robustness for Gaussian classifiers," in *2005 13th European signal processing conference*, 2005, pp. 1-4: IEEE.
[65] J. Walters-Williams and Y. Li, "Comparative study of distance functions for nearest neighbors," in *Advanced Techniques in Computing Sciences and Software Engineering*: Springer, 2010, pp. 79-84.
[66] D. H. Besset, *Object-oriented implementation of numerical methods: An introduction with Java and Smalltalk*. Morgan Kaufmann, 2001.
[67] E. Eyob, *Social Implications of Data Mining and Information Privacy: Interdisciplinary Frameworks and Solutions: Interdisciplinary Frameworks and Solutions*. IGI Global, 2009.
[68] V. K. Pathirana, "Nearest Neighbor Foreign Exchange Rate Forecasting with Mahalanobis Distance," 2015.
[69] "Mahalanobis Distance," in *Encyclopedia of Biometrics*, S. Z. Li and A. Jain, Eds. Boston, MA: Springer US, 2009, pp. 953-953.
[70] Z. Diao and Y. Wu, "A new SVM decision tree multi-class classification algorithm based on Mahalanobis distance," in *Control Conference (CCC), 2011 30th Chinese*, 2011, pp. 3124-3127: IEEE.
[71] G. Madzarov and D. Gjorgjevikj, "Evaluation of distance measures for multi-class classification in binary svm decision tree," in *International Conference on Artificial Intelligence and Soft Computing*, 2010, pp. 437-444: Springer.
[72] M. H. Bhavsar and A. Ganatra, "Support Vector Machine Classification using Mahalanobis Distance Function," *space,* vol. 6, p. 7.
[73] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition,* vol. 91, pp. 216-231, 2019.
[74] A. Hay, "The derivation of global estimates from a confusion matrix," *International Journal of Remote Sensing,* vol. 9, no. 8, pp. 1395-1398, 1988.
[75] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *biometrics,* pp. 159-174, 1977.
[76] F. Espinosa-Ortega *et al.*, "Comparison of autoantibody specificities tested by a line blot assay and immunoprecipitation-based algorithm in patients with idiopathic inflammatory myopathies," *Annals of the rheumatic diseases,* pp. annrheumdis-2018-214690, 2019.
[77] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: the kappa statistic," *Fam med,* vol. 37, no. 5, pp. 360-363, 2005.
[78] E. C. Cheshire, M. J. Biggs, F. E. Hollingbury, V. L. Fitzpatrick-Swallow, T. R. Prickett, and R. D. Malcomson, "Frequency of macroscopic intradural hemorrhage with and without subdural hemorrhage in early childhood autopsies," *Forensic Science, Medicine and Pathology,* pp. 1-7, 2019.
[79] A. Mustaqeem, S. M. Anwar, and M. Majid, "Multiclass classification of cardiac arrhythmia using improved feature selection and SVM invariants," *Computational and mathematical methods in medicine,* vol. 2018, 2018.
[80] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: Improving classification performance when training data is skewed," in *2008 19th International Conference on Pattern Recognition*, 2008, pp. 1-4: IEEE.
[81] Y. Qian, Y. Liang, M. Li, G. Feng, and X. Shi, "A resampling ensemble algorithm for classification of imbalance problems," *Neurocomputing,* vol. 143, pp. 57-67, 2014.
[82] T. Zhang, J. Jin, S. Yang, D. Hu, G. Li, and J. Jiang, "Synthesis and characterization of fluorinated PBO with high thermal stability and low dielectric constant," *Journal of Macromolecular Science, Part B,* vol. 48, no. 6, pp. 1114-1124, 2009.
[83] B. M. Shahshahani and D. A. Landgrebe, "The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon," *IEEE Transactions on Geoscience and remote sensing,* vol. 32, no. 5, pp. 1087-1095, 1994.
[84] M. Perez-Ortiz, P. Tino, R. Mantiuk, and C. Hervas-Martinez, "Exploiting Synthetically Generated Data with Semi-Supervised Learning for Small and Imbalanced Datasets," *arXiv preprint arXiv:1903.10022,* 2019.
[85] O. O. Olatunji, S. A. Akinlabi, M. P. Mashinini, S. O. Fatoba, and O. O. Ajayi, "Thermo-gravimetric characterization of biomass properties: A review," *IOP Conference Series: Materials Science and Engineering,* vol. 423, no. 1, p. 012175, 2018.
[86] K. A. Motghare, A. P. Rathod, K. L. Wasewar, and N. K. Labhsetwar, "Comparative study of different waste biomass for energy application," *Waste Manag,* vol. 47, no. Pt A, pp. 40-5, Jan 2016.
[87] S. Kamel, H. A. El-Sattar, D. Vera, and F. Jurado, "Bioenergy potential from agriculture residues for energy generation in Egypt," *Renewable and Sustainable Energy Reviews,* vol. 94, pp. 28-37, 2018.