Comparative Predictive Analysis through Machine Learning in Solar Cooking Technology

Original Scientific Paper

Karankumar Chaudhari*
Department of Mechanical Engineering, G H Raisoni College of Engineering, Nagpur, India
karan.chaudhari01@gmail.com

Pramod Walke
Department of Mechanical Engineering, G H Raisoni College of Engineering, Nagpur, India
pramod.walke@raisoni.net

Sagar Shelare
Department of Mechanical Engineering, Priyadarshini College of Engineering, Nagpur, India
Centre for Research Impact and Outcomes, Chitkara University, Rajpura, Punjab, India
sagmech24@gmail.com

*Corresponding author

Abstract – *Renewable energy technology has helped solve global environmental issues in recent years. Solar cooking technology is a sustainable alternative to conventional cooking, particularly in regions with ample sunlight. Although there is a growing interest into solar cooking, however, there is a lack of comprehensive comparison research upon the machine learning models predictive accuracy. Prior studies frequently concentrate upon individual models or fail to conduct comprehensive comparative analyses, resulting in a knowledge deficit regarding the most effective predictive methodologies for solar cooking technology. This research article compares solar cooking with special types of cooking utensils used for indoor cooking by predictive analysis of different kinds of machine learning models. To achieve proper cooking, the temperature of both pan and pot is to be monitored constantly. For this, a machine learning (ML) system model was constructed for predicting pan and pot temperature as a response parameter. By leveraging datasets encompassing time duration of the cooking, mass flow rate of heat transfer fluid, type of heat transfer fluid, and global solar radiations, a range of machine learning algorithms, including decision tree regressor, linear regression, extreme gradient boosting, and random forest regressor algorithms, are employed for predicting pan and pot temperature of solar cookers. Extreme gradient boosting is the best machine learning model for solar utensil temperature, with maximum R2 and minimum mean squared error, mean absolute error, and root mean squared error values that perfectly predict all answers. Also, extreme gradient boosting predicts well on training and testing datasets, whereas Random forest predicts well on training datasets but poorly on test data, causing overfitting. This research shows that machine learning could revolutionize solar cooking technology, promising a future for renewable energy and sustainable living.*

Keywords: *Solar Cooking, Machine Learning, Regression Analysis, XGBoost, Statistical Analysis*

Received: February 12, 2024; Received in revised form: April 22, 2024; Accepted: April 23, 2024

1. INTRODUCTION

In recent years, the global quest for sustainable and eco-friendly practices has gained unprecedented momentum, prompting a critical reevaluation of conventional processes across various sectors. Clean, renewable energy is necessary to combat climate change, environmental degradation, and the depletion of fossil fuels. One of the domains that needs a paradigm shift is cooking. Traditional methods use environmentally harmful non-renewable energy. Solar energy in culinary applications overcomes environmental concerns connected with conventional fuel sources, reduces climate change, and supports global sustainable development. Modern cooking consumes considerable energy, adding to greenhouse gas emissions and resource depletion. This article examines the constraints of conventional cooking and the potential benefits of solar energy to demonstrate how sustainable energy solutions improve the culinary sector. Through an examination of existing research, technological advancements, and successful case studies, we will explore the multifaceted advantages of integrating solar energy into the cooking field. From reducing carbon footprints and minimizing reliance on finite energy sources to fostering community empowerment and technological innovation, the use of solar energy in cooking holds the promise of a more sustainable and socially responsible culinary future. So, the research has been conducted in the field of solar cooking.

To demonstrate recent technological developments and the present state of solar-based cooking technology, Aramesh et al. [1] provide a comprehensive assessment of current experimental and analytical economics research on solar cookers. With exemplary examples from India, a methodology for estimating the level of many incentives necessary to ensure the financial appeal of institutional solar cooking is described. In terms of cost to the government, an accelerated depreciation is demonstrated to be least expensive method for an incentivizing institutional solar cooking, followed by viability gap financing, interest subsidy, and investment tax credit in that order. Solar Dish Stirling Systems (SDSS) design requirements, thermal performance analysis, opto-geometrical parameters, techno-economic factors, and thermodynamic optimization are discussed. SDSS applications include hybridization and storage, solar power plants, solar cookery, water desalination, and micro co-generation. Solar cooking is a viable option since it is both economical and expandable. Arunachala et al. [2] give a survey of such cookers to unveil the cost-effective solar cooker concepts. Materials used in solar thermal storage include fatty acids paraffin and non-paraffin, hydrated salts as well as material that use the thermo-chemical processes, sensible heat energy. Ndukwu et al. [3] discuss the various exergy methodologies used for various solar systems such as solar still, hybrid solar water heating, solar dryers-heaters, solar cookery systems and solar space heating. Because greater temperatures are attained in a shorter period of time, parabolic solar cookers outperform conventional box solar cookers. Lentswe et al. [4] provide an in-depth evaluation of a thermal energy storage (TES) based parabolic solar cookers, which are sustainable cooking option for some underdeveloped nations. This study predicts pan and pot temperatures. For optimal solar cooking system operation, prior temperature information is helpful. ML algorithms are the most advanced prediction systems today. Many researchers utilise ML.

Qahwaji et al. [5] investigate the use of sunspot relationships and ML for autonomous short term of a prediction of solar flare. It uses ML to anticipate automated short-term solar flare retrieval and convert McIntosh categorization of each sunspot into a numerical representation for ML algorithms. Colak et al. [6] provide short term predictions of a big solar flares using automated hybrid computer system. A ML based system will analyze years of a sunspot and flare data to generate associations Ahmed et al.'s [7] work uses feature selection, ML, and advanced feature extraction to forecast solar flares. Flare prediction is more accurate than SMART MFs and ML. Bobra et al. [8] utilize a machine learning algorithm called Support Vector Machine (SVM) and data of four years from the Solar Dynamics Observatory's Magnetic and Helioseismic Imager. Researchers want to forecast X- and M-class solar outbursts. Voyant et al.'s [9] provided an overview of ML-based solar irradiation forecasting techniques. ML has recently advanced to the point that a wide range of solar prediction works have been produced. In the continental United States, seven sites, five climatic zones, and three sky conditions [10] are employed to evaluate hourly predicting performances of total 68 ML algorithm. In their evaluation of several ML regression algorithms, Cornejo-Bueno et al. [11] tackle the topic of estimating worldwide solar radiation using data from geostationary satellites. Viscondi et al. [12] present a literature review utilizing big data model to forecast generation of solar photovoltaic electricity. The review considers the data used to solve the problem and each project proposal. Artificial Neural Network (ANN), support vector machine (SVM), deep learning (DL), and K-nearest neighbor (KNN) are the four ML methods used in the study. The analysis found that the ANN algorithm fits best. However, all examined algorithms can reliably anticipate daily global solar radiation statistics. Mahmood et al. [13] describe the fundamentals of ML and standard operating procedures. Additionally, the author has made several recommendations that may improve ML's value for companies researching organic solar cells.

The use of ML in solar engineering is from last decades. Most of the researchers has used ML in prediction of solar radiation, solar power, and application of solar energy. The table 1 shows the authors with their applied ML algorithm and evaluation metrics. Umit et al. [14] has forecasted daily global sun radiation using the KNN, SVM, DL, and ANN ML algorithms. The author finds R^2 values between 0.855 and 0.936 for all four techniques. Cetina et al. [15] has applied ANN, SVM, and linear regression (LR) used to predict daily solar global radiations. Author assessment measures include R^2, root mean square error (RMSE), mean average error (MAE), and mean square error (MSE). Linear regression (LR) ML's maximum R^2 is 0.9917. Tagnamas et al [16] has predicted the two parameters such as atmosphere temperature and thickness of beetroot using catboost ML algorithm. Author has used R^2, RMSE, MSE, and MAE for the evaluation of ML algorithm purpose. The author gets R^2 value of 0.9999 for this algorithms. Ledmaoui et al. [17] has applied total six algorithms i.e. ANN, Support Vector Regression (SVR), Decision Tree (DT), Generalized Additive Model (GAM) Random Forest (RF), and Extreme Gradient Boosting (XGBOOST) to predict the electricity production of solar energy. The R2, RMSE, MSE and MAE are the evaluation metrics considered by the author. The maximum R^2 value for the XGB ML algorithm is 0.99. Elgendi et al. [18] has predicted the yield of solar still using ANN and LR ML algorithm.

The experiment has conducted on the pyramid solar still. Author has used R^2, RMSE, and MAE for the evaluation of ML algorithm purpose. The author gets R^2 value of 0.956 for ANN algorithms.

Kameni et al. [19] has used six algorithms i.e. LR, DT, SVM, DL, RF and Gradient Boosted Trees (GBT) to predict global solar radiation. The maximum R^2 value for the GBT ML algorithm is 0.985. Oh et al. [20] has predicted the diffuse and direct solar radiation using XGB, Light Gradient Boosting Machine (LGBM), Kier and ANN ML algorithm. Author has used R^2, RMSE, and MAE for the evaluation of ML algorithm purpose. The author gets R^2 value of 0.955 for Kier algorithms.

Table 1. Summary of machine learning applications in solar energy

Authors	Parameters	Response	ML Algorithm	R2	Evaluation Parameter
Ağbulut et al. [14]	daily maximum and minimum ambient temperature, daily extraterrestrial solar radiation, cloud cover, solar radiation and day length	daily global solar radiation	SVM, ANN, KNN and DL	0.855 to 0.936	R^2, rRMSE, RMSE, MABE, MAPE, and MBE
Cetina et al. [15]	solar irradiance, Solar dryer type, ambient temperature, relative humidity, and wind velocity	daily global solar radiation	ANN, SVM, LR	0.9917	R^2, MSE, MAE, RMSE
Tagnamas et al. [16]	absorber plate, drying chamber outlet air temperatures and solar collector outlet air	temperature and thickness of the beetroot slices	Catboost model	0.9999	R^2, MSE, MAE, RMSE
Ledmaoui et al. [17]	the irradiation, total energy, daily energy, and the temperature	solar energy production	ANN, SVR, RF, DT, XGB and GAM	0.99	R^2, MSE, MAE, RMSE
Elgendi et al. [18]	the atmosphere temperature, relative humidity, air velocity	the yield of solar still	ANN and LR	0.956	MAE, R^2, and RMSE
Kameni et al. [19]	wind speed (va), daily air temperature(ta), solar radiation, and relative humidity(rh)	global solar radiation	LM, DT, SVM, DL, RF, AND GBT	0.985	ARE,AAE,RMSE, and R^2
Oh et al. [20]	Relative humidity, Dry-bulb temperature, Extraterrestrial irradiance, Solar azimuth angle, Solar zenith angle, Turbidity, Clearness index	direct and diffuse solar irradiance	XGB, LGBM, ANN, KIER	0.955	RMSE, MAE, and R^2
Khosravi et al. [21]	local time, pressure, temperature, relative humidity, and wind speed	hourly solar irradiance	MLFFNN, RBFNN, SVR, FIS, and ANFIS	0.9999	RMSE, MAE, and R^2
Muhammed et al. [22]	sunshine, temperature, meteorological parameters and day number	global horizontal solar irradiation	MLP, ANFIS and SVM	0.85	RMSE, MSE, and R^2
Alhamrouni et al. [23]	--	Day temperature and solar radiation	SVM LR, KNN, and RF	0.9948	--
Feng et al. [24]	air temperature	global solar radiation	ANN, MNEA, RF, AND WNN	0.885	RMSE, MSE, and R^2, RRMSE, MAE
Quej et al. [25]	daily minimum and maximum air temperature, rainfall, and extraterrestrial solar radiation	daily global solar radiation	ANN, ANFIS and SVM	0.737	MSE, RMSE, MAE and R^2
Citakoglu [26]	calendar month number (M), average air temperature (Tmean), extraterrestrial radiation (Ra), and average relative humidity (RHmean)	Solar radiation	RF, KNN, XGB	0.9436	MAE, RMSE, R^2,

Khosravi et al. [21] has used radial basis function neural network (RBFNN), multilayer feed forward neural network (MLFFNN), SVR, adaptive neuro-fuzzy inference system (ANFIS), and fuzzy inference system (FIS) for the prediction of hourly based solar radiation. The maximum R^2 value for the ANFIS ML algorithm is 0.9999. Muhammed et al. [22] has predicted the global horizontal solar irradiation using Multi-layer perceptron (MLP), ANFIS and SVM ML algorithm. The sunshine, temperature, meteorological parameters and day number were the parameters considered for the prediction purpose. The author gets R^2 value of 0.85 for SVM algorithms. Alhamrouni et al. [23] has used LR, RF, KNN and SVM for the prediction of solar radiation and temperature. R^2, RMSE, MSE and MAE are the evaluation metrics considered. The maximum R2 value for the SVM ML algorithm is 0.9948. Feng Yu et al. [24] to predict the global solar radiation using ANN, mind evolutionary algorithm (MNEA), RF, Wavelet neural network (WNN) ML algorithm. The author gets R^2 value of 0.885 for ANN algorithms. Victor et al. [25] has used ANFIS, ANN and SVM for the prediction of daily based global solar radiations. The maximum R^2 value for the SVM ML algorithm is 0.737. Citakoghu [26] has predicted the solar radiation using RF, KNN and XGB ML algorithm. The extraterrestrial radiation (Ra), calendar month number (M), average relative humidity (RHmean), and average air temperature (Tmean) were the parameters considered for the prediction purpose. Author has used RMSE, R^2 and MAE for the evaluation of ML algorithm purpose. The author gets R^2 value of 0.9436 for XGB algorithms.

Despite the tremendous improvement in renewable energy technologies, machine learning in solar cooking still needs to be explored. Although there is a growing interest into solar cooking, however there are lack of comprehensive comparison research upon the machine learning models predictive accuracy. Prior studies frequently concentrate upon individual models or fail to conduct comprehensive comparative analyses, resulting in a knowledge deficit regarding most effective predictive methodologies for solar cooking technology. Most of the researchers has used ML algorithm for prediction of solar radiation and solar dryers. It gives scope of the use of such algorithms for solar cooking also. Also, data-driven methods to analyze and enhance it are still being determined. By utilizing machine learning predictive analysis, anyone can bridge the gap and gain valuable insights through data-driven methods.

This study led to the development of split-type solar cooking. The experiment examined pan and pot cooking performance. The study found that sun intensity varies with time of day. Heat transfer fluid type and oil mass flow rate are most significant. The four ML methods investigated were linear regression (LR), decision tree (RF), random forest (RF), and extreme gradient boosting. These programmes predicted the future using available data. This research found that the extreme gradient boosting algorithm has the lowest mean square error, root mean square error, mean absolute error, and greatest R^2 value. It was suggested to use XGBoost for the project.

2. RESEARCH METHODOLOGY

The solar cooking system is a need of the future. The day to day development has been takes place in this system. In this research, the indirect solar-powered cooking system has been created. As per discussion in the introduction, different types of cooking system are available. But it has been observed that, the research on the cooking utensils was not conducted. In this research, special types of cooking utensils were developed which can cook the Indian food. The basic cooking utensils for the Indian food are pan and pot. So, the research was carried out to develop the cooking pan and pot for the indirect solar cooking system, which gives the comfort of cooking food inside house like cooking on LPG gas.

The testing of these utensils were conducted on the indirect solar cooking system. The indirect solar cooking system was developed as shown in Fig 1. The solar cooking system consist of parts like parabolic dish collector, solar receiver, pump, pipelines, pan and pot. In order to check the performance of system, the temperature indicators were installed. In this system, the solar energy was collected by the parabolic dish collector and transferred to the solar receiver. The receiver gets heated due to solar energy. The heat transferred fluid i.e. Therminol 55 and Soyabean oil was used to transfer heat from solar receiver to cooking utensils. Therminol 55 and soybean oil have optimal heat transfer characteristics. Therminol 55 is a high-quality heat transfer fluid due to its excellent thermal stability, lower viscosity, and higher thermal conductivity. Soybean oil is an ideal heat transfer fluid and having widespread availability, lower cost, and environmentally friendly with the good thermal properties. These heat transfer oils have working temperature range from 200°C to 2500°C, also has high specific heat. In order to transfer the fluid, 0.5 hp centrifugal pump was used. The heat absorbed from the solar receiver was transfer to the utensils and utensils were gets heated. The heated utensils (heat from the utensils) were used to cook the food. The pan was used to cook the Indian food like roti, chapatti, paratha, dosa etc., while pot was used to cook the Indian food like, dal, rice, curry etc.

The testing was conducted in the month of April 2023 at Nagpur, India. The testing was conducted from 9 am to 5 pm. The solar intensity were recorded The observations were recorded after equal interval of one hour. As discussed earlier two different heat transfer fluid were used i.e. Therminol 55 and Soyabean oil. The specific heat is the important parameter for the selection of these fluid. The examinations were conducted by varying the mass flow rate of fluid. The table 2 shows parameters used for the prediction system. The original dataset of 54 size used for the study and to predict the temperature of pan and pot. The maximum obtained temperature of pan is 1920°C for 5 hours of heating till 1 PM with a solar intensity of 635 w/m2 and mass flow rate of 12 lpm.

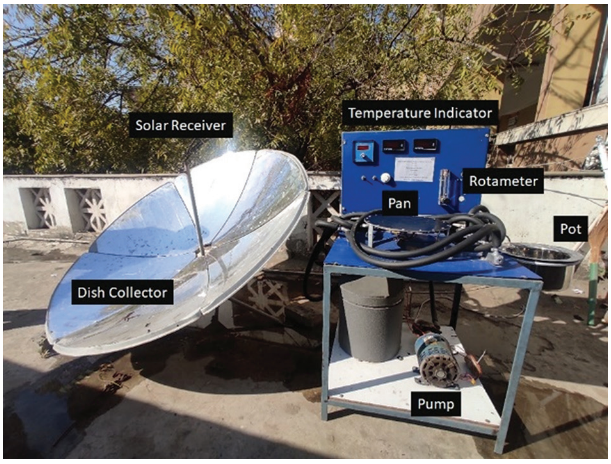

Fig.1. The Experimental Setup of Indirect Solar Cooking System with utensils

Table 2. Parameters used for the experimentation

	P1	P2	P3	P4	P5	P6	P7	P8	P9
Time (hr)	0	1	2	3	4	5	6	7	8
Solar Intensity (W/m2)	300	350	400	450	500	550	600	650	700
Mass Flow Rate (LPM)	6	9	12	-	-	-	-	-	-
HTF (kJ/KgK)	1.98	2.54	-	-	-	-	-	-	-

3. MACHINE LEARNING APPROACH

In this study, ML approach including LR, DT, RF and XGB were used to predict the temperature of cooking utensils i.e. pan and pot and select the suitable algorithm for accurate prediction. The suggested regression learning methodology utilized in the ML.

3.1. LINEAR REGRESSION

Two hypotheses make use of the linear regression approach [27]. Researchers first apply linear regression analyses in forecasting and prediction, where they closely resemble the use of ML. Linear regression analysis is a useful tool in some scenarios to ascertain the relationship between the dependent and independent variable. It is important to consider that regressions demonstrate relationships between dependent variables and a defined dataset comprising various factors.

Linear regression models [28] predict the dependent variables based on the independent variables. Linear regression analysis is used to estimate value of dependent variable, y, since independent variables, x, has a range of values [29, 30]. There are two categories of regression [31]: simple linear regression and polynomial linear regression. In this study, simple linear regression is used.

3.1.1. Simple Linear Regression

A simple linear regression is a model with only one independent variable [32]. Simple linear regression defines the variable's dependence as $y=\beta 0+\beta 1x+\varepsilon$. The effect of independent factors is distinguished from interaction of dependent variables by simple regression. Fourier multivariate regression (MLR) is statistical method that uses many explanatory factors to predict answer variable's outcome. Modeling the linear connection between independent variables x and dependent variable y that will be examined is the goal of MLR.

3.2. DECISION TREE REGRESSOR

Among many ML approaches is the decision tree. Although this approach is usually applied to classification data, it may also be applied to regression data. An approach that is transparent and simple to comprehend is the DT technique, as opposed to employing an artificial neural network as a black box.

The aim in this study is a continuous value since a decision tree technique is utilized for a regression problem. In order to minimize the impurity function and choose the best sites for future data splits, regression criteria such as mean squared error (MSE) and mean absolute error (MAE) may be used. The mean values for MSE can be used to minimize an error [33].

However, this approach has problems with stability, scalability, and robustness when it comes to large-scale data processing [34]. The utilization of extensive data samples leads to increased complexity, which must be addressed. To reduce the complexity of a decision tree, metrics such as total number of leaves, the total number of nodes, number of attributes, and tree depth can be adjusted [35]. Ensemble DT are utilized instead as they are more reliable and can handle these problems in some situations.

3.3. RANDOM FOREST REGRESSION

A supervised learning technique called Random Forest may be used to solve decision tree and classification issues. A "Random Forest" is a collection of numerous trees, where each tree depends on the value of a random vector, which is equally and independently sampled from each tree in the forest. [36]. By combining many decision trees, the random forest method may greatly improve the decision tree's predictive performance [37].

Two reasons contribute to the randomness of this algorithm:(1) each split node in the DT formation process selects sample chunk of m variables from the original data set, and the best one is used in that node; (2) every tree develops at random on a distinct bootstrap sample derived from the training set. A useful ML technique for prediction is Random Forest. The RFR Model is suggested by Harrison et al. [38] for nutrient concentration estimation utilising high-frequency sensor data. Since the method is suitable for multivariate datasets with multicollinearity among predictors, nonlinear correlations between predictor and response variables, and highly skewed data, it is well-suited for this application. The benefit of Random Forest Regression over least squares regression, according to the study in [39], is higher R squared (R^2) value.

3.4. EXTREME GRADIENT BOOSTING

Chen and Guestrin [40] developed the XGBoost method. Given its efficacy as a tree-based ensemble learning method, data scientists view it as a potent instrument. Based on gradient boosting architecture [41], XGBoost estimates the outcomes makes use of a variety of complement functions.

3.5. PERFORMANCE EVALUATION

Three performance statistical error functions, including the coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE), were taken into consideration in order to assess the performance of the ML models under examination. Generally speaking, five-fold cross-validation (CV) involves randomly dividing all of the data into k folds ($k = 5$ in this example), training the model on the k - 1 folds, and leaving one fold for testing. There are k repetitions of this process. But before any data is utilised in this study, it is divided into training and testing datasets, with the purpose of using the training dataset for cross-validation. The 95% confidence intervals (CI) and model accuracy are estimated using the repeated cross-validation procedure [55]. For every model, the 5-fold cross-validation is carried out 50 times. CV accuracy is the average of all repetitions, and 95% confidence intervals are computed from the repeated cross-validation data. The last step in assessing the model's performance is to determine if the testing accuracy falls within the 95% confidence interval. The model is deemed acceptable if the testing accuracy is within the 95% confidence interval. If the testing accuracy falls outside of this range and the difference is statistically significant, underfitting or overfitting is thought to be present. The cross-validation process does not use the testing data, which is a separate dataset.

4. RESULTS AND DISCUSSION

As previously stated, the goal of this research is to predict the temperature of the pan (Tpn) and the pot (Tpt) in a solar cooking system using four ML models: linear regression (LR), decision tree (DT), random forest (RF), and XGBoost (XGB). First, using the Scikit-Learn Python module, the following two PR-based meta-models are created based on the training dataset.

$$T_{pn} = 11.493 \times T + 0.301 \times E - 1.415 * mf - 8.396 \times htf - 25.162 \quad (1)$$

$$T_{pt} = 10.09 \times T + 0.265 \times E - 1.24 * mf - 7.192 \times htf - 19.527 \quad (2)$$

The first ML model applied for the data given in table 2 is Linear regression model. In this model, the sklearn library was used. In order to consider intercept for this model, fit intercept is considered as a True in nature. The normalize is kept deprecated to fit the model intercepted. For the faster computation, number of jobs is considered as a 2. The 80% of data i.e. 43 samples are used for the training purpose while 20% of data i.e. 11 samples are used for the testing purpose. A machine learning model was developed and tested for predicting the temperature of a pan and pot. The regression equation obtain from the model can be seen in equation 1 and 2 for the T_{pn} and T_{pt} respectively.

Decision Tree Machine Learning algorithm also applied on the given data. The decision tree is one of the advanced technique of the regression model. It is node based algorithm. In order to apply the algorithm on the data, the criteria for evaluation was considered as a "squared error", which helps to reduce the variance as a feature selection and minimize the $L2$ loss. The "best" strategy is considered for the split at node. The maximum depth of the tree is restricted to 10, in order to avoid overfitting of model. The minimum sample split is set default as 1. This all parameters are considered while developing decision tree algorithm. Here also, the data is divided as 80% for training and 20% for testing.

A machine learning algorithm has also been applied to the provided data. The decision tree is one of the advanced technique of the regression model. It is node based algorithm. In order to apply the algorithm on the data, the criteria for evaluation was considered as a "squared error", which helps to reduce the variance as a feature selection and minimize the $L2$ loss. The "best" strategy is considered for the split at node. The maximum depth of the tree is restricted to 10, in order to avoid overfitting of model. The minimum sample split is set default as 1. This all parameters are considered while developing decision tree algorithm. Here also, the data is divided as 80% for training and 20% for testing.

The random forest is a bagging techniques. The bagging techniques helps to improve the accuracy and overcome the problem of overfitting in the decision tree. The random forest regression model was applied on the data with spilt of 80:20 for training and testing. "n_estimator" is set to 100 with criterion "squared error" to run multiple decision trees in parallel and determine the final outcome. The depth of the tree was restricted to 10 with minimum sample split 2 and minimum sample leaf as 1. The maximum features considered as an "auto" means all available features are considered for the model. In order to avoid the overfitting, pruning takes place. All these parameters were considered for the development of random forest regression ML model.

The other method to improve the performance of decision tree algorithm is boosting techniques. In boosting techniques series approach was used. The output of one tree is used for the nest decision tree. One of such algorithm was used for testing of data. The XGBoost algorithm is one of the most advance boosting algorithm. For this algorithm, "n_estimator" considered as 100 while criterion as "squared error". The depth of the tree was restricted to 10 with minimum sample split 2 and minimum sample leaf as 1. The maximum features considered as an "auto" means all available features are considered for the model. The learning rate as 0.1 and "n_job" as a 10 in order to speed up the performance of the algorithm.

An effort is now made to estimate the values of T_{pn} and T_{pt} for the solar kitchenware once all created ML models have been properly trained using the dataset under consideration. For each of these ML models, Table 2 shows a predicted and target responses values. Plotting a predicted and target values for T_{pn} and T_{pt}, respectively, allows for a more clear understanding of the prediction performance of the ML models. These numbers show that within a ± 15% error band, all of the created ML models can accurately anticipate both of these answers.

As can be seen from Fig. 2, all of the ML models had almost excellent estimates for Tpn prediction on training data, with the majority of the data points either hugging or resting upon a diagonal identity line. But, in LR ML model, some training points are not on the line while three points are beyond the ±15% error band. But in case of RF only 3 points are found such that they are not on line but are in the ±15% error band. The other two ML model DT and XGB are tuning perfectly with the line and almost all the data points are on the line. This shows that the DT and XGB model has good generalisation and no overtraining. Both ML model shows identical performance.

When ML models predict Tpt, a trend comparable to that of Tpn is observed. The LR ML model, have quite similar prediction as seen as for Tpn. There are most of the data points are on the line and some are beyond the line. Out of some distracted data points only three data points are outside the error band which can be seen poor prediction towards the loser data points. The better prediction of Tpt can be seen for the RF model than the LR model. In RF no data points are beyond the error band. The DT and XGB shows the best performance model than LR and RF. Most of the data points of Tpt can be seen on the line.

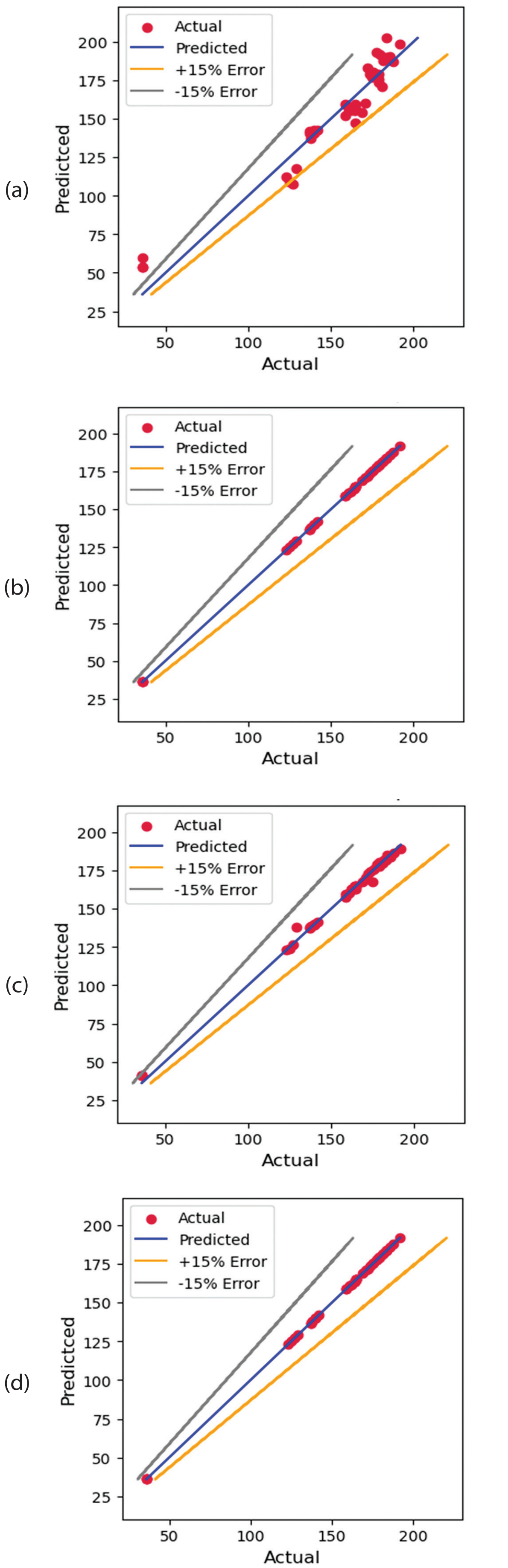

Fig. 2. Target vs. predicted T_{pn} values comparison for (a) LR, (b) DT, (c) RF and (d) XGB

Fig. 3 shows the residuals, or differences between the goal and anticipated response values, for each of the created ML models. The zero line in Fig. 3 denotes zero prediction error, while the points above and below it shows underprediction (i.e., predicted value less than target value) and overprediction (i.e., projected value more than target value), respectively. This is important to notice. All of the ML models' test data performances—aside from LR's—are generally comparable to their related data performances. This suggests that the instruction is sufficient. Additionally, there is no discernible pattern in the residuals' dispersion, which suggests that there is no bias.

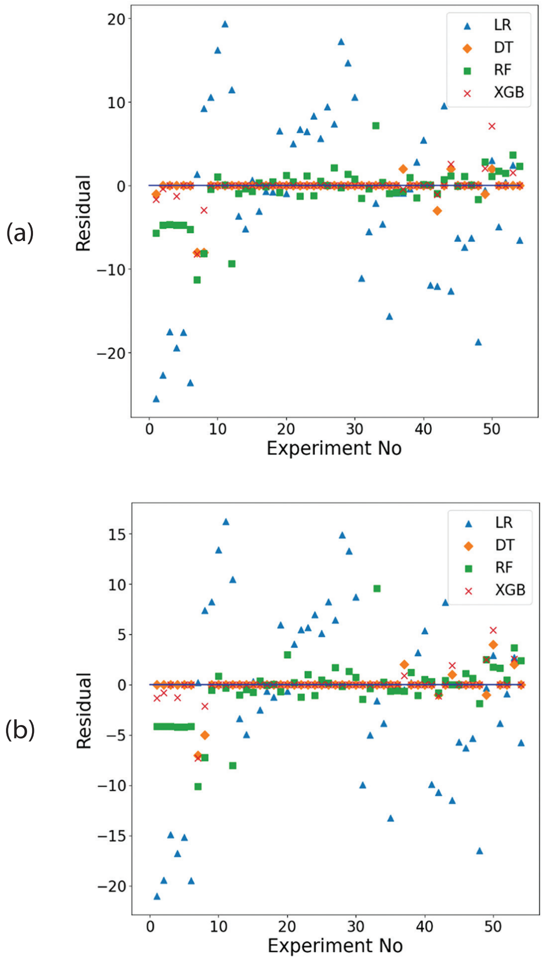

Fig 3. (a) Residuals of predicted T_{pn} and (b) residuals of predicted T_{pt}

Now that Figs. 2, and 3 have been closely examined, it is clear that the RF, DT, and XGB ML models all perform rather well in terms of T_{pn} prediction. The predicted T_{pn} values for each of these three ML models exhibit incredibly little deviations from the matching goal values. With minor variations of the projected Tpn values from the goal, the implementation of the LR ML model yields average prediction results. T_{pt} response is also observed in a similar manner. It is challenging to identify which of the produced ML models has the greatest prediction performance for the case under consideration simply by looking at the aforementioned numbers. To do this, the values of four model accuracy metrics—MSE, RMSE, MAE, and R^2—are calculated, as shown in Table 3. It is important to note that lower values of MSE, RMSE, and MAE and higher values of R^2 are always preferred for any of the predictive models [42].

Table 3. Metrics representing the accuracy of the models for both responses

Response	ML Model	Dataset	MSE	RMSE	MAE	R2
Tpn	LR	Testing	176.89	13.30	9.98	0.96
		Overall	117.13	10.82	8.51	0.94
	DT	Testing	13.00	3.61	2.82	1.00
		Overall	2.65	1.63	0.57	1.00
	RF	Testing	26.89	5.19	4.09	0.99
		Overall	10.10	3.18	1.99	1.00
	XGB	Testing	13.23	3.64	2.66	1.00
		Overall	2.69	1.64	0.54	1.00
Tpt	LR	Testing	128.82	11.35	8.49	0.96
		Overall	85.78	9.26	7.31	0.95
	DT	Testing	13.18	3.63	2.00	1.00
		Overall	2.69	1.63	0.41	1.00
	RF	Testing	20.90	4.57	3.56	0.99
		Overall	8.89	2.98	1.86	0.99
	XGB	Testing	10.04	3.17	2.48	1.00
		Overall	2.05	1.43	0.51	1.00

Table 3 shows that, when it comes to Tpn, the XGB ML model has the best R2 values, coming in at 0.9997 on the training dataset. Based on the training dataset, the MSE, RMSE, and MAE corresponding values of 1.9423, 1.3936, and 0.3718, respectively, further support the outstanding performance of the XGB ML model. However, DT has the highest R2 values at 0.9977 when taking into account the performance in relation to the test dataset. The minimal MSE, RMSE, and MAE values for the identical test data are 13, 3.605, and 2.4212, respectively. Based on the entire dataset (training and testing), Tpn prediction shows that XGB is the best ML model, with the highest R2 (0.9987) and lowest MSE (2.6945), RMSE (1.6415), and MAE (0.5424) values. Additionally, the LR ML model has the lowest R2 (0.9424) and the poorest MAE (8.5133), RMSE (10.8226), and MSE (117.13) values based on the entire dataset. In summary, the DT and RF ML models come in second and third place, respectively, when it comes to predicting Tpn values throughout the whole dataset.

Similar findings are also observed when estimating Tpt values for the solar cooking tool. On the training dataset with the maximum R2, minimum MSE, RMSE, and MAE values, and maximum R2 accuracy, the XGB performs best. However, XGB proves to be the most accurate ML model when it comes to predicting Tpt values using the test data, with the highest R2 (0.9967) and lowest MSE (10.04), RMSE (3.1688), and MAE (2.4817) values. In terms of all model accuracy measures, XGB performs the best across the board for the dataset, with DT and RF ML models following closely behind. Out of all four measures, the LR displays the poorest results for the whole dataset. As a result, it is seen that LR's performance is very variable for both Tpn and Tpt replies. It's interesting to note that while it performs well on training datasets, it performs poorly on testing and general datasets. On the other hand, for the two replies that are being examined, XGB consistently possesses an accuracy level over the whole dataset.

5. CONCLUSION

This work develops 4 ML models—linear regression, decision tree, random forest, and extreme gradient boosting to accurately predict solar cooking utensil temperatures. For the utensils temperature that is temperature of Pan, and temperature of pot; duration of time, solar intensity, type of heat transfer fluid, and mass flow rate of heat transfer fluid are treated as the input parameters. The prediction performance of the four developed ML models is compared in terms of four model accuracy metrics—R2, mean squared error, root mean squared error, and mean absolute error using these experimental datasets as a basis. Based on the comprehensive comparative analysis of the Machine Learning models' overall performance, the following conclusions can be drawn:

- In case of both the temperature of solar utensils, extreme gradient boosting emerges out as the best machine learning model with maximum R2 and a minimum value of mean squared error, mean absolute error and root mean squared error values perfectly predict all of the answers that are being considered.
- For prediction, extreme gradient boosting consistently yields good results on training and testing datasets.
- Moreover, while random forest performs exceptionally well in predictions on training datasets, its accuracy on test data is low, which causes the model to become over fit.
- Extreme gradient boosting has a wide range of tuning parameters, yet it may be argued that, for forecasting response values of the temperature of utensils under consideration, it is an effective prediction tool. Finding the ideal mix of those tuning parameters to get the highest prediction performance out of the extreme gradient boosting machine learning model is still a difficult challenge.

In the future, it may be possible to investigate the application potential of more machine learning models, such as the Adaboost regressor, support vector regressor, and Naïve Bias regressor, and compare how well they predict response values for utensil temperature. The demonstrative examples provided use tiny datasets, each with a mere 54 experimental observations, to validate the prediction performance of all the ML models. A data repository with a sizable amount of experimental data may be created in order to improve the image and be used for the training and testing of those ML models.

6. REFERENCES

[1] M. Aramesh et al. "A review of recent advances in solar cooking technology", Renewable Energy, Vol. 140, 2019, pp. 419-435.

[2] U. C. Arunachala, A. Kundapur, "Cost-effective solar cookers: A global review", Solar Energy, Vol. 207, pp. 903-916, 2020.

[3] M. C. Ndukwu, L. Bennamoun, M. Simo-Tagne, "Reviewing the exergy analysis of solar thermal systems integrated with phase change materials", Energies, Vol. 14, No. 3, 2021, p. 724.

[4] K. Lentswe, A. Mawire, P. Owusu, A. Shobo, "A review of parabolic solar cookers with thermal energy storage", Heliyon, Vol. 7, No. 10, 2021.

[5] R. Qahwaji, T. Colak, "Automatic short-term solar flare prediction using machine learning and sunspot associations", Solar Physics, Vol. 241, 2007, pp. 195-211.

[6] T. Colak, R. Qahwaji, "Automated Solar Activity Prediction: A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares", Space Weather, Vol. 7, No. 6, 2009.

[7] O. W. Ahmed, R. Qahwaji, T. Colak, P. A. Higgins, P. T. Gallagher, D. S. Bloomfield, "Solar flare prediction using advanced feature extraction, machine learning, and feature selection", Solar Physics, Vol. 283, 2013, pp. 157-175.

[8] M. G. Bobra, S. Couvidat, "Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm", Astrophysical Journal, Vol. 798, No. 2, 2015, p. 135.

[9] C. Voyant et al. "Machine learning methods for solar radiation forecasting: A review", Renewable Energy, Vol. 105, 2017, pp. 569-582.

[10] G. M. Yagli, D. Yang, D. Srinivasan, "Automatic hourly solar forecasting using machine learning models", Renewable and Sustainable Energy Reviews, Vol. 105, 2019, pp. 487-498.

[11] L. Cornejo-Bueno, C. Casanova-Mateo, J. Sanz-Justo, S. Salcedo-Sanz, "Machine learning regressors for solar radiation estimation from satellite data", Solar Energy, Vol. 183, 2019, pp. 768-775.

[12] G. de Freitas Viscondi, S. N. Alves-Souza, "A Systematic Literature Review on big data for solar photovoltaic electricity generation forecasting", Sustainable Energy Technologies and Assessments, Vol. 31, 2019, pp. 54-63.

[13] A. Mahmood, J.-L. Wang, "Machine learning for high performance organic solar cells: current scenario and future prospects", Energy & Environmental Science, Vol. 14, No. 1, 2021, pp. 90-105.

[14] Ü. Ağbulut, A. E. Gürel, Y. Biçen, "Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison", Renewable and Sustainable Energy Reviews, Vol. 135, 2021, p. 110114.

[15] A. J. Cetina-Quiñones, G. Santamaria-Bonfil, R. A. Medina-Esquivel, A. Bassam, "Techno-economic analysis of an indirect solar dryer with thermal energy storage: An approach with machine learning algorithms for decision making", Thermal Science and Engineering Progress, Vol. 45, 2023, p. 102131.

[16] Z. Tagnamas, A. Idlimam, A. Lamharrar, "Predictive models of beetroot solar drying process through machine learning algorithms", Renewable Energy, Vol. 219, 2023, p. 119522.

[17] Y. Ledmaoui, A. El Maghraoui, M. El Aroussi, R. Saadane, A. Chebak, A. Chehri, "Forecasting solar energy production: A comparative study of machine learning algorithms", Energy Reports, Vol. 10, 2023, pp. 1004-1012.

[18] M. Elgendi, M. Atef, "Calculating the impact of meteorological parameters on pyramid solar still yield using machine learning algorithms", International Journal of Thermofluids, Vol. 18, 2023, p. 100341.

[19] M. K. Nematchoua, J. A. Orosa, M. Afaifia, "Prediction of daily global solar radiation and air temperature using six machine learning algorithms; a case of 27 European countries", Ecological Informatics, Vol. 69, 2022, p. 101643.

[20] M. Oh et al. "Analysis of minute-scale variability for enhanced separation of direct and diffuse solar irradiance components using machine learning algorithms", Energy, Vol. 241, 2022, p. 122921.

[21] A. Khosravi, R. N. N. Koury, L. Machado, J. J. G. Pabon, "Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms", Journal of Cleaner Production, Vol. 176, 2018, pp. 63-75.

[22] M. A. Hassan, A. Khalil, S. Kaseb, M. A. Kassem, "Potential of four different machine-learning algorithms in modeling daily global solar radiation", Renewable Energy, Vol. 111, 2017, pp. 52-62.

[23] Y. Zahraoui, I. Alhamrouni, S. Mekhilef, M. R. B. Khan, "Machine learning algorithms used for short-term PV solar irradiation and temperature forecasting at microgrid", Applications of AI and IoT in Renewable Energy, Elsevier, 2022, pp. 1-17.

[24] Y. Feng, D. Gong, Q. Zhang, S. Jiang, L. Zhao, N. Cui, "Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation", Energy Conversion and Management, Vol. 198, 2019, p. 111780.

[25] V. H. Quej, J. Almorox, J. A. Arnaldo, L. Saito, "ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm subhumid environment", Journal of Atmospheric and Solar-Terrestrial Physics, Vol. 155, 2017, pp. 62-70.

[26] H. Citakoglu, "Comparison of artificial intelligence techniques via empirical equations for prediction of solar radiation", Computers and Electronics in Agriculture, Vol. 118, 2015, pp. 28-37.

[27] D. Wu, J. Liu, Y. Yang, Y. Zheng, "Nitrogen/Oxygen Co-Doped Porous Carbon Derived from Biomass for Low-Pressure CO2Capture", Industrial & Engineering Chemistry Research, Vol. 59, No. 31, 2020, pp. 14055-14063.

[28] H. Roopa, T. Asha, "A linear model based on principal component analysis for disease prediction", IEEE Access, Vol. 7, 2019, pp. 105314-105318.

[29] G. A. F. Seber, A. J. Lee, "Linear regression analysis", John Wiley & Sons, 2003.

[30] D. C. Montgomery, E. A. Peck, G. G. Vining, "Introduction to linear regression analysis", John Wiley & Sons, 2021.

[31] S. Kavitha, S. Varuna, R. Ramya, "A comparative analysis on linear regression and support vector regression", Proceedings of the Online International Conference on Green Engineering and Technologies, 2016, pp. 1-5.

[32] A. Abdulazeez, B. Salim, D. Zeebaree, D. Doghramachi, "Comparison of VPN Protocols at Network Layer Focusing on Wire Guard Protocol", International Journal of Interactive Mobile Technologies, Vol. 14, No. 18, 2020.

[33] A. J. Smola, B. Schölkopf, "A tutorial on support vector regression", Statistics and Computing, Vol. 14, 2004, pp. 199-222.

[34] T. Aluja-Banet, E. Nafria, "Stability and scalability in decision trees", Computational Statistics, Vol. 18, No. 3-4, 2003, pp. 505-520.

[35] S. B. Kotsiantis, "Decision trees: a recent overview", Artificial Intelligence Review, Vol. 39, 2013, pp. 261-283.

[36] L. Breiman, "Random forests", Machine Learning, Vol. 45, 2001, pp. 5-32.

[37] G. James, D. Witten, T. Hastie, R. Tibshirani, "An introduction to statistical learning", Vol. 112, Springer, 2013.

[38] J. W. Harrison, M. A. Lucius, J. L. Farrell, L. W. Eichler, R. A. Relyea, "Prediction of stream nitrogen and phosphorus concentrations from high-frequency sensors using Random Forests Regression", Science of The Total Environment, Vol. 763, 2021, p. 143005.

[39] M. Schonlau, R. Y. Zou, "The random forest algorithm for statistical learning", Stata Journal, Vol. 20, No. 1, 2020, pp. 3-29.

[40] T. Chen, C. Guestrin, "Xgboost: A scalable tree boosting system", Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13-17 August 2016, pp. 785-794.

[41] J. H. Friedman, "Greedy function approximation: a gradient boosting machine", The Annals of Statistics, Vol. 29, No. 5, 2001, pp. 1189-1232.