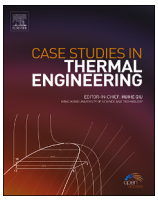
Case Studies in Thermal Engineering 60 (2024) 104743

Contents lists available at ScienceDirect

Case Studies in Thermal Engineering

journal homepage: www.elsevier.com/locate/csite



# Development of advanced machine learning for prognostic analysis of drying parameters for banana slices using indirect solar dryer

Van Giao Nguyen [a], Prabhu Paramasivam [b, **], Marek Dzida [c], Sameh M. Osman [d], Duc Trong Nguyen Le [e, ***], Dao Nam Cao [f], Thanh Hai Truong [f], Viet Dung Tran [f, *]

[a] Institute of Engineering, HUTECH University, Ho Chi Minh City, Viet Nam
[b] Department of Research and Innovation, Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu, 602105, India
[c] Institute of Naval Architecture, Gdańsk University of Technology, Gdansk, Poland
[d] Chemistry Department, College of Science, King Saud University, P.O. Box 2455, Riyadh, 11451, Saudi Arabia
[e] Faculty of Automotive Engineering, Dong A University, Danang, Viet Nam
[f] PATET Research Group, Ho Chi Minh City University of Transport, Ho Chi Minh City, Viet Nam

## HIGHLIGHTS

- XGBoost and LighGBM were employed to model-predict the banana chips drying process.
- Prediction results reveal that XGBoost with $R^2 = 0.9971$ was superior to LightGBM.
- Game theory-based Shapley functions were employed for feature analysis.
- XAI revealed that temperature and product types are the most influential features.

## ARTICLE INFO

*Keywords:*
Banana slice
Solar drying
Machine learning
XGBoost
LightGBM
Energy consumption

## ABSTRACT

In this study, eXtreme Gradient Boosting (XGBoost) and Light Gradient Boosting (LightGBM) algorithms were used to model-predict the drying characteristics of banana slices with an indirect solar drier. The relationships between independent variables (temperature, moisture, product type, water flow rate, and mass of product) and dependent variables (energy consumption and size reduction) were established. For energy consumption, XGBoost demonstrates superior performance with an $R^2$ of 0.9957 during training and 0.9971 during testing, alongside minimal MSE of 0.0034 during training and 0.0008 during the testing phase indicating high predictive accuracy and low error rates. Conversely, LGBM shows lower $R^2$ values (0.9061 training, 0.8809 testing) and higher MSE of 0.0747 during training and 0.0337 during testing, reflecting poorer performance. Similarly, for shrinkage prediction, XGBoost outperforms LGBM, evidenced by higher $R^2$ (0.9887 training, 0.9975 testing) and lower MSE (0.2527 training, 0.4878 testing). The comparative statistics showed that XGBoost regularly outperformed LightGBM. The game theory-based Shapley functions revealed that temperature and product types were the most influential features of the energy consumption model. These findings illustrate the practical applicability of the XGBoost and LightGBM models in food drying operations towards optimizing drying conditions, improving product quality, and reducing energy consumption.

* Corresponding author. PATET Research Group, Ho Chi Minh City University of Transport, Ho Chi Minh City, Viet Nam.
** Corresponding author. Department of Research and Innovation, Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu, India.
*** Corresponding author. Faculty of Automotive Engineering, Dong A University, Danang, Viet Nam.
*E-mail addresses:* lptprabhu@gmail.com (P. Paramasivam), nguyenldt@donga.edu.vn (D.T.N. Le), dungtv@ut.edu.vn (V.D. Tran).

https://doi.org/10.1016/j.csite.2024.104743
Received 27 February 2024; Received in revised form 20 June 2024; Accepted 22 June 2024
Available online 23 June 2024

## 1. Introduction

Climate change is one of the major problems for humankind at present [1], thus achieving the targets set in the sustainable development goals (SDGs) is a major challenge [2]. It is observed that climate change may lead to a reduction in agricultural yields by 2030 to a certain extent [3], resulting in unpredictable consequences for food safety [4]. Climate change caused by greenhouse gases (GHGs) is already negatively affecting food production systems [5], in which the effects on food systems are predicted to be vast, complicated, and geographically and temporally diverse [6,7]. Due to this reason, the use of green and clean energy [8] and zero-carbon fuels [9] is considered as one of the leading solutions to address climate change-related global problems and mitigate GHG. Among the green and clean energy sources, solar energy is considered as a promising and potential renewable source [10] because the solar energy is the most abundant energy on earth [11]. Solar energy could be utilized to produce heat [12], electricity [13], and hydrogen [14]. In addition, it could be used for transportation [15], food processing [16], and water distillation applications [17]. For preserving the process of fruits and vegetables, the power of the sun has been employed to preserve grain and other farming items [18]. Crops are traditionally dried under the open sky using sunlight. This method of food preservation is popular and cost-effective as opposed to thermal drying [19], it is thus still widely employed in many developing countries [20]. If the drying is not completed properly, the quantity and quality of the dry products will be poor [21]. It is worth mentioning here, that the drying process consumes a major chunk of energy in the food processing industry [22]. According to Simal et al. [23], vegetables and fruits are more perishable in comparison with grains, since they comprise a considerable amount of moisture. The drying process helps to avoid undesirable changes, such as microbial spoilage and enzymatic reactions, to remove the moisture. Goyal et al. [24], reported that drying leads to a reduction of weight and volume of food products. It also helps in the reduction of the costs associated with packaging, storing, and transportation [25].

Heat and mass transport processes in a system are respectively controlled by the Fourier law and Fick's second law of diffusion. Hussain et al. [26], reported that these laws offer simplified mathematical representations of the physical reality of heat and mass transfer. It is very helpful in the field of food engineering to use mathematical modeling as a tool for the design and control of drying operations [27]. The distribution of moisture content and temperature may cause several unwelcome changes to take place in meals either during the drying process or after the drying process has been completed [28]. For this reason, it is possible that modeling and forecasting the temperature and moisture concentrations in meals as an indicator of drying time might assist us in avoiding unwelcome changes in foods while they are being dried or preserved [29].

Bananas that are still fresh are renowned for having a high moisture content. Drying the fruit after harvesting eliminates moisture from the fruit, decreases water activity, limits the formation of microbes, and minimizes physicochemical deterioration [30]. Drying bananas is one strategy that may be used to reduce the amount of microbial activity and degradation that occurs while also extending the amount of time that bananas can be stored [31]. To prevent the development of microorganisms, render enzymes inactive, and reduce unwelcome alterations in the color, taste, and texture of dried objects, the drying time and temperature are important factors to consider [32]. Furthermore, the overall quality of dried banana slices is affected by the pretreatment processes that are employed [33]. The traditional numerical techniques used to estimate banana slice drying have various limitations that restrict their usefulness in tackling the complex difficulties presented by climate change and the requirement for sustainable food preservation measures [34,35]. For starters, classic numerical modeling techniques often depend on simplified mathematical illustrations of heat and mass transport processes, such as Fourier's equation and Fick's second rule of diffusion. While these equations give important information, they may oversimplify the complex dynamics of heat and moisture transport throughout the food matrix, resulting in incorrect predictions of drying kinetics and the quality of the product [36].

The traditional numerical techniques often fail to account for the wide range of elements that influence drying processes, such as differences in ambient conditions, food qualities, and equipment design [37]. As a consequence, these models may fail to represent all of the interactions and events that occur throughout banana slice drying, restricting their prediction power and practical value [38]. Furthermore, the growing focus on ecological and environmental stewardship in food processing mandates the development of creative techniques to reduce energy consumption and waste creation while preserving product quality [39]. Conventional drying procedures often depend on high temperatures [40], which may result in unwanted changes in taste, appearance, and nutritional value, as well as the creation of potentially toxic chemicals [41]. These shortcomings highlight the need for alternative drying technologies that provide more sustainable and ecologically friendly alternatives. Machine learning offers a viable solution for overcoming the constraints of traditional numerical approaches in simulating banana slice drying. In the literature, machine learning based-prediction and optimization models are being implemented successfully in several domains of engineering applications like energy (such as modelling thermal performance [42], harnessing renewable energy [43], optimizing energy system [44]), fuels (optimizing parameters of dual-fuel engine using biogas [45], evaluating characteristics of engine fueled with biodiesel [46], optimizing biofuel production [47]), environment (classification of waste [48] and sorption capacity prediction of hazardous elements [49]), industrial applications (physically-based cloth simulation [50], optimizing the manufacturing products [51,52]), nano-based technologies (predicting the thermophysical properties of nanofluid [53] and optimizing the heat transfer of nanofluid [54]), society and health (prediction of in-hospital risk [55], improving learning environment [56], managing health systems [57]), agriculture (managing land-use change [58], evaluating the rice seeds quality [59]), and transport (event-triggered adaptive control for ships [60], managing maritime and logistics [61], modelling ship main engine [62]). Machine learning systems, which use advanced algorithms and computational methods, may analyze large amounts of information, find underlying patterns, and generate accurate predictions in a variety of situations [63,64]. Furthermore, machine learning models may adapt and improve over time via continual learning and optimization [65], resulting in more accurate and trustworthy forecasts of drying kinetics and product quality [66]. Machine learning algorithms may combine several sources of data, such as sensor readings, process parameters, and ambient variables, to capture the

complex character of banana slice drying and improve predicted accuracy [67]. By bringing machine learning into the modeling process, researchers may create more complete and sophisticated models that better represent the intricate interaction of variables that influence drying [68,69]. A summary of recent studies using machine learning for solar-based drying of food items is listed in Table 1.

The research introduces a novel method for predicting the drying characteristics of banana slices using state-of-the-art machine learning algorithms integrated into an indirect solar dryer system. The work creates a prognostic model that incorporates advanced machine learning methods for ML-based model prediction of the drying process. A more flexible and resilient modeling strategy, able to handle the intricacies of drying processes, is guaranteed by this departure from traditional numerical approaches. This makes the article's emphasis on efficient and environmentally friendly drying methods all the more relevant.

## 2. Materials and methods

### 2.1. Test setup and methodology

In the experimental part, the test samples of fresh bananas were purchased from a banana plantation in Thoothukudi, Tamil Nadu, India. The test samples were prepared in two different sizes employing an automated cutting machine. To guarantee that the product volume remained constant, each experiment included cutting the banana into two geometries. The oven technique was used to calculate the initial moisture content of banana slices at 105 °C for 24 h [79,80]. An indirect solar dryer featuring a control device to control the flow of water and temperature was used for the research. Measurements were taken using portable instruments like a weight scale, and an infrared temperature gun. Using a fresh food product as a pre-treatment was not necessary. Different drying cabin temperatures, flow of water rates, and banana slice sizes were tested. To ensure a consistent product volume, the bananas were divided into two shapes for every experiment: square and circular. Periodically, the digital scale was utilized to evaluate the amount of weight lost. Uneven dehydration persisted until the product achieved the desired equilibrium moisture content (EMC). To calculate EMC, the Guggenheim-Anderson-de Boer model was used. At dehydration temperatures between 72 and 90 °C, its dry basis (db) value was approximately 0.01 gm water/gram dry matter. A series of experiments were carried out using different product masses and water flow rates. The length of drying is calculated employing to calculate the amount of energy used to dry banana slices in a drier at different input temperatures and velocity levels as Eq. (1) [81].

**Table 1**
A summary of recent works in the domain of solar-based food items drying.

| Food article and Type of solar drying | ML used | Main outcome | References |
|---|---|---|---|
| Sweet potatoes (Convective type tray dryer) | Neuro-fuzzy, Random Forest (RF), and Support Vector Machine (SVM) | **1.** Drying time was the most influential factor.<br>**2.** Neuro-fuzzy was the best ML technique in this case. | [70] |
| Potato slice (Hot air dryer) | Artificial Neural Network (ANN), Response Surface Methodology (RSM), and Adaptive Neuro-Fuzzy Inference System (ANFIS) | **1.** Temperature and drying time were the most significant factors.<br>**2.** ANFIS was best best-performing ML technique | [71] |
| Peanut Novel drying chamber) | SVM, Decision Tree (DT), Random Tree (RT) and Multiple Linear Regression (MLR) | **1.** The novel design was more effective in drying.<br>**2.** RF-based model provided the best predictions. | [72] |
| Watermelon rind (Convective dryer) | ANN optimized with genetic algorithm (GA) | **1.** The peak temperature achieved in the dryer was 45 °C.<br>**2.** ANN results were superior to the fitted empirical approach. | [73] |
| Mushroom slice (Hot air impingement drying) | Extreme Learning Machine Integrated Bayesian Methods (ELMIBM), Back Propagation Neural Network (BPNN), and Basic Extreme Learning Machine | **1.** ELMIBM was superior in this case.<br>**2.** Relative prediction errors were less than 8.5 %. | [74] |
| Mint (Greenhouse solar dryer) | Gaussian process regression (GPR), Multilayer Perceptron Neural Network, Radial Basis Function (RBF) | **1.** Models could be developed for mass and temperature.<br>**2.** The RDF model offered higher accuracy compared to other ML techniques in this case. | [75] |
| Tomato (Solar dryer with controllable temperature system) | Feed Forward Neural Network (FFNN) optimized with GA | **1.** A drying model was developed.<br>**2.** A correlation coefficient higher than 0.99 was achieved. | [76] |
| Pomelo fruit (Microwave drying, Freeze drying, and Forced convection drying) | ANN | **1.** Mass and drying time prediction models with ANN were developed.<br>**2.** R-value above 0.993 could be achieved for all cases. | [77] |
| Apple cubes (Convective dryer) | ANN, GA | **1.** The optimized setting was 1 m/s air velocity and 65 °C drying temperature.<br>**2.** The prediction was less than 3.24 %. | [78] |

$$q = \rho . A . V . C_p . \Delta T . t \tag{1}$$

The variables q denotes energy consumption (kJ), A represents the inlet area of the dryer ($m^2$), V denotes the velocity of air (m/s), $\rho$ represents the density of air ($kg/m^3$), t denotes time consumed in each drying cycle (s), $\Delta T$ difference of temperature between banana slice and air (°C), and $C_p$ represents specific heat of air (kJ/kg C).

The oven method evaluates the sample moisture content. The samples were dehydrated in the oven until the moisture content remained constant. The masses of the slices were calculated before and after surgery using electronic weighing equipment. The sample's moisture content was determined using the expression in Eq. (2) [82]:

$$Z_t = \left(M_t - M_d\right) / \left(Md\right) \tag{2}$$

Herein, Zt represents the moisture content in the test sample at a specified time 't' of the process. $M_t$ represents the weight of the test sample at time 't' and $M_d$ represents the hydrated weight of the banana slice. A total of 30 experiments were conducted. The schematics of the test setup are depicted in Fig. 1a.

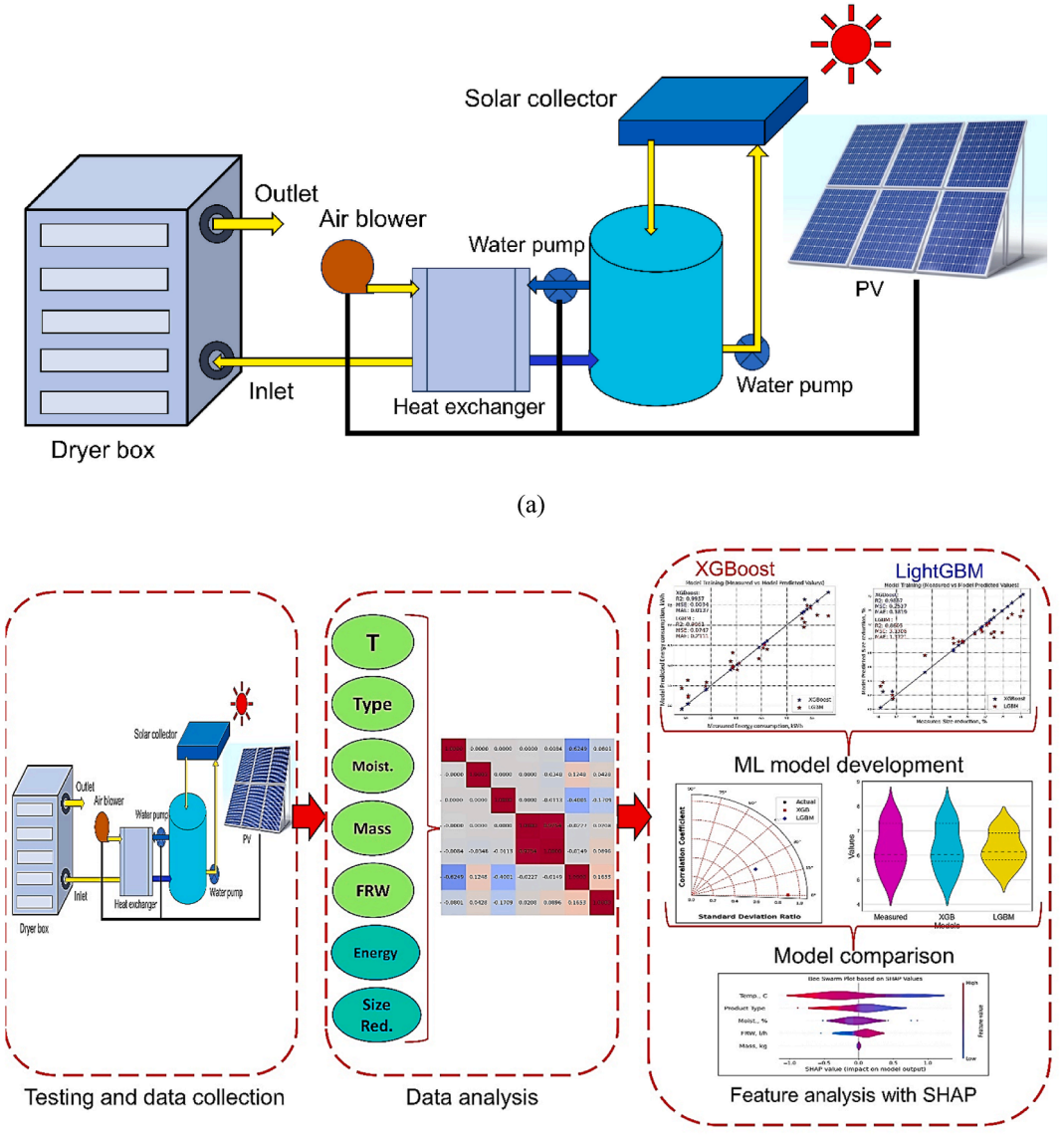



**Fig. 1.** (a) Schematics of test setup; (b) Testing and ML framework used in this study.

### 2.2. Machine learning

The process of modeling, simulating, and making predictions regarding the complicated and non-linear behavior of solar-based drying of perishable food products is a work that is both difficult and time consuming, as was mentioned before. Modern machine learning approaches, on the other hand, present a one-of-a-kind potential to provide the resolution in this respect. As a result, the current investigation makes use of two contemporary machine learning-based algorithms for prognostic purposes: extreme gradient boosting and lightGBM. The data analysis process was started by importing Python libraries for various projects. Pandas were introduced as pd to handle data manipulation and organization, numpy as np for numerical calculations, and matplotlib.pyplot as plt to visualize data graphically. In addition, train_test_split was introduced from sklearn.model_selection to split the data set into training test subsets, an important step in model development to use this library for gradient boosting algorithms. The Lightgbm process was implemented by importing lgb library while XGBRegressor was used for XGBoost implementation. In the Jupyter Notebook environment, using default hyperparameters, regression-based model estimation was performed. This advanced approach by seamlessly integrating these libraries simplified intensive regression analysis within the Python ecosystem, making the search effort more efficient and accurate. The ML process adopted in this study is depicted in Fig. 1b.

#### 2.2.1. XGBoost

Extreme Gradient Boosting (XGBoost) belongs to the ensemble learning technique family and works by iteratively creating an ensemble of weak learners, most notably decision trees, to generate a strong prediction model [83]. XGBoost is effective in both classification and regression because it employs regularization to decrease overfitting and minimizes a specified loss function [84]. Two of its main features are gradient boosting, which successively corrects prediction errors created by earlier models, and approximation tree learning, which accelerates training by creating trees level-wise, these are two of its most important aspects. XGBoost's success may be attributed to its ability to provide extremely accurate predictions as well as its adaptability to a wide range of applications [85,86]. The method went as follows:

Suppose the test data set employed in the problem is specified as [87,88]:

$$D = \left\{ \left( x_i, y_i \right) \right\}_{i=1}^{n}; \tag{3}$$

In this case, the vector feature is denoted as $x_i$ and the label for $i_{th}$ sample is represented by $y_i$. XGBoost ML is designed to train an additive ensemble of weak learners:

$F\ (x) = \sum_{k=1}^{K} f_k\,(x)$ ; herein, $f_k(x)$ denotes regression tree.

The following expression can be used as the objective function in this case [87,88]:

$$Obj = \sum_{i=1}^{n} L\ \left( y_i, \hat{y}_i \right) + \sum_{k=1}^{K} \Omega\left( f_k \right) \tag{4}$$

Herein, $L\,(y_i, \widehat{y_i})$ is used for loss function $y_i$ represents the estimated value and $\widehat{y_i}$ is used for model predicted value.

The model is trained iteratively by introducing weak learners within the ensemble. In every cycle, a new regression tree undergoes training to minimize the ensuing objective function [87,88]:

$$Obj_k = \sum_{i=1}^{n} L\ \left( y_i, \hat{y}_i^{\,t} \right) + \sum_{i=1}^{t} \Omega\left( f_i \right) + \gamma . T + \frac{1}{2}\ \lambda\ \sum_{j=1}^{T} \omega_j^2 \tag{5}$$

The complexity of the tree is defined by the regularization factor λ, and the total number of leaves γ is represented by T in Eq. (5). When all of the regression trees in an ensemble have finished training, their predictions are combined to get the most accurate prediction for a fresh sample x.

$F\ (x) = \sum_{k=1}^{K} f_k\,(x)$; herein, the $f_k(x)$ is the prediction of $k_{th}$ tree.

#### 2.2.2. LightGBM

Light Gradient Boosting Machine (LightGBM) is an ML framework that is renowned for its scalability and efficiency when it comes to the management of big datasets. Efficiency and efficacy are two aspects of gradient boosting that stand out when compared to other implementations of the technique [89]. One of the unique properties of LightGBM is that it natively supports categorical characteristics, and it does so without the need for one-hot information encoding. The training pace is increased, but the consumption of memory for training is minimized [90,91].

A gradient-based tree-boosting approach is utilized by LightGBM. This algorithm was developed with the particular purpose of optimizing both efficiency and scalability. First, let's have a look at the mathematical foundations that support LightGBM. Typically, an objective function is specified as the sum of a loss function $(y_i, \widehat{y_i})$, and LightGBM is able to minimize this function. It uses a regularization term $\Omega(f)$. The objective function in the case of LightGBM can be defined as [92,93]:

$$Obj = \sum_{i=1}^{n} L\ \left( y_i, \hat{y}_i \right) + \sum_{k=1}^{K} \Omega\left( f_k \right) \tag{6}$$

In contrast to depth-wise tree development, which involves growing the tree level by level, LightGBM uses a tree growth approach that develops the tree leaf by leaf. This results in fewer levels, which in turn helps to reduce the amount of memory that is used and speeds up the training process. For the sake of this discussion, let T(x) stand for the decision tree model, wherein x is the input feature. To minimize the objective function, LightGBM constructs a succession of decision trees iteratively. In the case of each iteration m, it

fits a new tree to the negative gradient of the loss function, represented as $g_m\ (x)$. The forecast $\hat{y}_i^m$ for a given case $x_i$, at m iteration is updated as [92,94]:

$$\hat{y}_i^{(m)} = \hat{y}_i^{(m-1)} + learning\ rate \times T\left(x_i\right) \tag{7}$$

Herein, $\hat{y}_i^{(m-1)}$ is a prediction from the preceding iteration.

### 2.3. Statistical evaluation of models

Statistical methodologies like coefficient of determination ($R^2$) were used for model evaluation and comparison. $R^2$ quantifies the variation in the dependent variable (the variable being predicted) and is explained by variations between the independent variables (predictors) in the regression model. It ranges from 0 to 1, with '1′ properly explaining the model and '0′ not explaining anything. The optimal number should fall anywhere between 0 and 1, preferably close to 1 [95,96]. The expression of $R^2$ is given as:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{8}$$

Herein, $SS_{tot}$ represents the total sum of squares while $SS_{res}$ denotes residual's sum of squares.

The average absolute difference between the data that was seen and the data that was projected is the mean absolute error (MAE). MAE offers a measurement of the mean quantity of errors, but it penalizes huge errors with less severity. The MAE is determined by the following formula [97]:

$$MAE = \frac{1}{n}\sum_{i-1}^{n}\left|y_i - \hat{y}_i\right| \tag{9}$$

Mean Squared Error (MSE) estimates the mean of squared difference between predicted and measured data values. It estimates the mean size of the model's errors, with greater MSE representing bigger errors. The following expression was employed using the following expression [98]:

$$MSE = \frac{1}{n}\sum_{i-1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{10}$$

In this case, n is the total number of observations, $y_i$ represents the measured value of $i_{th}$ observation, $\hat{y}_i$ denotes the forecasted value of $i_{th}$ observation.

## 3. Results and discussion

### 3.1. Data analysis and preprocessing

Preprocessing and analyzing the data is known as an essential step in enhancing the performance of ML-based prediction models [99]. In this process, the correlation matrix offered depicts the correlations between different variables in a system. Each column in the matrix represents the correlation coefficient between two variables, which can vary from −1 to 1. A score of 1 implies perfect positive correlation, −1 shows perfect negative correlation, and 0 indicates no connection. As shown in Fig. 2, the correlation coefficient

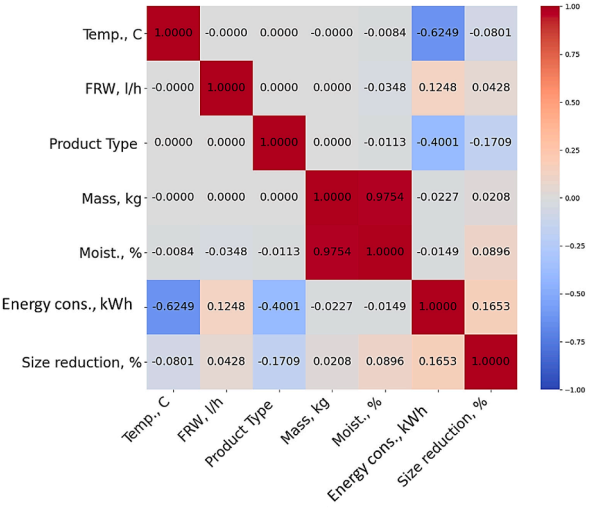



**Fig. 2.** Correlation heatmap.

(R) value between temperature and flow rate of water is negligible, indicating that these two variables have no linear connection. Similarly, the relationship between temperature and product type, as well as temperature and mass, is minimal.

The value of R is negative (−1.11E-16) between flow and temperature, close to zero between flow and productivity, indicating no significant relationship between these variables. The correlation between volume and moisture is high (0.975), and it shows a positive relationship. This indicates an increase in moisture content with concentration. Similarly, the correlation between mass and energy consumption is slightly negative (−0.0227), indicating that more mass leads to less energy consumption. The correlation between water and energy consumption is −0.015, indicating a slightly negative relationship between these two variables. Finally, the correlation between size reduction and energy consumption is positive (0.165), indicating that as the percentage of significant reduction increases, energy consumption also increases there was a slightly positive correlation (0.09) between size decrease and moisture content. This indicates that a larger size decrease corresponds to an increase in moisture content.

Descriptive statistics provide information on the central tendency, dispersion, and shape of the distribution for each variable in the dataset. As depicted in Table 2, beginning with temperature (Temp., C), the average temperature is 80 °C with a standard variation of 7.88 °C. The temperatures vary from 70 °C to 90 °C, with 80 °C being the median (50th percentile). The distribution looks to be symmetrical, with a minor negative skewness (−1.35) and zero kurtosis, indicating a generally normal distribution. The average flow rate of water (FRW, l/h) is 20 l/h, with the same standard variation as temperature (7.88 l/h). The readings vary from 10 l/h to 30 l/h, with a median of 20 l/h. Similar to temperature, the distribution of flow rates appears to be reasonably symmetrical, with the same skewness and kurtosis. The product type variable has a mean of 0, suggesting a balanced distribution centered on zero. The standard deviation is 0.788, indicating moderate variability in the data. The numbers range from −1 to 1 and represent the various categories of goods. The distribution appears to be generally symmetrical, with kurtosis and skewness that match those of temperature and flow rate. Moving on to mass (kg), the average mass is 0.62 kg, with a standard deviation of 0.299 kg. The mass values vary from 0.24 kg to 1 kg, and the distribution appears to be generally symmetrical based on skewness and kurtosis.

The moisture content variable has an average of 29.98 % and a standard deviation of 21.38 %. The moisture content varies from 5.52 % to 61.65 %, with a median of 27.05 %. The distribution looks to be strongly skewed (skewness = 0.36), with a little larger kurtosis (−1.37), indicating a distribution with heavier tails and a more prominent peak than a normal distribution. The average energy usage (kWh) is 6.26 kWh, with a standard deviation of 0.849. The values vary between 4.92 kWh and 7.82 kWh, with a median of 6.035 kWh. The skewness and kurtosis readings suggest that the distribution is substantially symmetrical. Finally, the average size reduction percentage is 69.5 %, with a standard deviation of 4.887 %. The percentages vary from 60.21 % to 76.24 %, with a median of 71.23 %. The distribution looks to be negatively skewed (−0.79) and has considerable kurtosis (−0.56), indicating a distribution with lighter tails and a flatter peak than a normal distribution.

### 3.2. Energy consumption model

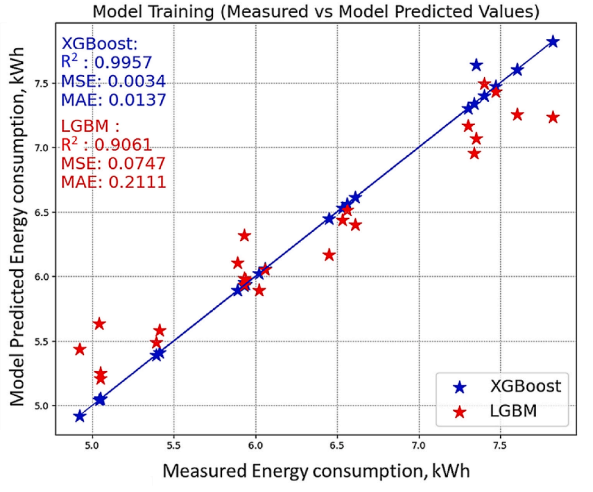
Post data analysis and correlation evaluation, the prediction models for energy consumption during the drying process were developed using two modern ML techniques namely XGBoost and LightGBM. The data was divided into two parts, training as well as testing in a 70:30 ratio. The open libraries in Python were used.

Once the models were ready these were employed for prediction of data both during the training and testing process. Fig. 3 depicts the model performance for XGBoost and LightGBM-based models during model training as well as testing, respectively. It can be observed that both during training and testing, the XGBoost-based model performs superior to the LightGBM-based model since most of the results in the form of data points are close to the best-fit line. To make things more clear, the models were compared based on statistical metrics as marked on the plots.

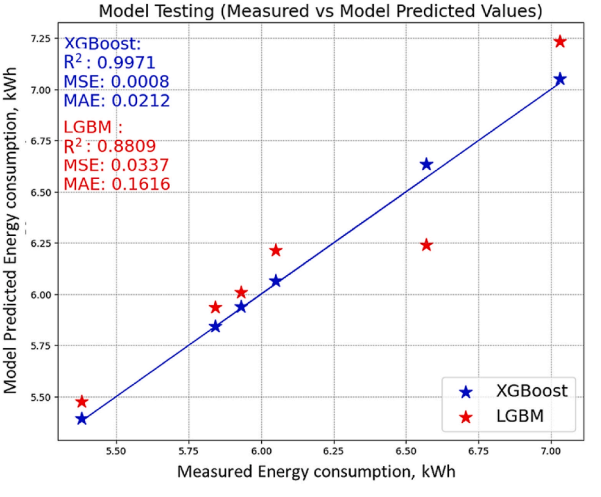
These indicators are critical for assessing the efficacy and generalizability of the models. Starting with XGBoost, in the training phase, it gets a very high coefficient of determination ($R^2$) of 0.9957, meaning that the model can explain 99.57 % of the data variance. The mean squared error (MSE) is astonishingly low at 0.0034, suggesting that the actual and projected values are quite similar. Furthermore, the mean absolute error (MAE) is minimal, at 0.0137, indicating high prediction accuracy. Moving on to the testing step for XGBoost, the model retains its impressive performance. The $R^2$ score rises slightly to 0.9971, indicating high generalization ability. The MSE falls further to 0.0008, showing that the model continues to perform well on unknown data. However, MAE has increased slightly to 0.0212, indicating a somewhat larger average error than in the training phase, although it is still within acceptable bounds. Moving on to LightGBM, in the training phase, it obtains an $R^2$ value of 0.9061, which is lower than XGBoost, suggesting that it explains 90.61 % of the variance in the data. The MSE is greater (0.0747), indicating a bigger mistake than XGBoost. Similarly, the MAE is greater, at 0.2111, suggesting lower prediction accuracy than XGBoost. LightGBM's performance metrics fall somewhat during the testing phase as compared to the training phase. The $R^2$ value falls to 0.8809, suggesting a decline in explanatory power for un-

**Table 2**
Descriptive statistics.

| | Mean | Std | Min | 25 % | 50 % | 75 % | Max | Kurtosis | Skewness |
|---|---|---|---|---|---|---|---|---|---|
| Temp., C | 80 | 7.88 | 70 | 70 | 80 | 90 | 90 | −1.35 | 0 |
| FRW, l/h | 20 | 7.88 | 10 | 10 | 20 | 30 | 30 | −1.35 | 0 |
| Product Type | 0 | 0.788 | −1 | −1 | 0 | 1 | 1 | −1.35 | 0 |
| Mass, kg | 0.62 | 0.299 | 0.24 | 0.24 | 0.62 | 1 | 1 | −1.35 | 0 |
| Moist., % | 29.98 | 21.38 | 5.52 | 6.77475 | 27.05 | 56.57 | 61.65 | −1.37 | 0.36 |
| Energy cons., kWh | 6.26 | 0.849 | 4.92 | 5.8525 | 6.035 | 6.93 | 7.82 | −0.98 | 0.22 |
| Size reduction, % | 69.5 | 4.887 | 60.21 | 68.285 | 71.23 | 72.67 | 76.24 | −0.56 | −0.79 |




(a)




(b)

**Fig. 3.** Energy consumption model measured vs predicted values during; (a) Training, (b) Testing.

known data. The MSE drops to 0.0337, an improvement over training but still higher than XGBoost. The MAE also drops to 0.1616, suggesting that the average error has decreased relative to the training phase but remains greater than XGBoost. In summary, XGBoost outperforms LightGBM in both the training and testing phases, with higher $R^2$ values and lower error metrics, indicating improved prediction accuracy and generalization ability, these results are agreed by other researchers in the literature [100].

### *3.3. Size reduction (shrinkage) model*

Following data collection and correlation evaluation, prediction models for shrinkage of chip size during the drying process were constructed utilizing two recent machine learning techniques: XGBoost and LightGBM. The data was separated into two parts: training and testing in a 70:30 ratio. Open libraries in Python were utilized. Once the models were completed, they were used to predict data during both the training and testing phases. Fig. 4 shows the model performance of XGBoost and LightGBM-based models during training and testing, respectively. Both during training and testing, the XGBoost-based model outperforms the LightGBM-based model because the majority of the data points are near the best-fit line. To make things clearer, the models were compared using statistical measures as shown in the graphs.

These measures are often used to assess the performance of regression models. Starting with XGBoost, during the training phase, the model acquired a high $R^2$ value of 0.9887, suggesting that it explains about 98.87 % of the variation in the target variable. The MSE for the training data is 0.2527, indicating that the model's predictions are generally accurate. Furthermore, the MAE for the training data is 0.1819, which is the average absolute difference between the predicted and observed values. Moving on to the testing step for XGBoost, the $R^2$ score stays high at 0.9775, showing that the model generalizes well to new data. However, both MSE and

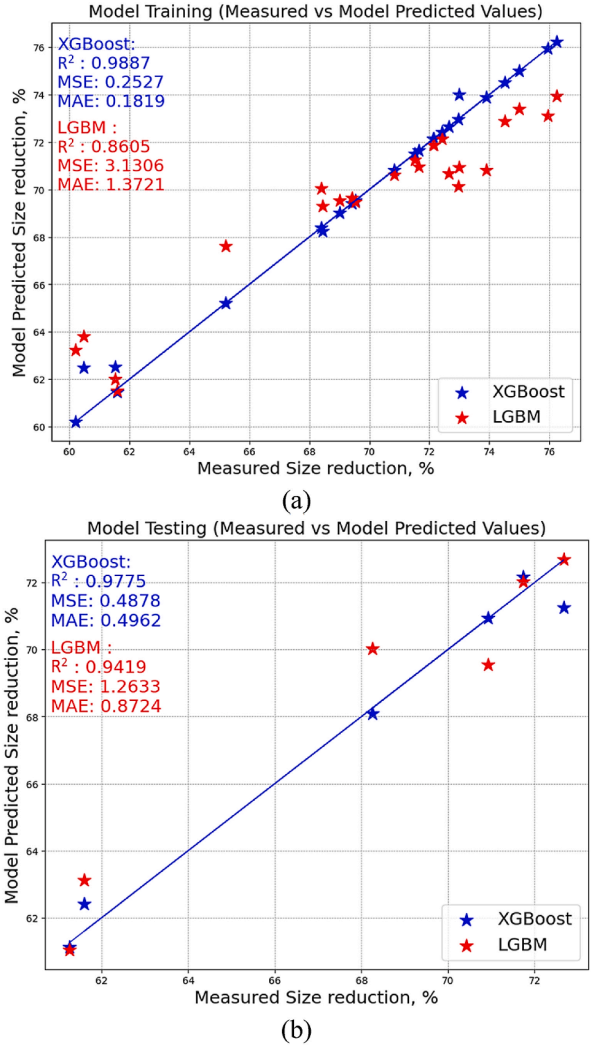



**Fig. 4.** Size reduction (Shrinkage) model measured vs predicted values during: (a) - Training, (b) - Testing.

MAE have somewhat increased from the training phase, with values of 0.4875 and 0.4962, respectively. This rise indicates that the model's performance on unknown data is somewhat poorer than on training data, but it still has a high predictive potential.

In the case of LightGBM in the training phase, the $R^2$ value is lower than XGBoost, at 0.8605, suggesting that the model explains roughly 86.05 % of the variation in the target variable. The MSE for training data is greater, at 3.1306, indicating a bigger average squared difference between predicted and actual values than XGBoost. Similarly, the MAE for training data is 1.3721, showing a larger average absolute difference than XGBoost. However, throughout the testing phase, LightGBM shows a significant improvement in performance. The $R^2$ score rises significantly to 0.9419, showing that the model generalizes better to unknown data than it did in the training phase. Furthermore, both MSE and MAE fall dramatically to 1.2633 and 0.8724, respectively, showing that the model's predictions on unexplored data are more accurate and closer to the true values than its performance on training data. Overall, although both XGBoost and LightGBM perform well in predicting the target variable, LightGBM outperforms in the testing phase, with higher generalization and smaller prediction errors on unknown data [101].

### *3.4. Model comparison using Taylor's and violin diagrams*

Taylor diagrams and violin plots are two examples of approaches that may be utilized to give valuable information on the performance of various machine learning models when comparing between them, this is also agreed by other works [102,103]. It is possible to simultaneously evaluate models based on a number of metrics, including correlation, root-mean-square error (RMSE), and standard deviation, with the help of Taylor diagrams, which are a complete visualization tool [104]. On the diagram, each model is represented as a point, and its distance from a reference point is used to show correlation, while its placement along the x-axis is used to communicate standard deviation. The degree of correlation that exists between the model and the reference may be determined by the angle that is formed between the model point and the reference point. Comparisons and model selection are both made possible through the use of Taylor diagrams, which offer a concise method for evaluating a large number of models using a variety of metrics.

Taylor's diagrams in the case of the energy consumption model during training and testing are depicted in Fig. 5a and b, respectively. Taylor's diagrams in the case of size reduction during training and testing are depicted in Fig. 6a and b, respectively. These plots also reveal the XGBoost-based model's superiority in predictions.

Comparing models may also be done using violin plots, which are an additional method in addition to Taylor diagrams. The distribution of prediction errors and other critical metrics among models may be efficiently displayed using violin charts, which are highly successful in this regard, this result is also similar to previous work [105]. Each model is shown in the form of a "violin," which illustrates the distribution of the metric, and which typically includes a kernel density plot superimposed on top of it. The spread and shape of distributions among models may be easily compared with the help of these graphs, which also provide insights into the consistency of model predictions and demonstrate differences in performance variability.

Before attempting to make an accurate comparison between several machine learning models employing Taylor diagrams and violin plots, it is necessary to first analyze the performance of each model by employing pertinent metrics. The computation of correlation coefficients, root mean square error (RMSE), standard deviation, and prediction errors might fall under this category. Following the acquisition of these measurements, Taylor diagrams may be utilized to illustrate the overall performance of each model in contrast to a reference model or observations. During this time, it is possible to generate violin charts in order to examine the distribution of prediction errors across the various models. Fig. 7 exhibits the violin plots for the sizer reduction model throughout training and testing. Fig. 8 shows violin plots for size decrease during training and assessment. These charts further demonstrate the XGBoost-

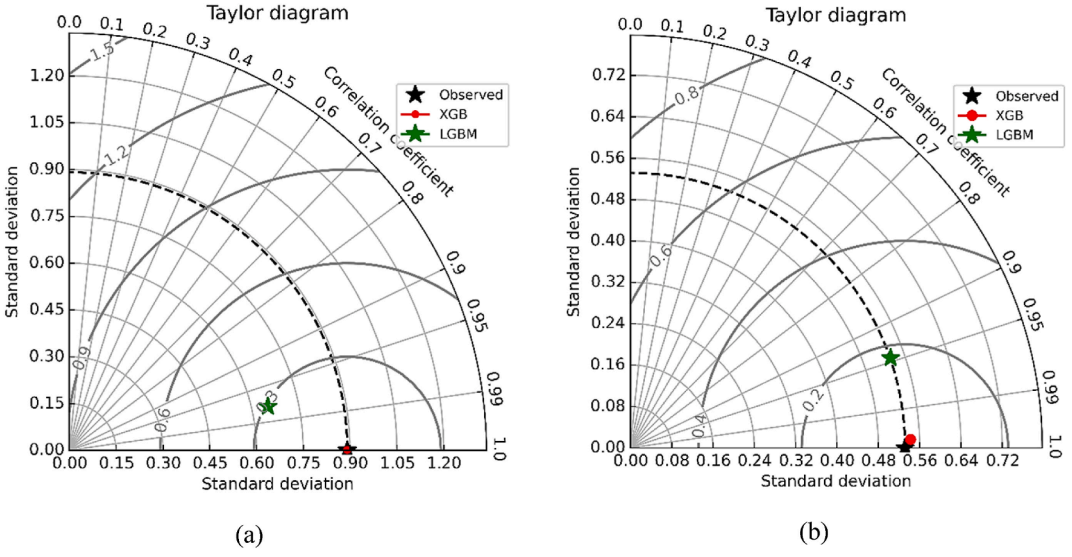



**Fig. 5.** Taylor's diagram for energy consumption model during: (a) Training, (b) Testing.

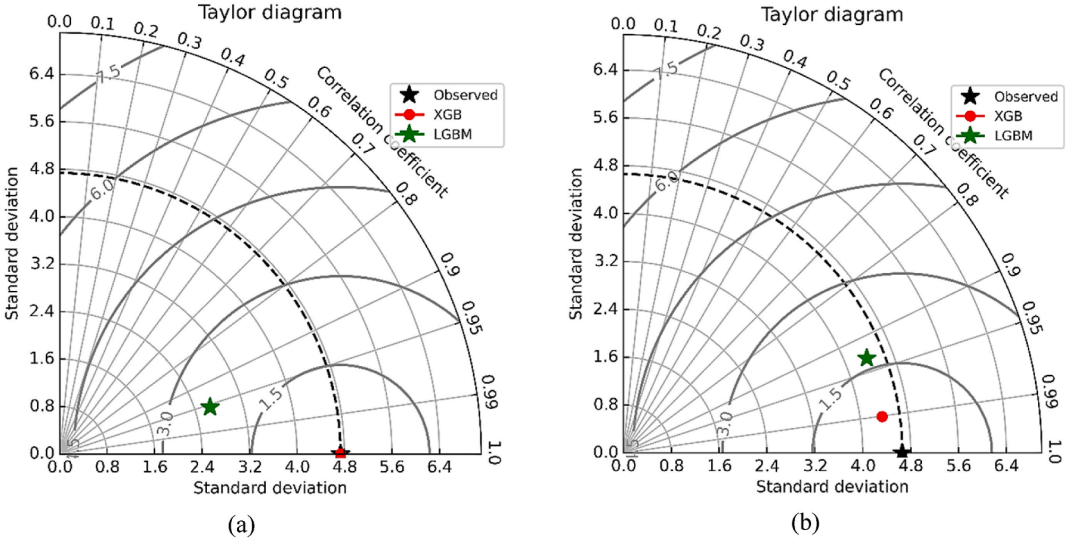
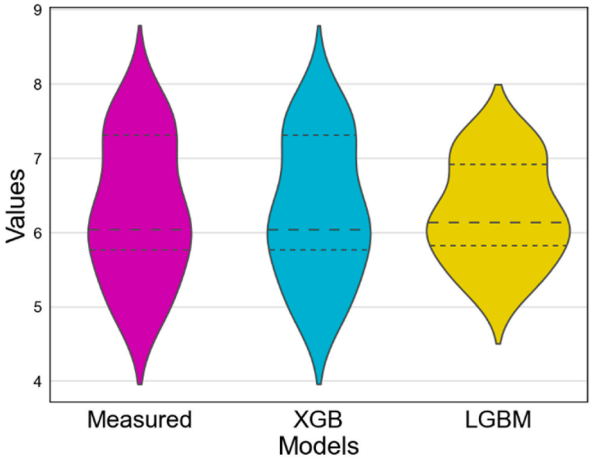
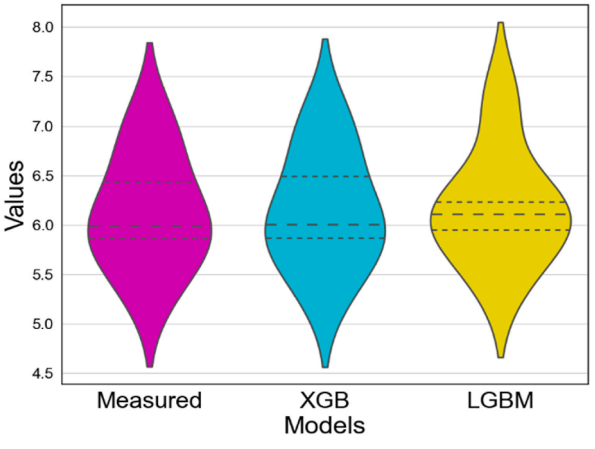



**Fig. 6.** Taylor's diagram for size reduction model during: (a) Training, (b) Testing.




(a)




(b)

**Fig. 7.** Violin plots for energy consumption model during: (a) Training, (b) Testing.

based model's advantage in prediction. It can be observed that violin plots of XGBoost-based models almost resemble the measured data.

By analyzing the Taylor diagrams, one may locate models that have a high correlation and a low root mean square error (RMSE) in comparison to the reference. On the other hand, by analyzing the violin plots, one can evaluate the variety and shape of prediction error distributions among the models. Through the integration of data from both representations, stakeholders have the potential to make decisions that are more well-informed on the selection of models, additional optimization, and areas for development. All things considered, the combination of Taylor diagrams and violin plots offers a comprehensive approach to the evaluation of machine learning models. This approach enables a more comprehensive comprehension of the performance characteristics of these models over a wide range of metrics and datasets.

### *3.5. eXplainable ML methods*

In the present study, two modern and highly precise ML techniques XGBoost and LightGBM were employed. However, both ML techniques are regarded as black-box methods. So, to improve on this aspect, the employment of eXplainable methods was considered, in which explainable artificial intelligence was an attractive option to address the issue of poor interoperability in black-box methods [106,107]. Using eXplainable methods could improve the interpretability, trustworthiness, and comprehension of black-box models [108]. Since XGBoost was superior in providing predictions in this case [109], it was selected for this purpose, and two different methods of eXplainable ML were selected. The use of Shapley plots, in conjunction with XGBoost predictions for energy consumption and size reduction, offers valuable insights and validation in the area of banana chip drying optimization. A better understanding of the relative significance of input features may be achieved via the use of the Shapley plot. This plot reveals which components have a significant influence on the amount of energy consumed and the decrease in size throughout the drying process [110]. When researchers look at the values of Shapley, they have the opportunity to identify important factors and focus their efforts to increase both

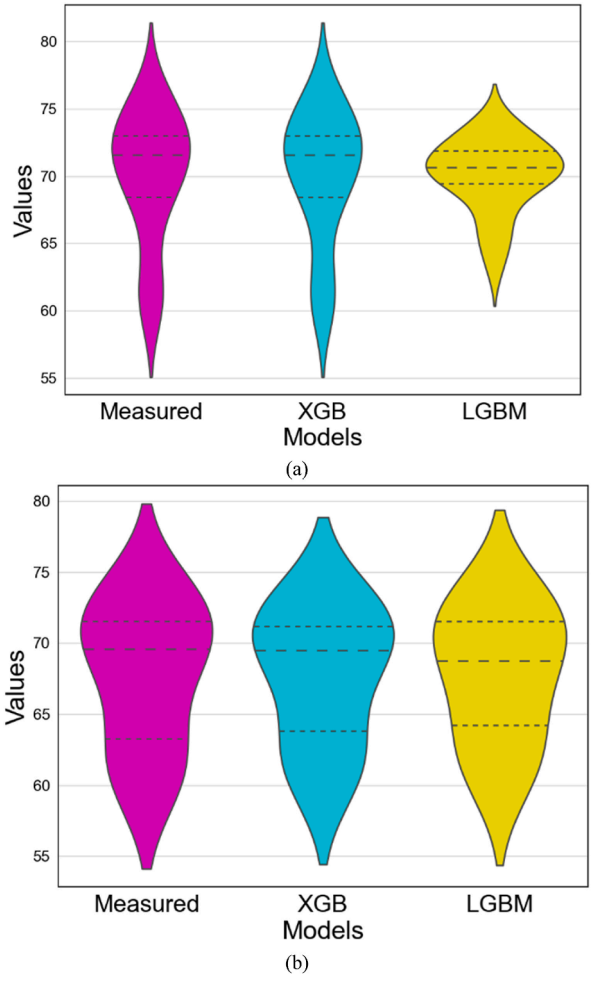



**Fig. 8.** Violin plots for size reduction model during: (a) Training, (b) Testing.

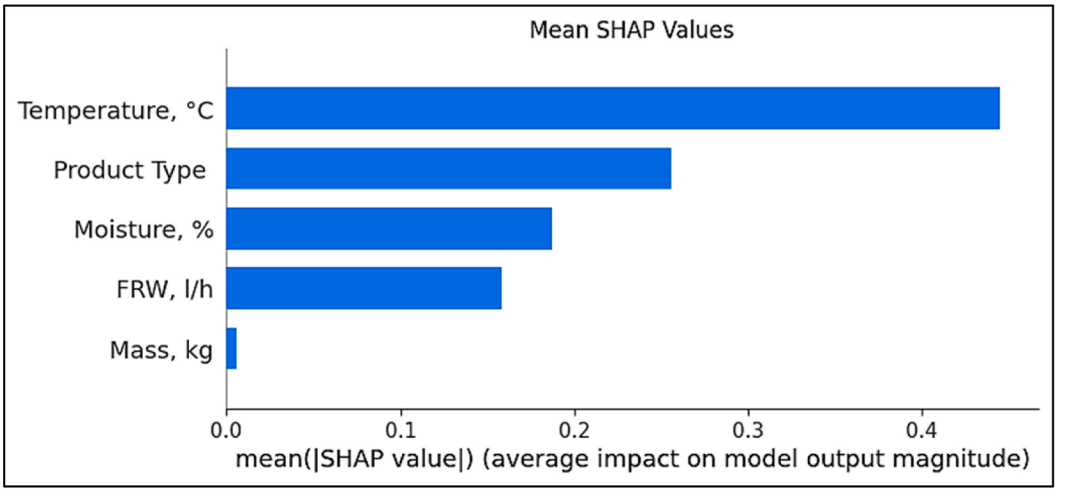
efficiency and product quality [111]. Through the process of comparing the projected value distribution to the observed values, researchers have the opportunity to evaluate the performance of the model and identify any potential disparities or biases that may exist. This validation step ensures that the XGBoost model generates accurate predictions, hence generating trust in the model's utility as a tool for making decisions in the context of banana chip drying operations. In a nutshell, the combination of Shapley plots with XGBoost predictions enables researchers to not only comprehend the fundamental dynamics of the drying process but also to assess and update predictive models [112], which ultimately leads to optimal drying conditions and improved product quality [113].

### *3.5.1. Shapley plots for energy consumption model*

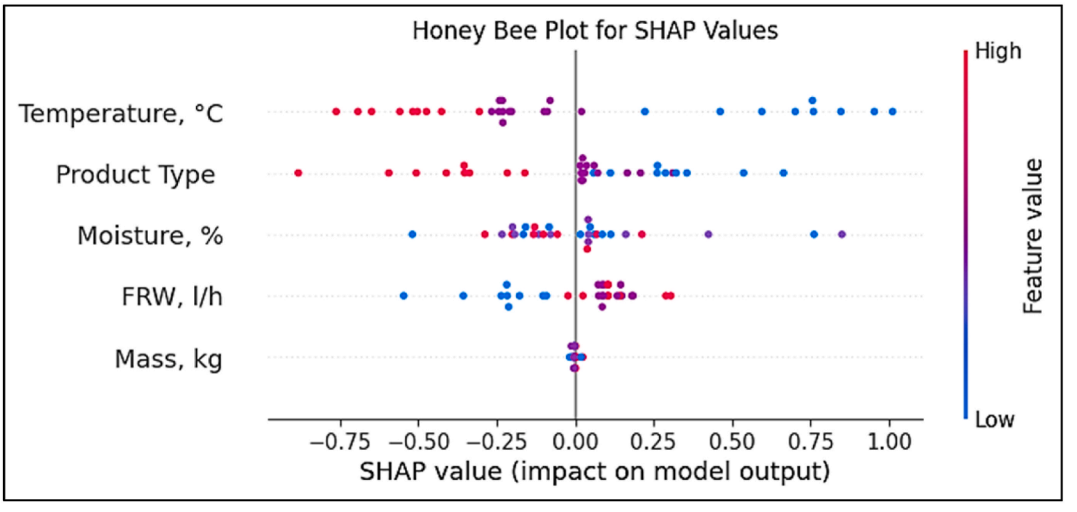
The mean Shapley demonstrates the effect on the prediction output of the model is depicted in Fig. 9a and the Bee swarm cum violin plot is depicted in Fig. 9b. SHapley Additive exPlanations (SHAP) values compute the average influence of each feature on the model's output (in this case, estimated energy consumption) [114,115]. It can be observed that temperature is the most influential feature that affects the model of energy consumption followed by the product type (its shape and size). The moisture content too has a substantial effect on the model but not as large as the previous two. The length of the bars and the size of the data point for each SHAP value indicate the relative relevance of each attribute. Longer bars as in the case of temperature and product type indicate a greater influence on energy use. The bee swarm cum violin plot showing kernel density of the plot helps in the identification of the nature of effects on the prediction. It can be observed from Fig. 9b that temperature and moisture have more negative effects, but product type has a positive effect on the model prediction of energy consumption.

### *3.5.2. Shapely plots for size reduction model*

Fig. 10a depicts the mean Shapley displays the influence on the prediction output of the model, and Fig. 10b depicts the Bee swarm cum violin plot. In this particular instance, the SHAP values calculate the average effect that each feature has on the output of the model, which is the projected amount of energy consumption. It is possible to observe that product type and moisture are the most significant factors that affect the model of energy consumption, followed by temperature. The flow rate of water also has a significant impact on the model, but it is not quite as significant as the two things that came before it. Both the length of the bars and the




(a)




(b)

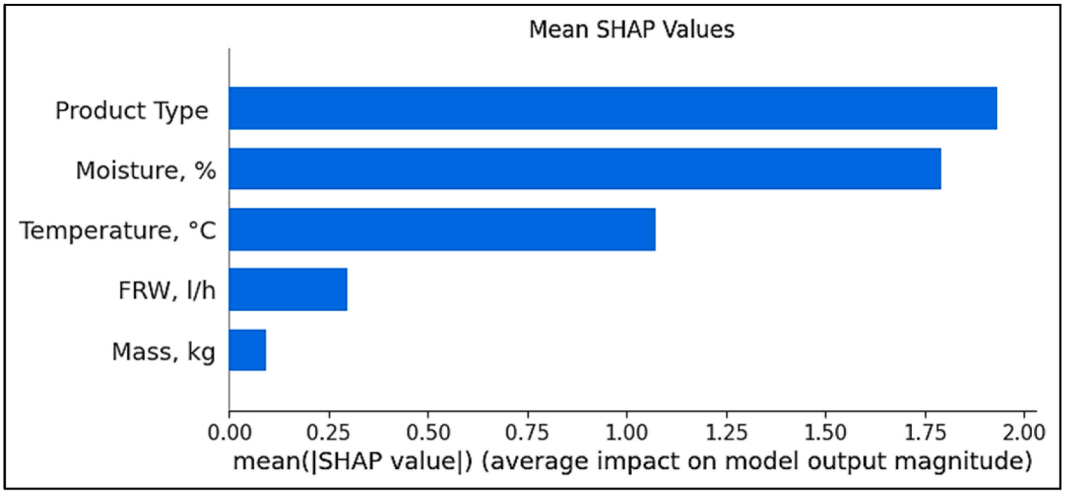
**Fig. 9.** Energy consumption model: (a) Mean Shapley values of each feature; (b) Bee swarm cum violin plots of each feature.

size of the data point for each SHAP value indicate the relative importance of each feature. When it comes to temperature and product kind, when the bars are longer, it indicates that there is a bigger effect on energy use. To assist in the identification of the nature of the impacts on the forecast, the bee swarm cum violin plot, which displays the kernel density of the plot, is used. As can be seen in Fig. 10b, the model's ability to forecast energy consumption is positively influenced by product type, and moisture content in the product has a greater impact on the model's ability to predict size reduction model.

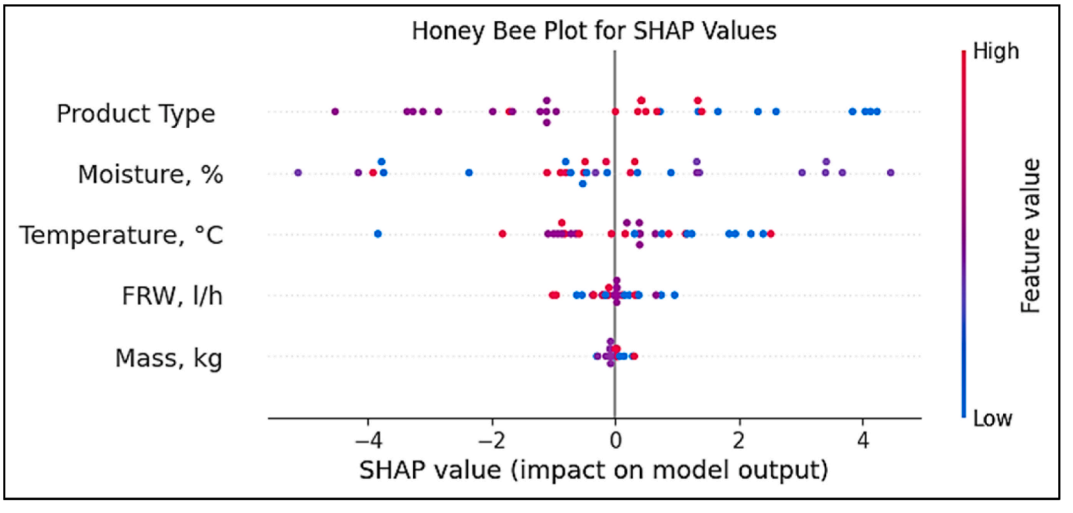
## 4. Conclusion

Some important implications regarding the application of machine learning algorithms, particularly XGBoost and LightGBM, in modeling and forecasting the drying process of banana slices produced by an indirect sun drier may be drawn from the study's findings.

- In both the training and testing phases, XGBoost routinely outperforms LightGBM in terms of $R^2$ values, Mean Squared Error (MSE), and Mean Absolute Error (MAE). In the training and testing phases, the $R^2$ values for XGBoost are 0.9957 and 0.9971, respectively; for LightGBM, they are 0.9061 and 0.8809.
- Compared to LightGBM, XGBoost yields much higher $R^2$ values, indicating better model performance and a higher percentage of variance explained by the independent variables. MSE results show that XGBoost had fewer errors in both the training and




(a)




(b)

**Fig. 10.** Size reduction model: (a) Mean Shapley values of each feature; (b) Bee swarm plots of each feature.

testing phases than LightGBM, with XGBoost scoring 0.0034 (training) and 0.0008 (testing) compared to 0.0747 (training) and 0.0337 (testing) for LightGBM.

- Comparatively, LightGBM has a Mean Absolute Error (MAE) of 0.1616 (testing) and 0.0137 (training), whereas XGBoost has 0.0137 and 0.0212, indicating that XGBoost produces predictions that are more accurate and less erroneous. LightGBM nevertheless provides respectable predictive capabilities, with $R^2$ values more than 0.88 in both the training and testing phases, even if it performs less well overall than XGBoost.
- When it comes to accurately predicting the drying process for banana slices in an indirect solar drier, both XGBoost and LightGBM show promising results. The results show how machine learning models, namely XGBoost, may be used practically to optimize drying conditions, enhance product quality, and use fewer resources while drying food.
- An explainable AI approach using Shapley values helped in the identification of temperature and product type as a major influential feature on the model prediction in the case of the energy consumption model.

To sum up, the results of this investigation show that machine learning algorithms - specifically, XGBoost—are effective in addressing the problems associated with drying banana chips. These models provide accurate projections for enhancing product quality and resource efficiency in addition to providing important insights into the relationships between the various components that affect the drying process. To increase the effectiveness and efficiency of banana chip drying processes, more research and testing may investigate novel machine learning techniques and modify model parameters.

## CRediT authorship contribution statement

**Van Giao Nguyen:** Writing – review & editing, Methodology, Conceptualization. **Prabhu Paramasivam:** Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marek Dzida:** Writing – review & editing. **Sameh M. Osman:** Writing – review & editing. **Duc Trong Nguyen Le:** Writing – review & editing, Conceptualization. **Dao Nam Cao:** Writing – review & editing. **Thanh Hai Truong:** Writing – review & editing. **Viet Dung Tran:** Writing – review & editing, Conceptualization.

## Declaration of competing interest

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

## Data availability

Data will be made available on request.

## Acknowledgment

This work has been supported by the Researchers Supporting Project RSP2024R405, King Saud University, Saudi Arabia.

## References

[1] B.H. Desai, B.K. Sidhu, Climate change as a common concern of humankind: some reflections on the international law-making process, in: Res. Handb. Glob. Clim. Const., Edward Elgar Publishing, 2019, pp. 153–175.
[2] W. Leal Filho, S.K. Tripathi, J. Andrade Guerra, R. Giné-Garriga, V. Orlovic Lovren, J. Willats, Using the sustainable development goals towards a better understanding of sustainability challenges, Int. J. Sustain. Dev. World Ecol. 26 (2019) 179–190.
[3] A.O. Akanmu, A.M. Akol, D.O. Ndolo, F.R. Kutu, O.O. Babalola, Agroecological techniques: adoption of safe and sustainable agricultural practices among the smallholder farmers in Africa, Front. Sustain. Food Syst. 7 (2023), https://doi.org/10.3389/fsufs.2023.1143061.
[4] C.V. Trappey, A.J.C. Trappey, H.-J. Lin, A.-C. Chang, Comparative analysis of food related sustainable development goals in the north Asia Pacific region, Food Ethics 8 (2023) 21, https://doi.org/10.1007/s41055-023-00132-8.
[5] J.L. Appelt, D.C. Garcia Rojas, P.H. Verburg, J. van Vliet, Socioeconomic outcomes of agricultural land use change in Southeast Asia, Ambio 51 (2022) 1094–1109, https://doi.org/10.1007/s13280-022-01712-4.
[6] Alami A. Hai, A. Ghani Olabi, S. Khuri, H. Aljaghoub, S. Alasad, M. Ramadan, et al., 3D printing in the food industry: recent progress and role in achieving sustainable development goals, Ain Shams Eng. J. 15 (2024) 102386, https://doi.org/10.1016/j.asej.2023.102386.
[7] J. Verschuuren, Agriculture, forestry and other land use (AFOLU), in: Res. Handb. Clim. Chang. Mitig. Law, Edward Elgar Publishing, 2022, pp. 433–456, https://doi.org/10.4337/9781839101595.00025.
[8] A.T. Hoang, P. Ashok, W.H. Chen, et al., Hydrogen production by water splitting with support of metal and carbon-based photocatalysts, ACS Sustain Chem Eng 11 (2023) 1221–1252, https://doi.org/10.1021/acssuschemeng.2c05226.
[9] A.T. Hoang, A. Pandey, F.J. Martinez De Osés, W.-H. Chen, Z. Said, K.H. Ng, et al., Technological solutions for boosting hydrogen role in decarbonization strategies and net-zero goals of world shipping: challenges and perspectives, Renew. Sustain. Energy Rev. 188 (2023) 113790, https://doi.org/10.1016/j.rser.2023.113790.
[10] A. Elahi Gol, M. Ščasný, Techno-economic analysis of fixed versus sun-tracking solar panels, Int. J. Renew. Energy Dev. 12 (2023) 615–626, https://doi.org/10.14710/ijred.2023.50165.
[11] N. Kannan, D. Vakeesan, Solar energy for future world: - a review, Renew. Sustain. Energy Rev. 62 (2016) 1092–1105, https://doi.org/10.1016/j.rser.2016.05.022.
[12] A.B. Al-Aasama, A. Ibrahim, U. Syafiq, K. Sopian, B.M. Abdulsahib, M. Dayer, Enhancing the performance of water-based PVT collectors with nano-PCM and twisted absorber tubes, Int. J. Renew. Energy Dev. 12 (2023) 891–901, https://doi.org/10.14710/ijred.2023.54345.
[13] L. Zhang, Y. Qiu, Y. Chen, A.T. Hoang, Multi-objective particle swarm optimization applied to a solar-geothermal system for electricity and hydrogen production; Utilization of zeotropic mixtures for performance improvement, Process Saf. Environ. Protect. 175 (2023) 814–833, https://doi.org/10.1016/j.psep.2023.05.082.
[14] A.T. Hoang, A. Pandey, E. Lichtfouse, V.G. Bui, I. Veza, H.L. Nguyen, et al., Green hydrogen economy: prospects and policies in Vietnam, Int. J. Hydrogen Energy 48 (2023) 31049–31062, https://doi.org/10.1016/j.ijhydene.2023.05.306.
[15] Y. Shi, W. Luo, Application of solar photovoltaic power generation system in maritime vessels and development of maritime tourism, Pol. Marit. Res. 25 (2018) 176–181, https://doi.org/10.2478/pomr-2018-0090.
[16] M.I. Ismail, N.A. Yunus, H. Hashim, Integration of solar heating systems for low-temperature heat demand in food processing industry – a review, Renew. Sustain. Energy Rev. 147 (2021) 111192, https://doi.org/10.1016/j.rser.2021.111192.
[17] A.M. Gandhi, S. Shanmugan, R. Kumar, A.H. Elsheikh, M. Sharifpur, A.K. Bewoor, et al., SiO2/TiO2 nanolayer synergistically trigger thermal absorption inflammatory responses materials for performance improvement of stepped basin solar still natural distiller, Sustain. Energy Technol. Assessments 52 (2022) 101974, https://doi.org/10.1016/j.seta.2022.101974.
[18] A.G.M.B. Mustayen, S. Mekhilef, R. Saidur, Performance study of different solar dryers: a review, Renew. Sustain. Energy Rev. 34 (2014) 463–470, https://doi.org/10.1016/j.rser.2014.03.020.
[19] S.K. Amit, M.M. Uddin, R. Rahman, S.M.R. Islam, M.S. Khan, A review on mechanisms and commercial aspects of food preservation and processing, Agric. Food Secur. 6 (2017) 51, https://doi.org/10.1186/s40066-017-0130-8.
[20] F. Chemat, N. Rombaut, A. Meullemiestre, M. Turk, S. Perino, A.-S. Fabiano-Tixier, et al., Review of green food processing techniques. Preservation, transformation, and extraction, Innovat. Food Sci. Emerg. Technol. 41 (2017) 357–377, https://doi.org/10.1016/j.ifset.2017.04.016.
[21] A. Zomorodian, D. Zare, H. Ghasemkhani, Optimization and evaluation of a semi-continuous solar dryer for cereals (Rice, etc), Desalination 209 (2007) 129–135, https://doi.org/10.1016/j.desal.2007.04.021.
[22] D. Pimentel, M. Pimentel, Energy use in food processing for nutrition and development, Food Nutr. Bull. 7 (1985) 1–10, https://doi.org/10.1177/156482658500700212.
[23] S. Simal, E. Deyá, M. Frau, C. Rosselló, Simple modelling of air drying curves of fresh and osmotically pre-dehydrated apple cubes, J. Food Eng. 33 (1997) 139–150, https://doi.org/10.1016/S0260-8774(97)00049-6.
[24] R.K. Goyal, A.R.P. Kingsly, M.R. Manikantan, S.M. Ilyas, Thin-layer drying kinetics of raw mango slices, Biosyst. Eng. 95 (2006) 43–49, https://doi.org/10.1016/j.biosystemseng.2006.05.001.
[25] J.A. Hernández, G. Pavón, M.A. García, Analytical solution of mass transfer equation considering shrinkage for modeling food-drying kinetics, J. Food Eng. 45 (2000) 1–10, https://doi.org/10.1016/S0260-8774(00)00033-9.

[26] M.M. Hussain, I. Dincer, Numerical simulation of two-dimensional heat and moisture transfer during drying of a rectangular object, Numer. Heat Tran. 43 (2003) 867–878, https://doi.org/10.1080/713838150.
[27] E. Tohidi, Application of Chebyshev collocation method for solving two classes of non-classical parabolic PDEs, Ain Shams Eng. J. 6 (2015) 373–379, https://doi.org/10.1016/j.asej.2014.10.021.
[28] N. Wang, J.G. Brennan, A mathematical model of simultaneous heat and moisture transfer during drying of potato, J. Food Eng. 24 (1995) 47–60, https://doi.org/10.1016/0260-8774(94)P1607-Y.
[29] S.M. Vahidhosseini, E. Barati, J.A. Esfahani, Green's function method (GFM) and mathematical solution for coupled equations of transport problem during convective drying, J. Food Eng. 187 (2016) 24–36, https://doi.org/10.1016/j.jfoodeng.2016.04.017.
[30] D. Alp, Ö. Bulantekin, The microbiological quality of various foods dried by applying different drying methods: a review, Eur. Food Res. Technol. 247 (2021) 1333–1343, https://doi.org/10.1007/s00217-021-03731-z.
[31] A.O. Omolola, A.I.O. Jideani, P.F. Kapila, DRYING KINETICS OF BANANA (Musa spp.), Interciencia 40 (2015) 374–380.
[32] F. Sarpong, X. Yu, C. Zhou, P. Oteng-Darko, L.P. Amenorfe, B. Wu, et al., Drying characteristic, enzyme inactivation and browning pigmentation kinetics of controlled humidity-convective drying of banana slices, Heat Mass Tran. 54 (2018) 3117–3130, https://doi.org/10.1007/s00231-018-2354-y.
[33] P. Jeet, G. Immanuel, O. Prakash, Effects of blanching on the dehydration characteristics of unripe banana slices dried at different temperature, Agric. Eng. Int. CIGR J. 17 (2015) 168–175.
[34] W.P. da Silva, C.M.D.P.S. e Silva, J.P. Gomes, Drying description of cylindrical pieces of bananas in different temperatures using diffusion models, J. Food Eng. 117 (2013) 417–424, https://doi.org/10.1016/j.jfoodeng.2013.03.030.
[35] C. Tunckal, İ. Doymaz, Performance analysis and mathematical modelling of banana slices in a heat pump drying system, Renew. Energy 150 (2020) 918–923, https://doi.org/10.1016/j.renene.2020.01.040.
[36] A. Lingayat, C. V P, Numerical investigation on solar air collector and its practical application in the indirect solar dryer for banana chips drying with energy and exergy analysis, Therm. Sci. Eng. Prog. 26 (2021) 101077, https://doi.org/10.1016/j.tsep.2021.101077.
[37] N. Kutlu, A. Isci, Drying characteristics of zucchini and empirical modeling of its drying process, Int J Food Stud 6 (2017) 232–244, https://doi.org/10.7455/ijfs/6.2.2017.a9.
[38] K.N. Cerci, D.B. Saydam, E. Hurdogan, Drying of mushroom slices in a new type solar drying system and under open sun: experimental and mathematical investigation, Eur Mech Sci 6 (2022) 221–232, https://doi.org/10.26701/ems.1144456.
[39] M.I.H. Khan, C.P. Batuwatta-Gamage, M.A. Karim, Y. Gu, Fundamental understanding of heat and mass transfer processes for physics-informed machine learning-based drying modelling, Energies 15 (2022) 9347, https://doi.org/10.3390/en15249347.
[40] S.M. Oliveira, T.R.S. Brandão, C.L.M. Silva, Influence of drying processes and pretreatments on nutritional and bioactive characteristics of dried vegetables: a review, Food Eng. Rev. 8 (2016) 134–163, https://doi.org/10.1007/s12393-015-9124-0.
[41] L.-Z. Deng, A.S. Mujumdar, Q. Zhang, X.-H. Yang, J. Wang, Z.-A. Zheng, et al., Chemical and physical pretreatments of fruits and vegetables: effects on drying characteristics and quality attributes – a comprehensive review, Crit. Rev. Food Sci. Nutr. 59 (2019) 1408–1432, https://doi.org/10.1080/10408398.2017.1409192.
[42] Z. Said, P. Sharma, A.K. Tiwari, V.V. Le, Z. Huang, V.G. Bui, et al., Application of novel framework based on ensemble boosted regression trees and Gaussian process regression in modelling thermal performance of small-scale Organic Rankine Cycle (ORC) using hybrid nanofluid, J. Clean. Prod. 360 (2022) 132194, https://doi.org/10.1016/j.jclepro.2022.132194.
[43] T. Han, P. Paramasivam, V.H. Dong, H. Cuong, D. Chuan, Harnessing a better future : exploring AI and ML applications in renewable energy, JOIV Int J Informatics Vis 8 (2024).
[44] D. Shetty, J.N. Sabhahit, Grey wolf optimization and incremental conductance based hybrid MPPT technique for solar powered induction motor driven water pump, Int. J. Renew. Energy Dev. 13 (2024) 52–61, https://doi.org/10.14710/ijred.2024.57096.
[45] P. Sharma, B.B. Sahoo, Z. Said, H. Hadiyanto, X.P. Nguyen, S. Nižetić, et al., Application of machine learning and Box-Behnken design in optimizing engine characteristics operated with a dual-fuel mode of algal biodiesel and waste-derived biogas, Int. J. Hydrogen Energy 48 (2023) 6738–6760, https://doi.org/10.1016/j.ijhydene.2022.04.152.
[46] A. Tuan Hoang, S. Nižetić, H. Chyuan Ong, W. Tarelko, V. Viet Pham, Le T. Hieu, et al., A review on application of artificial neural network (ANN) for performance and emission characteristics of diesel engine fueled with biodiesel-based fuels, Sustain. Energy Technol. Assessments 47 (2021) 101416, https://doi.org/10.1016/j.seta.2021.101416.
[47] G. Muhammad, A.D. Potchamyou Ngatcha, Y. Lv, W. Xiong, Y.A. El-Badry, E. Asmatulu, et al., Enhanced biodiesel production from wet microalgae biomass optimized via response surface methodology and artificial neural network, Renew. Energy 184 (2022) 753–764, https://doi.org/10.1016/J.RENENE.2021.11.091.
[48] J. Chaoraingern, V. Tipsuwanporn, A. Numsomran, Artificial intelligence for the classification of plastic waste utilizing TinyML on low-cost embedded systems, Int. J. Adv. Sci. Eng. Inf. Technol. 13 (2023) 2328–2337, https://doi.org/10.18517/ijaseit.13.6.18958.
[49] P.P. Biswas, W.-H. Chen, S.S. Lam, Y.-K. Park, J.-S. Chang, A.T. Hoang, A comprehensive study of artificial neural network for sensitivity analysis and hazardous elements sorption predictions via bone char for wastewater treatment, J Hazard Mater Adv 465 (2024) 133154, https://doi.org/10.1016/j.jhazmat.2023.133154.
[50] M. Mao, H. Va, M. Hong, Coefficient prediction for physically-based cloth simulation using deep learning, Int. J. Adv. Sci. Eng. Inf. Technol. 13 (2023) 1510–1517, https://doi.org/10.18517/ijaseit.13.4.19019.
[51] M. Bertolini, D. Mezzogori, M. Neroni, F. Zammori, Machine Learning for industrial applications: a comprehensive literature review, Expert Syst. Appl. 175 (2021) 114820.
[52] R. Rai, M.K. Tiwari, D. Ivanov, A. Dolgui, Machine learning in manufacturing and industry 4.0 applications, Int. J. Prod. Res. 59 (2021) 4773–4778, https://doi.org/10.1080/00207543.2021.1956675.
[53] A. Maleki, A. Haghighi, I. Mahariq, Machine learning-based approaches for modeling thermophysical properties of hybrid nanofluids: A comprehensive review, J Mol Liq. 322 (2021) 114843, https://doi.org/10.1016/j.molliq.2020.114843.
[54] T. Ma, Z. Guo, M. Lin, Q. Wang, Recent trends on nanofluid heat transfer machine learning research applied to renewable energy, Renew. Sustain. Energy Rev. 138 (2021) 110494, https://doi.org/10.1016/j.rser.2020.110494.
[55] M. Chae, H. Lee, A prediction of in-hospital cardiac arrest risk scoring based on machine learning, Int. J. Adv. Sci. Eng. Inf. Technol. 13 (2023) 895–900, https://doi.org/10.18517/ijaseit.13.3.17343.
[56] P. Asthana, B. Hazela, Applications of Machine Learning in Improving Learning Environment, Multimed Big Data Comput IoT Appl Concepts, Paradig Solut, 2020, pp. 417–433.
[57] T. Panch, P. Szolovits, R. Atun, Artificial intelligence, machine learning and health systems, J Glob Health 8 (2018).
[58] Z.K. Nisa, A.G. Pradipta, L.N. Sholikah, B.F. Pratama, A.S. Prihanantya, Ngadisih, et al., Recognition of agricultural land-use change with machine learning-based for regional food security assessment in kulon progo plains area, Int. J. Adv. Sci. Eng. Inf. Technol. 13 (2023) 54–61, https://doi.org/10.18517/ijaseit.13.1.16550.
[59] S.S. Hidayat, D. Rahmawati, M.C.A. Prabowo, L. Triyono, F.T. Putri, Determining the rice seeds quality using convolutional neural network, JOIV Int J Informatics Vis 7 (2023) 527, https://doi.org/10.30630/joiv.7.2.1175.
[60] W. Su, Q. Zhang, Y. Liu, Event-triggered adaptive neural network trajectory tracking control for underactuated ships under uncertain disturbance, Pol. Marit. Res. 30 (2023) 119–131, https://doi.org/10.2478/pomr-2023-0045.
[61] V.V. Vu, P.T. Le, T.M.T. Do, T.T.H. Nguyen, N.B.M. Tran, P. Paramasivam, et al., An insight into the application of AI in maritime and logistics toward sustainable transportation, JOIV Int J Informatics Vis 8 (2024) 158–174, https://doi.org/10.62527/joiv.8.1.2641.
[62] F. Okumuş, A. Ekmekçioğlu, S.S. Kara, Modelling ships main and auxiliary engine powers with regression-based machine learning algorithms, Pol. Marit. Res. 28 (2021) 83–96, https://doi.org/10.2478/pomr-2021-0008.
[63] L.D. Jathar, K. Nikam, U.V. Awasarmol, R. Gurav, J.D. Patil, K. Shahapurkar, et al., A comprehensive analysis of the emerging modern trends in research on

photovoltaic systems and desalination in the era of artificial intelligence and machine learning, Heliyon 10 (2024) e25407, https://doi.org/10.1016/j.heliyon.2024.e25407.
[64] W.-H. Chen, R. Aniza, A.A. Arpia, H.-J. Lo, A.T. Hoang, V. Goodarzi, et al., A comparative analysis of biomass torrefaction severity index prediction from machine learning, Appl. Energy 324 (2022) 119689, https://doi.org/10.1016/j.apenergy.2022.119689.
[65] A.M. Nassef, S. Md Atiqure Rahman, H. Rezk, Z. Said, ANFIS-based modelling and optimal operating parameter determination to enhance cocoa beans drying-rate, IEEE Access 8 (2020) 45964–45973, https://doi.org/10.1109/ACCESS.2020.2977165.
[66] R.P. Farias, R.S. Gomez, E.S. Lima, W.P. Silva, I.B. Santos, M.J. Figueredo, et al., Geometric and thermo-gravimetric evaluation of bananas during convective drying: an experimental investigation, Agriculture 12 (2022) 1181, https://doi.org/10.3390/agriculture12081181.
[67] W. Hao, H. Zhang, S. Liu, Y. Lai, Design and prediction method of dual working medium solar energy drying system, Appl. Therm. Eng. 195 (2021) 117153, https://doi.org/10.1016/j.applthermaleng.2021.117153.
[68] R.A. Chayjan, N. Dibagar, B. Alaei, Drying characteristics of zucchini slices under periodic infrared-microwave vacuum conditions, Heat Mass Tran. 53 (2017) 3473–3485, https://doi.org/10.1007/s00231-017-2081-9.
[69] Z. Tagnamas, A. Idlimam, A. Lamharrar, Predictive models of beetroot solar drying process through machine learning algorithms, Renew. Energy 219 (2023) 119522, https://doi.org/10.1016/j.renene.2023.119522.
[70] E.O. Oke, B.I. Okolo, O. Adeyi, F.N. Osuolale, P.C. Nnaji, C. Ude, et al., Parametric analysis and soft-computing prediction of sweet potatoes (Ipomoea batatas L) starch drying using machine learning techniques, SN Appl. Sci. 2 (2020) 1561, https://doi.org/10.1007/s42452-020-03378-7.
[71] C.E.K. Onu, P.T. Igbokwe, J. Nwabanne, O. Nwajinka C, E. Ohale P, Evaluation of optimization techniques in predicting optimum moisture content reduction in drying potato slices, Artif Intell Agric 4 (2020) 39–47, https://doi.org/10.1016/j.aiia.2020.04.001.
[72] E. Hürdoğan, K.N. Çerçi, D.B. Saydam, C. Ozalp, Experimental and modeling study of peanut drying in a solar dryer with a novel type of a drying chamber, Energy Sources, Part A Recover Util Environ Eff 44 (2022) 5586–5609, https://doi.org/10.1080/15567036.2021.1974126.
[73] M.P. Fabani, J.P. Capossio, M.C. Román, W. Zhu, R. Rodriguez, G. Mazza, Producing non-traditional flour from watermelon rind pomace: artificial neural network (ANN) modeling of the drying process, J. Environ. Manag. 281 (2021) 111915, https://doi.org/10.1016/j.jenvman.2020.111915.
[74] Z.-L. Liu, F. Nan, X. Zheng, M. Zielinska, X. Duan, L.-Z. Deng, et al., Color prediction of mushroom slices during drying using Bayesian extreme learning machine, Dry. Technol. 38 (2020) 1869–1881, https://doi.org/10.1080/07373937.2019.1675077.
[75] A. Daliran, M. Taki, A. Marzban, M. Rahnama, R. Farhadi, Experimental evaluation and modeling the mass and temperature of dried mint in greenhouse solar dryer; Application of machine learning method, Case Stud. Therm. Eng. 47 (2023) 103048, https://doi.org/10.1016/j.csite.2023.103048.
[76] P. Javadikia, S. Rafiee, A.T. Garavand, A. Keyhani, Modeling of moisture content in tomato drying procces by ANN-GA technique, in: 2011 1st Int. eConference Comput. Knowl. Eng., IEEE, 2011, pp. 162–165, https://doi.org/10.1109/ICCKE.2011.6413344.
[77] İ. Kırbaş, A.D. Tuncer, C. Şirin, H. Usta, Modeling and developing a smart interface for various drying methods of pomelo fruit (Citrus maxima) peel using machine learning approaches, Comput. Electron. Agric. 165 (2019) 104928, https://doi.org/10.1016/j.compag.2019.104928.
[78] R. Winiczenko, K. Górnicki, A. Kaleta, A. Martynenko, M. Janaszek-Mańkowska, J. Trajer, Multi-objective optimization of convective drying of apple cubes, Comput. Electron. Agric. 145 (2018) 341–348, https://doi.org/10.1016/j.compag.2018.01.006.
[79] C.K. Sankat, F. Castaigne, R. Maharaj, The air drying behaviour of fresh and osmotically dehydrated banana slices, Int. J. Food Sci. Technol. 31 (1996) 123–135, https://doi.org/10.1111/j.1365-2621.1996.332-35.x.
[80] A.A. Khoozani, A.E.-D.A. Bekhit, J. Birch, Effects of different drying conditions on the starch content, thermal properties and some of the physicochemical parameters of whole green banana flour, Int. J. Biol. Macromol. 130 (2019) 938–946, https://doi.org/10.1016/j.ijbiomac.2019.03.010.
[81] H. Majdi, J.A. Esfahani, Energy and drying time optimization of convective drying: taguchi and LBM methods, Dry. Technol. 37 (2019) 722–734, https://doi.org/10.1080/07373937.2018.1458036.
[82] Laskar A. Ahmad, M. Ahmed, D.-V.N. Vo, A. Abdullah, M. Shahadat, M.H. Mahmoud, et al., Mathematical modeling and regression analysis using MATLAB for optimization of microwave drying efficiency of banana, Therm. Sci. Eng. Prog. 46 (2023) 102157, https://doi.org/10.1016/j.tsep.2023.102157.
[83] P. Zhang, Y. Jia, Y. Shang, Research and application of XGBoost in imbalanced data, Int. J. Distributed Sens. Netw. 18 (2022) 155013292211069, https://doi.org/10.1177/15501329221106935.
[84] Y. Qiu, J. Zhou, M. Khandelwal, H. Yang, P. Yang, C. Li, Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration, Eng. Comput. 38 (2022) 4145–4162, https://doi.org/10.1007/s00366-021-01393-9.
[85] X. Li, L. Ma, P. Chen, H. Xu, Q. Xing, J. Yan, et al., Probabilistic solar irradiance forecasting based on XGBoost, Energy Rep. 8 (2022) 1087–1095, https://doi.org/10.1016/j.egyr.2022.02.251.
[86] O. Sagi, L. Rokach, Approximating XGBoost with an interpretable decision tree, Inf. Sci. 572 (2021) 522–542, https://doi.org/10.1016/j.ins.2021.05.055.
[87] T. Chen, C. Guestrin, XGBoost, in: Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., ACM, New York, NY, USA, 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.
[88] J. Dong, Y. Chen, B. Yao, X. Zhang, N. Zeng, A neural network boosting regression model based on XGBoost, Appl. Soft Comput. 125 (2022) 109067, https://doi.org/10.1016/j.asoc.2022.109067.
[89] H. Liu, Q. Xiao, Y. Jin, Y. Mu, J. Meng, T. Zhang, et al., Improved LightGBM-based framework for electric vehicle lithium-ion battery remaining useful life prediction using multi health indicators, Symmetry 14 (2022) 1584, https://doi.org/10.3390/sym14081584.
[90] B. Li, G. Chen, Y. Si, X. Zhou, P. Li, P. Li, et al., GNSS/INS integration based on machine learning LightGBM model for vehicle navigation, Appl. Sci. 12 (2022) 5565, https://doi.org/10.3390/app12115565.
[91] Z. Xiaochen, Y. Qizhi, Z. Fuqin, Q. Weiwen, Z. Kuikui, Ship speed prediction model based on LightGBM, J. Dalian Marit. Univ. 49 (2023) 56–65.
[92] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, et al., LightGBM: a highly efficient gradient boosting decision tree, in: I. Guyon, U Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, et al. (Eds.), Adv. Neural Inf. Process. Syst. 30, NeurIPS Proceedings, 2017.
[93] Y. Ju, G. Sun, Q. Chen, M. Zhang, H. Zhu, M.U. Rehman, A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting, IEEE Access 7 (2019) 28309–28318, https://doi.org/10.1109/ACCESS.2019.2901920.
[94] A. Shehadeh, O. Alshboul, R.E. Al Mamlook, O. Hamedat, Machine learning models for predicting the residual value of heavy construction equipment: an evaluation of modified decision tree, LightGBM, and XGBoost regression, Autom. ConStruct. 129 (2021) 103827, https://doi.org/10.1016/j.autcon.2021.103827.
[95] P.K. Kanti, P. Sharma, M.P. Maiya, K.V. Sharma, The stability and thermophysical properties of Al2O3-graphene oxide hybrid nanofluids for solar energy applications: application of robust autoregressive modern machine learning technique, Sol. Energy Mater. Sol. Cells 253 (2023) 112207, https://doi.org/10.1016/j.solmat.2023.112207.
[96] P. Kumar Kanti, P. Sharma, K.V. Sharma, M.P. Maiya, The effect of pH on stability and thermal performance of graphene oxide and copper oxide hybrid nanofluids for heat transfer applications: application of novel machine learning technique, J. Energy Chem. 82 (2023) 359–374, https://doi.org/10.1016/j.jechem.2023.04.001.
[97] A. de Myttenaere, B. Golden, B. Le Grand, F. Rossi, Mean absolute percentage error for regression models, Neurocomputing 192 (2016) 38–48, https://doi.org/10.1016/j.neucom.2015.12.114.
[98] P.K. Kanti, P. Sharma, B. Koneru, P. Banerjee, K.D. Jayan, Thermophysical profile of graphene oxide and MXene hybrid nanofluids for sustainable energy applications: model prediction with a Bayesian optimized neural network with K-cross fold validation, FlatChem 39 (2023) 100501, https://doi.org/10.1016/j.flatc.2023.100501.
[99] V.G. Nguyen, P. Sharma, Ü. Ağbulut, H.S. Le, T.H. Truong, M. Dzida, et al., Machine learning for the management of biochar yield and properties of biomass sources for sustainable energy, Biofuels, Bioprod Biorefining 18 (2024) 567–593, https://doi.org/10.1002/bbb.2596.
[100] Alabdullah A. Abdulalim, M. Iqbal, M. Zahid, K. Khan, M. Nasir Amin, F.E. Jalal, Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis, Construct. Build. Mater. 345 (2022) 128296, https://doi.org/10.1016/j.conbuildmat.2022.128296.
[101] D.D. Rufo, T.G. Debelee, A. Ibenthal, W.G. Negera, Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM), Diagnostics 11 (2021) 1714,

https://doi.org/10.3390/diagnostics11091714.
[102] V. Demir, H. Citakoglu, Forecasting of solar radiation using different machine learning approaches, Neural Comput. Appl. 35 (2023) 887–906, https://doi.org/10.1007/s00521-022-07841-x.
[103] A.M. Handhal, A.M. Al-Abadi, H.E. Chafeet, M.J. Ismail, Prediction of total organic carbon at Rumaila oil field, Southern Iraq using conventional well logs and machine learning algorithms, Mar. Petrol. Geol. 116 (2020) 104347, https://doi.org/10.1016/j.marpetgeo.2020.104347.
[104] A.M.-C. Wadoux, D.J.J. Walvoort, D.J. Brus, An integrated approach for the evaluation of quantitative soil maps through Taylor and solar diagrams, Geoderma 405 (2022) 115332, https://doi.org/10.1016/j.geoderma.2021.115332.
[105] V.G. Nguyen, P. Sharma, Ü. Ağbulut, H.S. Le, V.D. Tran, D.N. Cao, Precise prognostics of biochar yield from various biomass sources by Bayesian approach with supervised machine learning and ensemble methods, Int. J. Green Energy (2023) 1–25, https://doi.org/10.1080/15435075.2023.2297776.
[106] V.N. Nguyen, W. Tarełko, P. Sharma, A.S. El-Shafay, W.-H. Chen, P.Q.P. Nguyen, et al., Potential of explainable artificial intelligence in advancing renewable energy: challenges and prospects, Energy Fuels 38 (2024) 1692–1712, https://doi.org/10.1021/acs.energyfuels.3c04343.
[107] I. Ullah, K. Liu, T. Yamamoto, R.E. Al Mamlook, A. Jamal, A comparative performance of machine learning algorithm to predict electric vehicles energy consumption: a path towards sustainability, Energy Environ. 33 (2022) 1583–1612, https://doi.org/10.1177/0958305X211044998.
[108] P. Linardatos, V. Papastefanopoulos, S. Kotsiantis, Explainable ai: a review of machine learning interpretability methods, Entropy 23 (2020) 18.
[109] M.R. Zafar, N. Khan, Deterministic local interpretable model-agnostic explanations for stable explainability, Mach Learn Knowl Extr 3 (2021) 525–541, https://doi.org/10.3390/make3030027.
[110] D. Fryer, I. Strumke, H. Nguyen, Shapley values for feature selection: the good, the bad, and the axioms, IEEE Access 9 (2021) 144352–144360, https://doi.org/10.1109/ACCESS.2021.3119110.
[111] Y. Wu, Y. Zhou, Hybrid machine learning model and Shapley additive explanations for compressive strength of sustainable concrete, Construct. Build. Mater. 330 (2022) 127298, https://doi.org/10.1016/j.conbuildmat.2022.127298.
[112] M.S. Sameera, G. Rao Kancharla, The selection of best fit model involving corelation in examination with QQ plot, in: 2022 Int. Conf. Edge Comput. Appl., IEEE, 2022, pp. 814–817, https://doi.org/10.1109/ICECAA55415.2022.9936255.
[113] J.I. Marden, Positions and QQ plots, Stat. Sci. 19 (2004), https://doi.org/10.1214/088342304000000512.
[114] Y. Gebreyesus, D. Dalton, S. Nixon, D. De Chiara, M. Chinnici, Machine learning for data center optimizations: feature selection using Shapley additive exPlanation (SHAP), Future Internet 15 (2023) 88, https://doi.org/10.3390/fi15030088.
[115] Amiri S. Shams, M. Mueller, S. Hoque, Investigating the application of a commercial and residential energy consumption prediction model for urban Planning scenarios with Machine Learning and Shapley Additive explanation methods, Energy Build. 287 (2023) 112965, https://doi.org/10.1016/j.enbuild.2023.112965.