



OPEN

Prediction of arabica coffee production using artificial neural network and multiple linear regression techniques

Yotsaphat Kittichotsatsawat^{1,2}✉, Nakorn Tippayawong³ & Korrakot Yaibuathet Tippayawong^{2,4}✉

Crop yield and its prediction are crucial in agricultural production planning. This study investigates and predicts arabica coffee yield in order to match the market demand, using artificial neural networks (ANN) and multiple linear regression (MLR). Data of six variables, including areas, productivity zones, rainfalls, relative humidity, and minimum and maximum temperature, were collected for the recent 180 months between 2004 and 2018. The predicted yield of the cherry coffee crop continuously increases each year. From the dataset, it was found that the prediction accuracy of the R^2 and RMSE from ANN was 0.9524 and 0.0784 tons, respectively. The ANN model showed potential in determining the cherry coffee yields.

Coffee is one of the most popular global beverages, and Thailand is among the world's top 25 coffee exporters, with over 30,000 tons per year¹. It can generate enormous earnings for the country's economy and farmers. It was found that high-standard coffee can get an asking price a few times more than standard products^{2,3}. However, coffee products have been encountering several problems, including unqualified beans, climate change and environmental effects, a decrease in the nutrition level in soils, pests and insects, a lack of farming knowledge management, as well as the unavailability of raw materials⁴⁻⁷. While the main focus should be on the quality of the products to satisfy customer requirements, the lack of management might be the leading cause of these problems and might affect the efficiency of the accuracy rate. This could also subsequently lead to the failure of marketing prices in the future^{8,9}. At present, the world's coffee consumption has grown by an average of over 2% a year in the past decade¹⁰. The Thai coffee industry has shown similar growth between 2016 to 2020. Thailand's average demand for coffee beans is almost 79,000 tons per year¹¹. Accurate prediction of coffee will improve farmers' or entrepreneurs' ability to plan their annual coffee production and reduce the risk of inconsistent demand, supply, and lack of trade opportunities.

Arabica coffee (*Coffea arabica* L.), one of the main coffee species, has been considered a suitable plant to promote as an opium substitute to stop shifting cultivation^{12,13}. However, Arabica is also one of the most popular species of coffee that customers look for because it features a particular characteristic flavor that differs from other types of coffee¹⁴. For this reason, farmers and entrepreneurs are likely to concentrate on Arabica coffee to meet the demand from customers. Hence, it is necessary to predict each year's yields in order to manage the insufficient supply and the overwhelming demands of customers so as to reduce the loss of trade opportunities. Forecasting applications have recently been used to predict demand, supply management, and resources to reduce mistakes in terms of overproduction and operating costs^{15,16}. Prediction in agriculture is mainly made in terms of rainfall estimates, area, production or productivity, and price^{17,18}. Thus, these factors should be used to predict the productivity and yield of agricultural products, possibly with a machine learning model, as shown in Fig. 1.

Machine learning (ML) is carried out to learn and make decisions without setting rules and conditions¹⁹ because it can learn through reading human-provided information and processing data. This data information can be applied to make the desired decisions²⁰. The decision-making of ML has various formats, for example, predicting and classifying a new database²¹, identifying and clustering with different data^{22,23}, detecting incorrect

¹Graduate Program in Industrial Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand. ²Excellence Centre in Logistics and Supply Chain Management, Chiang Mai University, Chiang Mai 50200, Thailand. ³Department of Mechanical Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand. ⁴Department of Industrial Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200, Thailand. ✉email: yotsaphat_jant@cmu.ac.th; korrakot@eng.cmu.ac.th

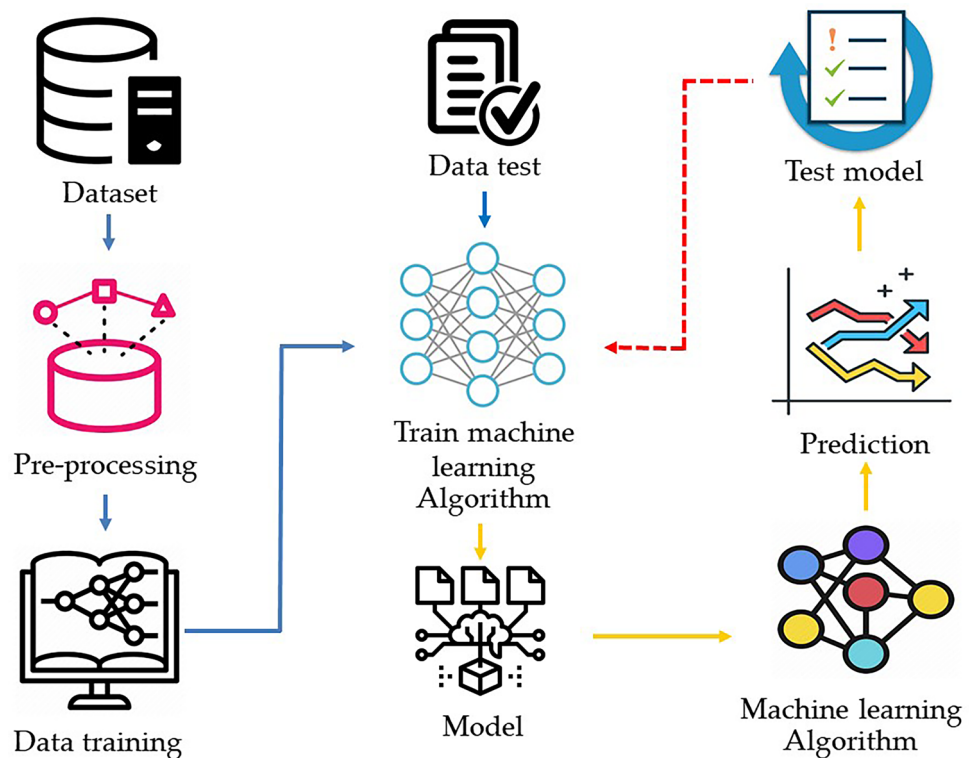


Figure 1. Machine learning predictive models.

data information²⁴, and introducing new data that are expected to interest humans by learning from previous interests^{25,26}. In our problem, the multilayer perceptron was adopted for prediction. Neurons in the network are generated in a layer. Then, the output of one layer supports the input into the next layer and conceivably others. The model parameter of this study can be optimized by using the feed-forward backpropagation training algorithm. The ML model has great potential for application to coffee production prediction. Therefore, this research uses artificial neural networks (ANN) in coffee cherry prediction because of its ability to learn the patterns of complex systems. Moreover, it can predict future productivity from the current database more effectively than the traditional statistical method.

ML is critical to studying programming and computer design, such as robotic systems, visual systems, or computer vision^{27–29}. A computer algorithm generates the machine learning program, and this program is then able to calculate and evaluate the condition, and the software will generate the decision-making. Then, the data is transferred to the dataset in order to generate the model. After that, the completed program is applied to determine work usability³⁰. ML can be categorized into two groups; (i) supervised learning, which comprises regression and classification groups, and (ii) unsupervised learning, which comprises clustering and non-clustering groups³¹. ML uses a variety of algorithms, for example, linear regression, logistics regression, support vector machines, decision trees, random forest, boosting, convolutional neural network, gradient-boosted trees, and ANN^{32–34}. In this work, we are interested in ANN.

The ANN model is now considered a vital data-modeling tool³⁵. It is a calculation model that activates the functional structures of biological neural networks. A neuron is one part of the process running unit that acquires inputs and consequently operates these inputs to achieve outputs. The construction of the ANN includes many architectures, consisting of single-layer perceptrons, multilayer perceptrons, recurrent ANN, and self-organization mapping. Each of these is appropriate for different problems³⁶.

The significant benefit of the ANN technique is the capability of modeling multiple outputs concurrently³⁷. Improvement of the ANN model requires data generation for training or testing, selecting the optimal configuration, and validating the model on an independent data set³⁸. The optimal configuration was found to be 1 and 2 hidden layers, 2–10 neurons, and 1000–10,000 learning runs³⁹. ANN is connected via a structure of single-layer perceptrons and consists of input, output, and hidden layers, which reconstruct the input through the output layer. ANN analysis was used to analyze the input and output data using the MATLAB Neural Network Toolbox. It includes input, hidden, and output layers, which are the central elements of each ANN structure⁴⁰. The input layer will obtain the input data and normalize from datasets through neurons. Then, all of the data will be transmitted to the hidden layer⁴¹. Every neuron of the subsequent layer(s) will calculate the output data, transmitting a linear combination of all the neurons in the previous layers⁴².

ANN is a technique that uses data mining. A mathematical model is used to connect the information processing and computation and simulate the function of neural networks⁴³. Neurons in computer systems include input and output, with neuron weights that can indicate the input and output weights. Then, the threshold value

is the decision for transferring the output data through the neurons. After the collaboration work is finished, the results are shown in numerical calculations format³⁶.

Previously, ANN and MLR have been employed in many various fields, such as irrigation groundwater quality parameters⁴⁴, forecasting PM levels⁴⁵, predicting SO₂ concentration⁴⁶, slope stability⁴⁷, intrinsic solubility of generic drugs⁴⁸, electrical conductivity⁴⁹, mosquito abundances in urban areas⁵⁰, estimating monthly rainfall distribution in Thailand⁵¹, prediction of human skin permeability⁵², etc. Presently, ANNs are used in medicine and health care^{53,54}, medical image denoising⁵⁵, face detection and recognition⁵⁶, road and traffic signaling⁵⁷, road sign detection and recognition^{58,59}, and in modeling conventional and solar earth⁶⁰. ANN and MLR are also currently used in agriculture, for instance, predicting organic potato yield³⁷ and the yield in the hullless barley⁶¹.

It can also be used in the agricultural field to detect and classify plant diseases, crops, and pests, especially in coffee bean species and coffee cherry⁶². Ahmad et al.⁶³ applied ANN to develop an automatic coffee sorting system based on image processing. Moreover, ANN was used to evaluate the degree of coffee roasting, predicting the sensory quality of a roasted coffee-flavored sterilized drink and recognizing the degree of coffee roasting⁶⁴. ANN could be applied to predict the amount of agricultural production by using previous datasets in order to obtain correct information through a machine learning-based algorithm for annual coffee production⁶⁵. In addition, ANN can predict not only agricultural products but also enter many types of data and extract knowledge outside of the primary dataset. Torkashvand et al.⁶⁶ applied the ANN and MLR model to evaluate and improve the accuracy rate in predicting kiwifruit firmness. Dhyani⁶⁷ used the ANN and MLR model to measure the rainfall to anticipate the crop in the following year. Furthermore, Etminan et al.⁶⁸ utilized an ANN model to identify the best drought tolerance indices in agronomy and plant breeding by providing this valuable model in other biological contexts.

According to the literature, ANN is one of the tools that is suitable for predicting agricultural products because it can simulate the physical behavior of a complex system from the input data. The advantages of ANN are as follows; (i) it can give a more accurate prediction than general methods, (ii) the model can still be used even if there are errors in the actual data, (iii) any conflicting results can still be implemented got either are continuous, (iv) the analysis of results is fast after the learning process⁶⁹. On the other hand, there are disadvantages of ANN; (i) a long training time can be necessary in order to find the best weight to learn, (ii) it is not easy to choose suitable parameters (the best topology), (iii) the learned function can be difficult to understand⁷⁰.

MLR is used to study the collaboration of the relationships of two or more variables that affect another value. The regression includes various linear regressions and logistic regression algorithms that control the learning model to predict the input data into the program⁷¹. Its algorithms are more complicated but are also used in agricultural production, for example, ordinary least squares regression, multivariate adaptive regression splines, multi-linear regression, and locally estimated scatterplot smoothing⁷².

From the literature review, it was noticed that ANN and MLR are not yet utilized widely in agricultural production. Their applications are relatively novel in the study of agricultural production. To the authors' knowledge, only a few works have deployed ANN and MLR to predict agricultural crop yield. So far, no work has been reported on applying machine learning to coffee cultivation. This study investigates and analyzes the prediction of cherry coffee and quantity control during plantation before processing the roasted coffee beans in Chiang Rai in northern Thailand, as shown in Fig. 2.

Results

Descriptive statistics. In this study, we provided the predicted yield of arabica coffee for 15 years from 2004 to 2018. Coffee is seasonally produced and available 6 months a year. Off-season, there is no yield. The dataset used in the study consists of factors related to crop yield, including the area, productivity zone, monthly rainfall, monthly RH, and monthly temperature. The data source is from the northern part of Thailand, which has been collected for 180 months. A summary of the statistical characteristics of the dataset is presented in Table 1. It was shown that the productivity zone had a positive skewness distribution of 0.413, and the temperature had a negative skewness distribution of -0.467 . The other variables were slightly skewed. The following subsection shows the results of the modeling. Productivity, also known as crop yield, is the variable to be predicted using the MLR. To apply MLR modeling, the correlation between pairs of input variables was examined to decide which variables to include in the model. The results are shown in Table 2.

In the MLR model, it is common to include the input variables with a high linear correlation with the response variable but to exclude ones that are highly correlated among themselves because of the multicollinearity problem. Table 2 shows that the area and productivity zone variables are highly positively correlated, while the variables Tmax and area and Tmax and RH are moderately positively and negatively correlated, respectively. These results might lead to a multicollinearity problem; however, in this study, we included all variables in the MLR, as the dataset is one hundred eighty observations and six variables. The results are shown in "Prediction results".

Prediction results. The MLR, in which the equation was derived from the full model of six variables, is used to predict crop yield. The MLR equation, R^2 , and the Durbin–Watson test are shown in Table 3. From Table 3, the model that includes six input datasets yields a high value of R^2 , which indicates that the coffee features substantially explain the change in the crop yield (92.35%). Meanwhile, the Durbin–Watson test is 1.56, which reveals a suitable autocorrelation for the predictive errors. After data analysis, it was found that the R^2 is 0.9235, RMSE is 0.0784, and MSE is 0.0061 ton, which is suitable for the model (see Fig. 3).

From Fig. 3, when the MLR model tested the input data and crop yield, it can be seen that the linear relationship was preserved between the variable input datasets and the crop yield. Thus, farmers and entrepreneurs could use this data set to predict cherry coffee productivity in the next period.

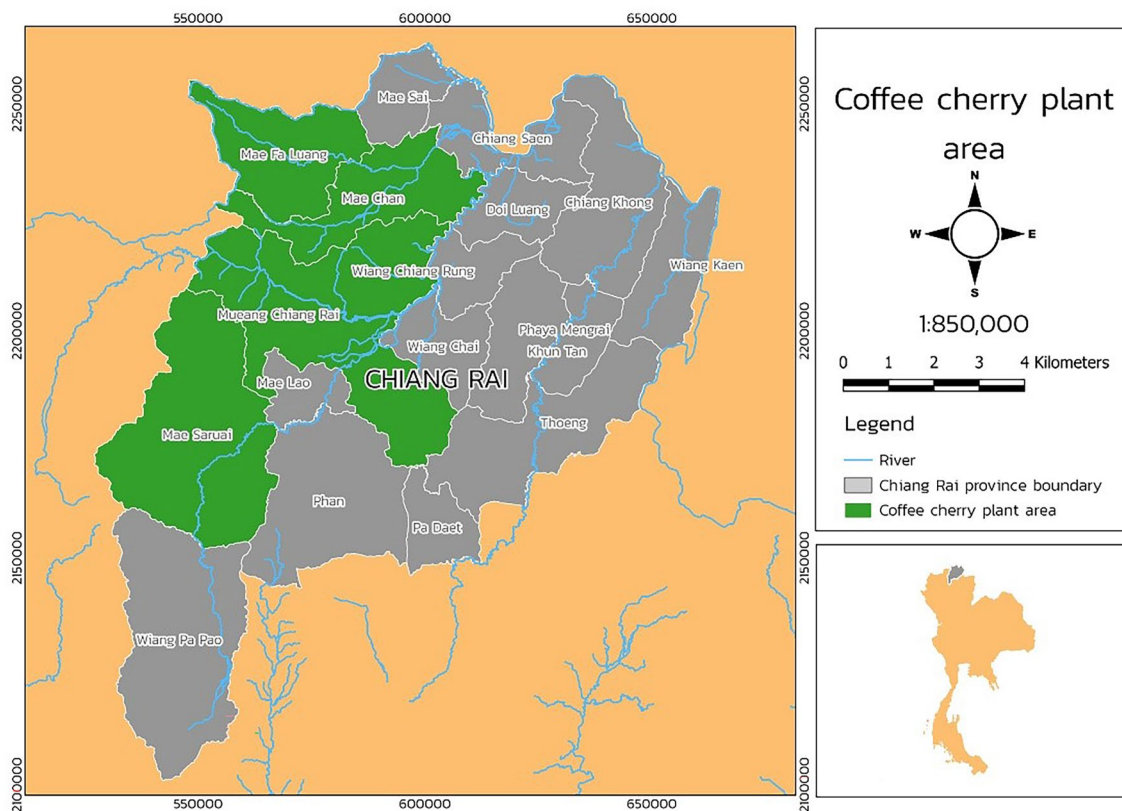


Figure 2. Coffee plantation area in Chiang Rai, Thailand. Figure has been generated by Mr. Yotsaphat Kittichotsatsawat, the first author of this manuscript using the Qgis program. Version 3.8.1. URL. <https://www.qgis.org/en/site/index.html>.

Variable	Maximum	Minimum	Mean	Median	SD	Skewness	Kurtosis
Area (hectare)	42,215	20,158	34,003	31,186.5	11,028.5	0.292	0.336
Productivity zone (hectare)	37,710	14,513	23,921	26,111.5	11,598.5	0.413	- 0.979
Monthly total rainfall (mm)	509.8	0	1830	254.9	254.9	- 0.055	- 0.739
Relative Humidity (%RH)	82.0	55.0	76.0	68.5	13.5	- 0.168	- 0.788
Temperature monthly (°C)	40.0	5.10	25.5	22.6	17.5	- 0.467	0.258
Productivity (tons)	4794.27	365.066	2505	2579.7	2214.6	- 0.243	- 1.106

Table 1. Summary of crop yield statistics.

Variable	Area	Productivity zone	Rainfall	RH	Tmax	Tmin
Area (hectare)	1.0000					
Productivity zone (hectare)	0.8119	1.0000				
Monthly total rainfall (mm)	0.0100	0.0677	1.0000			
Relative Humidity (%RH)	- 0.3366	- 0.1641	0.2527	1.0000		
Tmax monthly (°C)	0.6278	0.5140	- 0.1882	- 0.6542	1.0000	
Tmin monthly (°C)	0.3233	0.5962	0.1790	0.3080	0.3192	1.0000

Table 2. Correlation matrix of the input dataset.

The ANN analysis was performed based on the previous 180 months’ input dataset. In all ANN analyses, the input and output had eight neurons and one layer. The number of neurons in the hidden layer was defined using trial and error to find the best ANN configuration for predicting the crop yield of cherry coffee. The ANN configuration determined to be optimal should minimize the MSE and RMSE and optimize R². The ANN configuration from the network topology started from one and then consecutively increased.

R ²	RMSE	MSE	Durbin-Watson	Equation (coefficient)	P-value
0.9235	0.0784	0.0061	1.56	Crop yield = 0.2850 + 0.0402 Area + 0.3128 Productivity Zone - 0.0486 Rainfall - 0.4040 RH - 0.6144 Tmax + 1.1065 Tmin	0.001

Table 3. Results of MLR analysis to estimate crop yield of cherry coffee productivity.

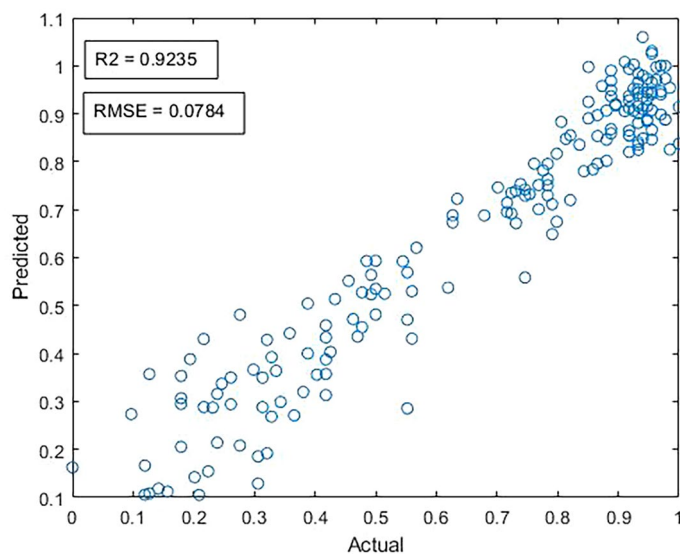


Figure 3. The R² and RMSE (ton) of the scatter plot on the MLR model between the actual and predicted crop yield.

Hidden layer (HI) number	Processing elements (PEs)	MSE	R
1	1	34,115.10	0.777
1	2	316,012.50	0.757
1	3	287,720.50	0.757
1	4	9384.50	0.746
1	5	2102.34	0.910
1	6	241,512.50	0.684
1	7	23,690.64	0.922
1	8	55,592.50	0.776
1	9	409,160.00	0.859
1	10	84,228.49	0.888
2	1	72,170.25	0.857
2	2	9383.68	0.684
2	3	50,069.96	0.859
2	4	66,932.68	0.871
2	5	87,831.30	0.541
2	6	53,020.86	0.851
2	7	5150.58	0.921
2	8	1027.99	0.975
2	9	103,884.14	0.855
2	10	113,635.85	0.761

Table 4. The R and MSE (ton) performances of various ANN configurations using once trained for the dataset.

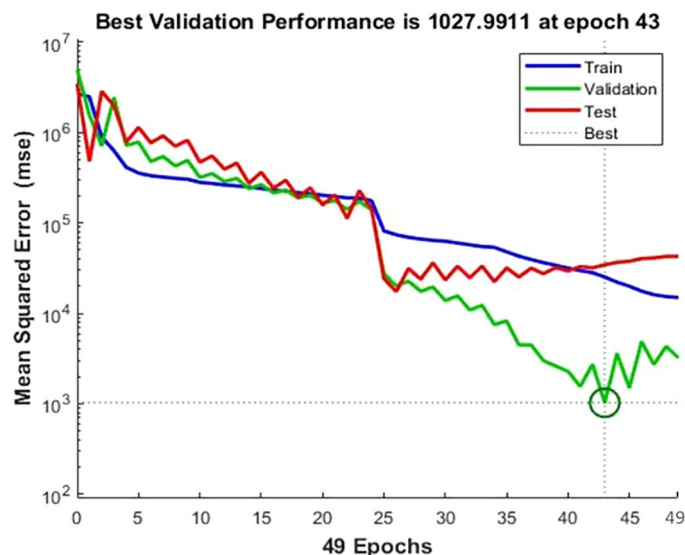


Figure 4. Mean squared error (MSE) of cherry coffee validation performance.

From Table 4, the network properties will set the treatments of the (i) type of network, (ii) input and target data, (iii) training and adaptation learning function, (iv) performance function, (v) layer number, (vi) number of neurons, and (vii) transfer function. Feed-forward backpropagation was used for setting the network type. The number of hidden layers is fixed as two, numbered H1 and H2, and the number of neurons will set the number of processing elements. Then, these data will be trained in order to obtain the best ANN performance. After training the ANN model, the best of the various ANN configurations is with the second hidden layer and the eighth processing element. Meanwhile, the MSE and R^2 are 1027.99 and 0.9524 tons, respectively, as shown in Figs. 4 and 5.

Figure 4 shows the MSE from the random data generated, which included the training, validation, and testing data. The learning rate and momentum were constant at 1000, 0.5, and 0.1, respectively. The results show that the learning cycles were 49 epochs, and the best validation performance was found at 43 epochs with the MSE at 1027.99. Table 5 shows the outcome of the coefficient (R^2) of the training and testing data phases. The neural network confirms the precise tools with perfect prediction accuracy in the responses of genotypes for coffee crop yield based on the measured indices. Thus, the ANN model was demonstrated to provide a perfect agreement between the exact and approximate coffee productivity. However, the prediction potential of productivity by ANN (0.9524) appeared to show higher accuracy than MLR (0.9235). MLR and ANN were compared through machine learning techniques using historical areas, productivity zones, rainfalls, relative humidity, and minimum and maximum temperature for the recent 180 months.

According to the study, MLR and ANN have focused on practical interpolation methods. RMSE and MSE cross-validation were accomplished to assess the efficiency of each month (168 months). It was also estimated through the average of RMSE and MSE. However, the cross-validation of MLR showed that the average values of RMSE and MSE are 0.0618 and 0.0038, respectively. Nonetheless, in comparison to the full model of coffee prediction for 180 months, RMSE was 0.0784, and MSE was 0.0061 tons, respectively. The results showed that both of them were not significantly different.

Discussion and conclusion

The crop yield of cherry coffee varies because of several factors such as the cultivated area, amount of rainfall, temperature, and RH. These conditions affect the cherry coffee yield from each month's crop. Prediction is necessary to forecast the productivity accurately and to satisfy customer demand. Previously, farmers and entrepreneurs lack knowledge regarding the future cherry coffee yield, as well as which factors can change the cherry coffee crop yield each month. For this reason, production planning to match the supply to the demand is rather difficult. However, if farmers and entrepreneurs have sufficient data and can utilize it using machine learning models, they may realize more accurate estimate of the crop yield.

In this work, MLR and ANN were used to analyze and predict the coffee crop yield using data from 2004 to 2018, in which input data for 132 months was used for training, 24 months for validation, and 24 months for testing. Calculation of the cherry coffee crop yield required MATLAB programming. The MLR equation was derived from the full model consisting of six variables. The data shows that the area (X_1) ranged between 42,216 and 20,158 acres with an average of 34,003 acres, playing an essential role in cherry coffee production. The available area is a crucial component for coffee tree planting because it determines the farmer's cherry coffee productivity. The productivity zone (X_2) shows data ranging between 37,710 and 14,513 acres, with an average of 23,921 acres. This is also essential because it indicates how much the farmer can earn from the cherry coffee yield⁷³. The rainfall data (X_3) is from 0 to 509.8, but the average rainfall is 1830 mm per month. The amount

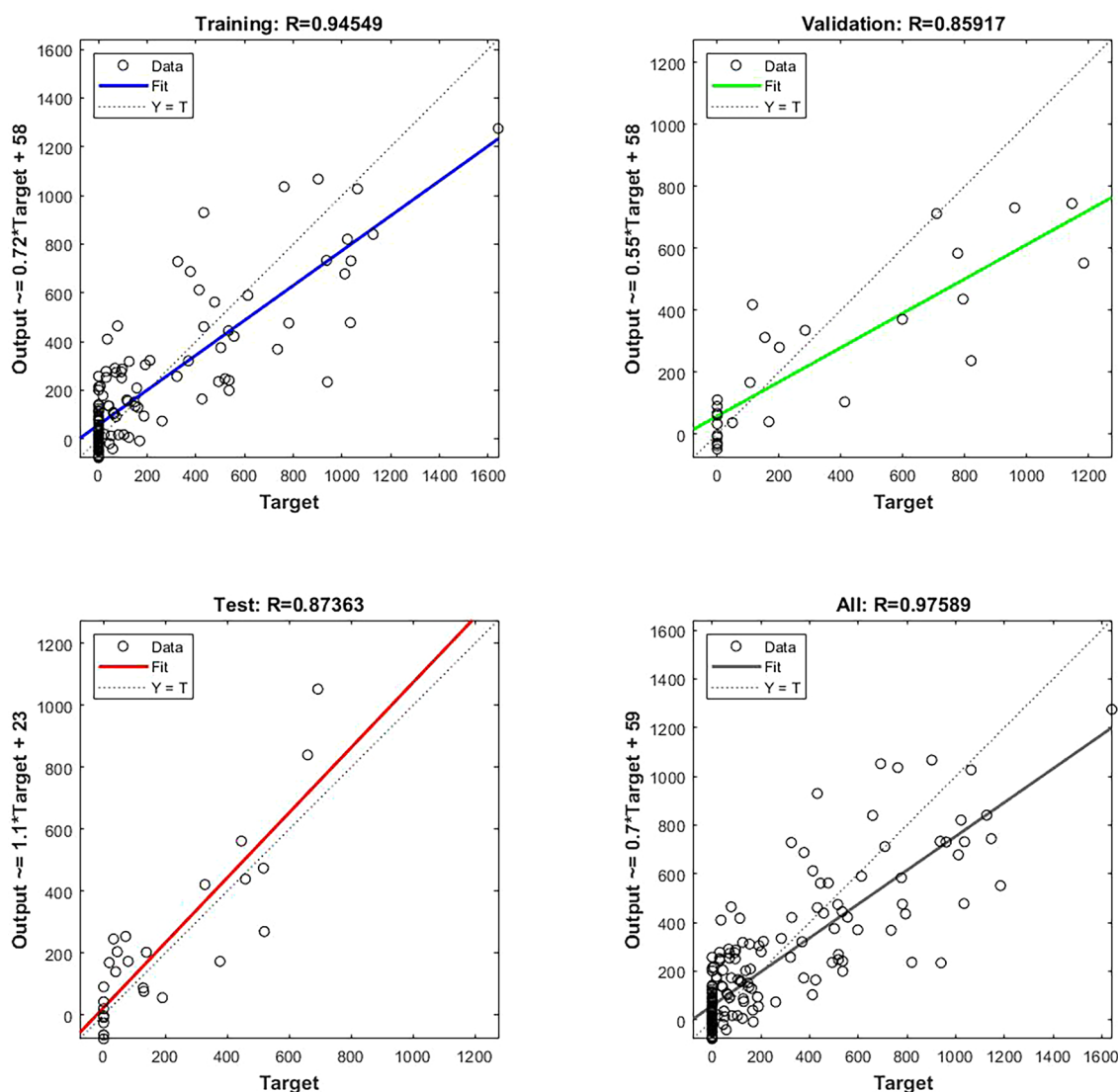


Figure 5. Prediction of the coffee cherry using ANN (MSE training and predicted values of cherry coffee by multilayer perceptron neural network versus data value, training and validation phases) (ton).

Model	R ²	RMSE	MSE
ANN	0.9524	0.0784	1027.99

Table 5. Error and R² (RSQ), RMSE (ton), and MSE (ton) in testing datasets of ANN model.

of RH (X_4) is between 55.0 and 82.0%RH, and the mean is 76.0%RH which is suitable for coffee cultivation⁷⁴. Moreover, the temperature (X_5 and X_6) is a factor that could affect coffee production and is shown in the data to be between 5.10 and 40.0 °C with an average of 25.5 °C⁷⁵.

The correlation was checked between independent variables. As in Table 2, there are some high correlations, such as those between the minimum temperature and productivity, the area and productivity zone, and the productivity zone and productivity, with values of 0.9175, 0.8119, and 0.7775, respectively. For this reason, it was necessary to put all the variables into the model or may represent them with VIF < 10 values (variance inflation factor). The assumption of MLR was adjusted and estimated from $\hat{Y} = 0.2850 + 0.0402X_1 + 0.3128X_2 - 0.0486X_3 - 0.4040X_4 - 0.6144X_5 + 1.1065X_6$ with 0.9235 for the standard error of estimation. The R² and Durbin–Watson values confirmed the appropriateness of the MLR equation, as shown in Table 3. R² and Durbin–Watson values were 0.9235 and 1.56, respectively which are suitable for this MLR model. Meanwhile, R² in Table 3 also indicates that the coffee variable substantially explains the changes in the crop yield, while the Durbin–Watson test reveals a low autocorrelation for the predictive errors.

The importance of each input variable on the yield was also analyzed based on the MLR mathematical expression. The equation coefficients showed that Tmin monthly variable (X_6) was the most important (1.1065). The

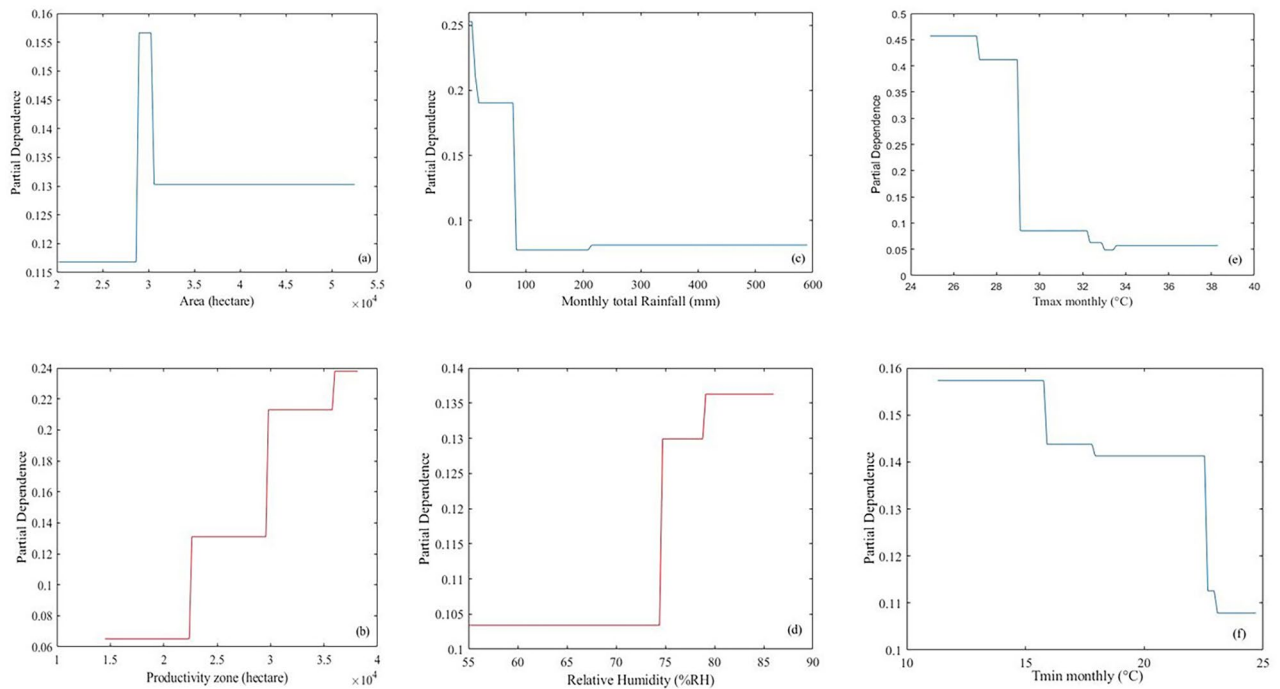


Figure 6. One-way PDPs of input variables on prediction of crop yields.

second-most important factor was Tmax monthly (X_5) with a coefficient of -0.6144 . The relative humidity (X_4), the productivity zone (X_2), and monthly total rainfall (X_3) were inferior, with values of -0.4040 , 0.3128 , and -0.0486 , respectively. Lastly, the area (X_1) was shown to be the least essential variable (0.0402).

Similarly, for the ANN model, the relative importance of each input variable was evaluated based on one-way partial dependence plots (PDPs), as shown in Fig. 6. It can be seen that the monthly Tmax variable (X_5) was the most important, in which the PDP value varied from 0.45 to 0.05 . Similarly, the second most influential variable was monthly total rainfall (X_3), with PDP values of 0.25 to 0.05 . The third variable was relative humidity (X_4) of 0.138 to 0.103 . Additionally, the last three variable groups were productivity zone (X_2), area (X_1), and Tmin monthly (X_6), showing the PDP values varied from 0.240 to 0.07 , 0.158 to 0.118 , and 0.160 to 0.110 , respectively^{76,77}. In Fig. 6e, when the Tmax monthly variable increased, cheery coffee's productivity was lowered, especially when it was higher than 29 °C. In other words, if the temperature was less than 29 °C, the productivity of cheery coffee was improved. From Fig. 6b, if we have more productivity zones, productivity will increase accordingly. From Fig. 6d, the productivity would be high if the cheery coffee experienced high relative humidity. On the contrary, if the amount of rain were too large, it would negatively affect productivity. Based on the PDPs in Fig. 6c, the cheery coffee should rain at less than 100 mm per month, leading to a high crop yield.

Likewise, the ANN method can obtain the input dataset from the available data. Then, the data was imported from the MATLAB workspace, and a variable was selected. A network or data was created to random the hidden layer until the performances of the various ANN configurations were known. The best performance is shown in Table 5, with values of 0.9524 and 0.0784 ton for R^2 and RMSE, respectively. The results of both models indicate that the cheery coffee crop yield can be accurately predicted if the farmers or entrepreneurs are able to utilize these data.

In comparing MLR and ANN performances with other agricultural-related predictions, the R^2 of the MLR was better than the ANN in some cases due to its limited number of input datasets. This was so for Bayat et al. work⁷⁸ on predicting tree survival and mortality in the Hyrcanian forest. The six input variables were used for 9 years of prediction, including the wind velocity and direction, diameter at breast height, basal area of larger trees, diameter, topographic wetness index, and temperature. The R^2 of ANN and MLR were 0.6790 and 0.9200 , respectively. Ustaoglu et al.⁷⁹ used MLR and ANN for temperature forecasts through daily mean, maximum, and minimum temperature in time series. The R^2 of ANN was 0.8590 , lower than that of MLR (0.9025).

The R^2 of ANN was generally found to be better than that of MLR when many datasets were used, for example, in predicting rainfall^{75,80}, and maize yield through the temperature and precipitation data⁸¹, monthly pan evaporation and soil temperatures^{82,83}, the relative humidity of climate in Turkey⁸⁴, and crop prediction^{85,86}, as shown in Table 6. For the cheery coffee crop yield by MLR and ANN in this work, the R^2 were 0.9235 and 0.9524 , respectively, while RMSE was 0.0784 ton, showing the suitability for crop yield prediction.

MLR and ANN were effective in forecasting the crop yield of cheery coffee. These methods can be utilized from upstream to downstream in coffee planning to respond to customer demand. Moreover, they can reduce the risk of insufficient production in the high season and improve future efficiency and effectiveness for industrial and agricultural products.

Concerning limitations, this research considered only six input datasets, which were the area, amount of rainfall, temperature, and RH that affected the efficiency of increased crop yield. More related data input such

No	References	Subject	Number of data	Variables	R ²		RMSE
					ANN	MLR	
1	Bayat et al. ⁷⁸	Predicting tree survival and mortality in the Hyrcanian forest of Iran	9 years	Wind velocity and direction, Diameter at breast height, Basal area of larger trees, Diameter, Topographic wetness index, Temperature	0.6790	0.9200	0.9850
2	Ustaoglu et al. ⁷⁹	Forecast of daily mean, maximum and minimum temperature time series	14 years	Temperature	0.8590	0.9025	0.9400
3	Ilaboya and Igbiniedion ⁷⁵	Prediction of monthly maximum rainfall in Benin city, Nigeria	34 years	Rainfall	0.9999	0.1755	0.7800
4	El-Shafie1 et al. ⁸⁰	Rainfall forecasting	10 years	Rainfall	0.8110	0.4160	0.5168–0.6800
5	Matsumura et al. ⁸¹	Maize yield forecasting	43 months	Temperature and precipitation data	0.8230–0.9540	0.7600–0.9350	0.8770
6	Patle et al. ⁸²	Monthly pan evaporation modeling	84 months	Temperature, Relative humidity, solar sunshine hours	0.8900	0.8800	0.3300
7	Kisi et al. ⁸³	Modeling soil temperatures at different depths	300 months	Temperature, Solar radiation, Wind speed, Relative humidity, and Soil temperature	0.9564–0.9940	0.8118–0.9761	–
8	Yasar et al. ⁸⁴	Estimation of relative humidity	144 months	Relative humidity	0.9624–0.9965	0.7379–0.9489	–
9	Li et al. ⁸⁵	Prediction of crop and soybean	1059 countries	Area	0.5329–0.9409	0.5041–0.9216	0.7501–0.9299
10	Han et al. ⁸⁶	Crop evapotranspiration prediction	24 months	Atmospheric pressure, Wind speed, Temperature, Relative humidity, Sunshine hours, and precipitation	0.8700	0.7900	0.9857

Table 6. Comparison of MLR and ANN models for other agricultural production.

as farm location, irrigation water depth applied, precipitation, solar radiation, potential evapotranspiration, soil moisture, land cultivated, pH, nitrogen, phosphorus, potassium, air pressure, wind speed, and climate variables should also be considered. The crop yield prediction could be improved further^{87–89}.

Once the crop yield and amount of cherry coffee are accurately predicted for future works, farmers or entrepreneurs can prepare their manufacturing resources for later stages of production planning accordingly. These may include (i) color sorters to classify the color shade of cherry coffee for higher added value and (ii) dryer machines to reduce the moisture content in the cherry coffee.

The ANN algorithm has been demonstrated to be effective in making predictions of production rates/yields from industries of coffee plantings. However, the dataset used may be imbalanced. It is essential to note that, in identifying the optimal granularity and refining the imbalanced dataset, combining granular computing and ML techniques is of great interest^{90,91}. In general, the predictions of developed ML models are based on knowledge learning from the complex multi-dimensional relationships between features and target(s)⁹². When an imbalanced, small dataset is employed to train and develop an ML model, the developed model would have a poor ability to make predictions. An innovative approach of combining granular computing with ML or deep learning may be considered further, in which the dataset's quality and efficiency could be improved via clustering examples in the best granularity, sampling, and refining into a more balanced example set, such as those deployed in services planning under social manufacturing context and bi-level model for risk classification of respiratory diseases^{90–94}. Additionally, reinforcement learning techniques may be considered to predict agricultural production. The learning process happens along the way with six main components of (i) agent, (ii) action, (iii) environment, (iv) state, (v) policy, and (vi) reward. It was successful for order acceptance decision of mass-individualized printed circuit board manufacturing⁹⁵. In this research, only six main input features were considered. Therefore, reinforcement learning with more complex datasets may be considered for cherry coffee and other agricultural industries in the near future.

Materials and methods

Research framework. This study focuses on predicting cherry coffee. The scope of this study covers four steps from plantation until harvesting (see Fig. 7).

Data collection and treatment. The data from the Climate Department included rainfall, RH, and minimum and maximum temperature. The Agricultural Economics Office and Meteorological Department provided the area and productivity zone data. The input dataset consisted of the area, rainfall, temperature, and (RH), while the output consisted of the crop yield of cherry coffee. The input and output data were normalized within the range 0–1 using

$$N = \frac{(X - X_{min})}{(X_{max} - X_{min})}, \quad (1)$$

where N is the normalized data; X is the measured value; X_{min} and X_{max} are the minimum and maximum values. The ANN performance must be estimated by means of a validation dataset. The selected ANN was tested, and its performance was compared using the coefficient of determination (R^2), the mean squared error (MSE), and the percentage error.

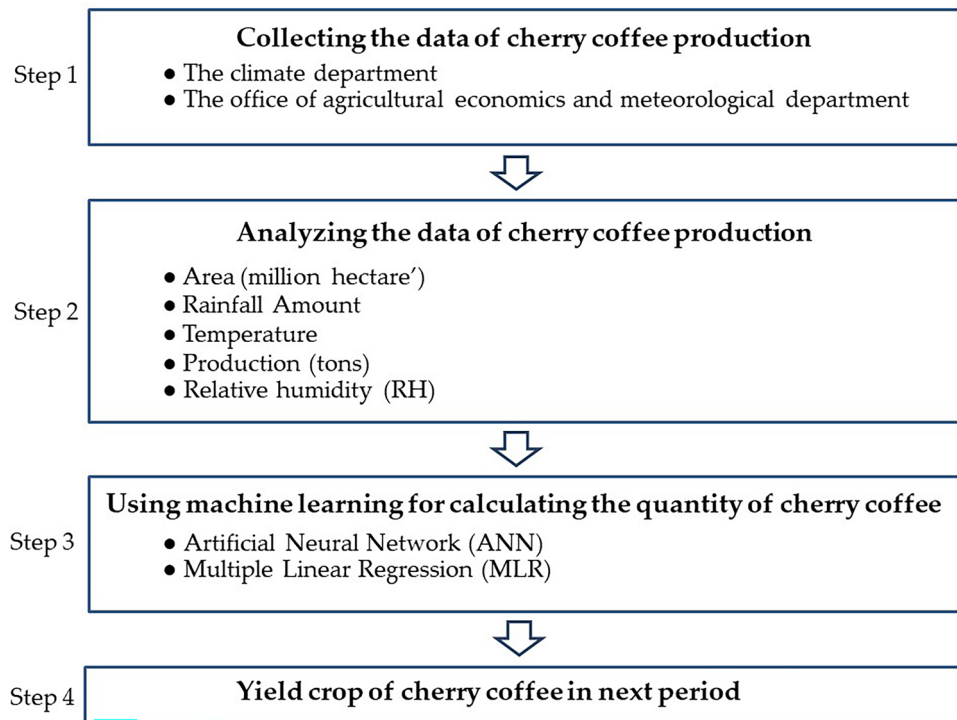


Figure 7. Research framework for cherry coffee prediction.

Statistical and MLR model. MLR is one of the supervised learning groups that can generate the model for constructing a relationship between two or more interpretive variables (independent) and response variables (dependent) via appropriate linear equations into the observed data. The model for MLR is:

$$y_i = b_0 + b_1X_{i,1} + b_2X_{i,2} + \dots + b_kX_{i,k} + e_i, \quad (2)$$

where Y_i is the dependent variable; b_0 is a constant (intercept); $X_{i,k}$ is an independent variable; b_k is the vector of regression coefficients (slope); and e_i is random measured errors.

Data set, training and selection of optimal onfiguration. The variables data were randomly split into three groups to prevent overfitting in the training step because a high number of neurons existed in the hidden layer. Training, validation, and testing of models were based on 70%, 15%, and 15% of the data, respectively⁶⁸. In the meantime, during neural network training, the algorithm provided the minimum or maximum performance via the shortest path to yield the network's size. At the same time, backpropagation was performed through the training set in order to update the rule to seek and find the minimum mean square error over the training set.

The best point of generalization form validation was obtained from a training stop when the error started to increase until the final increase point. Next, the networks were trained with the testing set, which was also compared with the output dataset with the desired output, and the model weight was fixed⁹⁶. After the training epoch, the mean square error of the network running of the validation set was calculated. Then, the neurons will apply a specific linear function in the hidden layer and a specific linear function in a hidden layer to collect the linear combination and bias. Finally, the output data will give the predicted model.

The validated model was evaluated based on the observed and predicted coffee crop yield difference. The 180 months of data on coffee production were randomly evaluated through a k-fold cross-validation technique but evenly partitioned into 168 months of the dataset. The dataset will be further randomly evaluated through three rounds of MLR and ANN (ten processing elements and two hidden layers) training. Each dataset was individually used as validation, while the enduring ones were treated as training datasets. From the k-fold cross-validation, RMSE and MSE were evaluated to determine the models' performance:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n [E(x_i) - M(x_i)]^2}, \quad (3)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\text{Error})^2, \quad (4)$$

where n is the sample size of the testing dataset, while $E(x_i)$ and $M(x_i)$ are interpolated and observed values, respectively.

Prediction of crop yield. In this model, the crop yield was considered to assess indices through the input and output data. A partial dependence plot (PDP) was performed to evaluate each variable as an independent variable during the neural network process^{76,77}. The relative importance of the parameters for each index was computed by dividing the value of its importance by the highest importance value. Consequently, the model with the largest R^2 and the smallest RMSE was considered the best.

The ratios of the correlation between standard errors and mean values were calculated by evaluating the accuracy in terms of repeatability and similarity for yield-based with the maximal indices up to 1000 bootstrap samples similar to the coefficient of variation (CV). A low CV indicates a selection index that is more accurate⁶⁵. All implementations were accomplished using the neural network toolbox in the MATLAB package. Therefore, in the present study, MATLAB version R2019b, with Windows 10 and a 64-bit machine, was used to run the MLR and ANN models.

Data availability

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Received: 30 March 2022; Accepted: 16 August 2022

Published online: 25 August 2022

References

- Lewin, B., Giovannucci, D. & Varangis, P. Coffee markets: New paradigms in global supply and demand. *World Bank Agric. Rural Dev. Discuss. Paper* <https://doi.org/10.2139/ssrn.996111> (2004).
- Panhuyens, S. & Pierrot, J. Coffee barometer 2014. *Hivos, IUCN Nederland, Oxfam, Novib, Solidaridad, WWF.* (2014).
- Torok, A., Mizik, T., Jambor, A. J. I. J. O. E. & Issues, F. The competitiveness of global coffee trade. *Int. J. Econ. Financ.* **8**, 1–6. <https://doi.org/10.32479/ijefi.6692> (2018).
- Degaga, J. & Alamerie, K. J. E. Supply and performance of coffee markets in Gololcha district of Oromia region, Ethiopia. *Ekonomika Poljoprivrede.* **67**, 797–816. <https://doi.org/10.5937/ekopolj2003797d> (2020).
- Genanaw, T. & Lameneu, W. Indigenous knowledge management framework for coffee production in Ethiopia. *Ethiop. e-J. Res. Innov. Foresight.* **6**, 53–61 (2019).
- Duarte, A., Uribe, J. C., Sarache, W. & Calderón, A. J. E. Economic, environmental, and social assessment of bioethanol production using multiple coffee crop residues. *Energy* **216**, 119170. <https://doi.org/10.1016/j.energy.2020.119170> (2021).
- Effendi, D. & Rismaya, M. Design and development of coffee production information system to support coffee production productivity in farmers group. *IOP Conf. Ser. Mater. Sci. Eng.* **879**, 012058. <https://doi.org/10.1088/1757-899x/879/1/012058> (2020).
- Baloi, D. & Price, A. D. J. I. J. O. P. M. Modelling global risk factors affecting construction cost performance. *Int. J. Proj. Manag.* **21**, 261–269. [https://doi.org/10.1016/s0263-7863\(02\)00017-0](https://doi.org/10.1016/s0263-7863(02)00017-0) (2003).
- Kittichotsawat, Y., Jangkrajarn, V. & Tippayawong, K. Y. J. S. Enhancing coffee supply chain towards sustainable growth with big data and modern agricultural technologies. *Sustainability.* **13**, 4593. <https://doi.org/10.3390/su13084593> (2021).
- International Coffee Organization. Developing a sustainable coffee economy (2020). http://www.ico.org/sustaindev_e.asp. Accessed 30 March 2022.
- Chuqian, W. A study on the situation and development of the coffee industry in Thailand (2018).
- Trébuil, G., Ekasingh, B. & Ekasingh, M. Agricultural commercialisation, diversification, and conservation of renewable resources in northern Thailand highlands. *Moussons.* **9–10**, 131–155. <https://doi.org/10.4000/moussons.2005> (2006).
- Sutthi, C. Highland agriculture: From better to worse. *Hill Tribes Today.* 107–142. (1989).
- Romani, S., Cevoli, C., Fabbri, A., Alessandrini, L. & Dalla Rosa, M. Evaluation of coffee roasting degree by using electronic nose and artificial neural network for off-line quality control. *J. Food Sci.* **77**, C960–C965. <https://doi.org/10.1016/j.foodres.2020.109667> (2012).
- Moon, M. A. Demand and supply integration: The key to world-class demand forecasting. *Walter de Gruyter* <https://doi.org/10.1515/9781501506024> (2018).
- Doucoure, B., Agbossou, K. & Cardenas, A. Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data. *Renew. Energy.* **92**, 202–211. <https://doi.org/10.1016/j.renene.2016.02.003> (2016).
- Ashokkumar, K., Chowdary, D. D. & Sree, C. D. Data analysis and prediction on cloud computing for enhancing productivity in agriculture. *IOP Conf. Ser. Mater. Sci. Eng.* **590**, 012014. <https://doi.org/10.1088/1757-899x/590/1/012014> (2019).
- Rajeswari, S. & Suthendran, K. Developing an agricultural product price prediction model using HADT algorithm. *Int. J. Eng. Adv. Technol.* **9**, 569–575. <https://doi.org/10.35940/ijeat.a1126.12915419> (2019).
- Janiesch, C., Zschech, P. & Heinrich, K. Machine learning and deep learning. *Electron. Mark.* **3**, 685–695. <https://doi.org/10.1007/s12525-021-00475-2> (2021).
- Tripathy, A., Agrawal, A. & Rath, S. K. Classification of sentiment reviews using n-gram machine learning approach. *Expert Syst. Appl.* **57**, 117–126. <https://doi.org/10.1016/j.eswa.2016.03.028> (2016).
- Sun, S., Cao, Z., Zhu, H. & Zhao, J. A survey of optimization methods from a machine learning perspective. *IEEE Trans. Cybern.* **8**, 3668–3681. <https://doi.org/10.1109/tycb.2019.2950779> (2019).
- Valletta, J. J., Torney, C., Kings, M., Thornton, A. & Madden, J. Applications of machine learning in animal behaviour studies. *Anim. Behav.* **124**, 203–220. <https://doi.org/10.1016/j.anbehav.2016.12.005> (2017).
- Dogan, A. & Birant, D. Machine learning and data mining in manufacturing. *Expert Syst. Appl.* **166**, 114060. <https://doi.org/10.1016/j.eswa.2020.114060> (2021).
- Wiens, J. & Shenoy, E. S. Machine learning for healthcare: On the verge of a major shift in healthcare epidemiology. *Clin. Infect. Dis.* **66**, 149–153. <https://doi.org/10.1093/cid/cix731> (2018).
- Howard, J. Artificial intelligence: Implications for the future of work. *Am. J. Ind. Med.* **11**, 917–926. <https://doi.org/10.1002/ajim.23037> (2019).
- Lheureux, A., Grolinger, K., Elyamany, H. F. & Capretz, M. A. Machine learning with big data: Challenges and approaches. *IEEE Access.* **5**, 7776–7797. <https://doi.org/10.1109/access.2017.2696365> (2017).
- Kim, D. et al. Review of machine learning methods in soft robotics. *PLoS ONE* **16**, e0246102. <https://doi.org/10.1371/journal.pone.0246102> (2021).

28. Abiodun, O. I. *et al.* State-of-the-art in artificial neural network applications: A survey. *Heliyon*. **11**, e00938. <https://doi.org/10.1016/j.heliyon.2018.e00938> (2018).
29. Lindsay, G. W. Attention in psychology, neuroscience, and machine learning. *Front. Comput. Neurosci.* **14**, 29. <https://doi.org/10.3389/fncom.2020.00029> (2020).
30. Dalal, S. R. *et al.* In model-based testing in practice. In *Proceedings of the 21st International Conference on Software Engineering—ICSE '99*. 285–294. <https://doi.org/10.1145/302405.302640> (1999).
31. Brownlee, J. Supervised and unsupervised machine learning algorithms. *Mach. Learn. Mastery*. **3** (2016). <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>. Accessed 30 March 2022.
32. Çınar, Z. M. *et al.* Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. *Sustainability*. **19**, 8211. <https://doi.org/10.3390/su12198211> (2020).
33. Onsree, T. & Tippayawong, N. Machine learning application to predict yields of solid products from biomass torrefaction. *Renew. Energy*. **167**, 425–432. <https://doi.org/10.1016/j.renene.2020.11.099> (2021).
34. Onsree, T., Tippayawong, N., Phithakkitnukoon, S. & Lauterbach, J. Interpretable machine-learning model with a collaborative game approach to predict yields and higher heating value of torrefied biomass. *Energy* **249**, 123676. <https://doi.org/10.1016/j.energy.2022.123676> (2022).
35. Agatonovic-Kustrin, S. & Beresford, R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J. Pharm. Biomed. Anal.* **5**, 717–727. [https://doi.org/10.1016/s0731-7085\(99\)00272-1](https://doi.org/10.1016/s0731-7085(99)00272-1) (2000).
36. Singh, Y. & Chauhan, A. S. Neural networks in data mining. *J. Theor. Appl. Inf. Technol.* **1**, 1–6 (2009).
37. Abrougui, K. *et al.* Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). *Soil Tillage Res.* **190**, 202–208. <https://doi.org/10.1016/j.still.2019.01.011> (2019).
38. May, R. J., Dandy, G. C., Maier, H. R. & Nixon, J. B. Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems. *Environ. Model. Softw.* **10–11**, 1289–1299. <https://doi.org/10.1016/j.envsoft.2008.03.008> (2008).
39. Anderson, D. & McNeill, G. Artificial neural networks technology. *Kaman Sci. Corp.* **6**, 1–83 (1992).
40. Zain, A. M., Haron, H. & Sharif, S. Prediction of surface roughness in the end milling machining using artificial neural network. *Expert Syst. Appl.* **37**, 1755–1768. <https://doi.org/10.1016/j.eswa.2009.07.033> (2010).
41. Togun, N. K. & Baysec, S. Prediction of torque and specific fuel consumption of a gasoline engine by using artificial neural networks. *Appl. Energy*. **87**, 349–355. <https://doi.org/10.1016/j.apenergy.2009.08.016> (2010).
42. Yosinski, J., Clune, J., Nguyen, A., Fuchs, T. & Lipson, H. Understanding neural networks through deep visualization. *Explain AI. Interpret. Explain. Vis. Deep Learn.* https://doi.org/10.1007/978-3-030-28954-6_4 (2015).
43. Krenker, A., Bešter, J. & Kos, A. Introduction to the artificial neural networks. *Artif. Neural Netw. Methodol. Adv. Biomed. Appl* <https://doi.org/10.5772/15751> (2011).
44. Kouadri, S., Pande, C. B., Panneerselvam, B., Moharir, K. N. & Elbeltagi, A. Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models. *Environ. Sci. Pollut. Res.* **29**, 21067–21091. <https://doi.org/10.1007/s11356-021-17084-3> (2022).
45. Ceylan, Z. & Bulkan, S. Forecasting PM10 levels using ANN and MLR: A case study for Sakarya City. *Glob. Nest J.* **20**, 281–290. <https://doi.org/10.30955/gnj.002522> (2018).
46. Shams, S. R., Jahani, A., Kalantary, S., Moenaddini, M. & Khorasani, N. The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO2 concentration. *Urban Clim.* **37**, 100837. <https://doi.org/10.1016/j.uclim.2021.100837> (2021).
47. Chakraborty, A. & Goswami, D. Prediction of slope stability using multiple linear regression (MLR) and artificial neural network (ANN). *Arab. J. Geosci.* **10**, 1–11. <https://doi.org/10.1007/s12517-017-3167-x> (2017).
48. Louis, B., Agrawal, V. K. & Khadikar, P. V. Prediction of intrinsic solubility of generic drugs using MLR, ANN and SVM analyses. *Eur. J. Med. Chem.* **45**, 4018–4025. <https://doi.org/10.1016/j.ejmech.2010.05.059> (2010).
49. Ekemen Keskin, T., Özler, E., Şander, E., Düğenci, M. & Ahmed, M. Y. Prediction of electrical conductivity using ANN and MLR: A case study from Turkey. *Acta Geophys.* **68**, 811–820. <https://doi.org/10.1007/s11600-020-00424-1> (2020).
50. Lee, K. Y., Chung, N. & Hwang, S. Application of an artificial neural network (ANN) model for predicting mosquito abundances in urban areas. *Ecol. Inform.* **36**, 172–180. <https://doi.org/10.1016/j.ecoinf.2015.08.011> (2016).
51. Chutsagulprom, N., Chaisee, K., Wongsajjai, B., Inkeaw, P. & Oonariya, C. Spatial interpolation methods for estimating monthly rainfall distribution in Thailand. *Theor. Appl. Climatol.* **148**, 1–12. <https://doi.org/10.21203/rs.3.rs-568778/v1> (2022).
52. Chen, L. J., Lian, G. P. & Han, L. J. Prediction of human skin permeability using artificial neural network (ANN) modeling. *Acta Pharmacol. Sin.* **28**, 591–600. <https://doi.org/10.1111/j.1745-7254.2007.00528.x> (2007).
53. Soroushmehr, S. R. & Najarian, K. Transforming big data into computational models for personalized medicine and health care. *Dialogues Clin. Neurosci.* <https://doi.org/10.31887/dcms.2016.18.3/ssoroushmehr> (2022).
54. Cervantes-Sanchez, F., Cruz-Aceves, I., Hernandez-Aguirre, A., Hernandez-Gonzalez, M. A. & Solorio-Meza, S. E. Automatic segmentation of coronary arteries in X-ray angiograms using multiscale analysis and artificial neural networks. *Appl. Sci.* **24**, 5507. <https://doi.org/10.3390/app9245507> (2019).
55. Priya, D. K., Sam, B. B., Lavanya, S. & Sajin, A. P. A survey on medical image denoising using optimisation technique and classification. *Int. Conf. Inf. Commun. Embed. Syst.* <https://doi.org/10.1109/icipes.2017.8070729> (2017).
56. Patil, J. V. & Bailke, P. Real time facial expression recognition using RealSense camera and ANN. *Int. Conf. Invent. Comput. Technol.* **2**, 1–6. <https://doi.org/10.1109/inventive.2016.7824820> (2016).
57. Islam, K. T. & Raj, R. G. Real-time (vision-based) road sign recognition using an artificial neural network. *Sensors*. **17**, 853. <https://doi.org/10.3390/s17040853> (2017).
58. Lahmyed, R., Ansari, M. E. & Kerkaou, Z. J. S. C. Automatic road sign detection and recognition based on neural network. *Soft Comput.* **26**, 1–22. <https://doi.org/10.1007/s00500-021-06726-w> (2022).
59. Kryvinska, N., Poniszewska-Maranda, A. & Gregus, M. An approach towards service system building for road traffic signs detection and recognition. *Proc. Comput. Sci.* **141**, 64–71. <https://doi.org/10.1016/j.procs.2018.10.150> (2018).
60. Chauhan, R., Dumka, P. & Mishra, D. R. Modelling conventional and solar earth still by using the LM algorithm-based artificial neural network. *Int. J. Ambient Energy*. **43**, 1389–1396. <https://doi.org/10.1080/01430750.2019.1707113> (2022).
61. Zaeefzadeh, M., Jalili, A., Khayatnezhad, M., Gholamin, R. & Mokhtari, T. Comparison of multiple linear regressions (MLR) and artificial neural network (ANN) in predicting the yield using its components in the hullless barley. *Adv. Environ. Biol.* **5**, 109–114 (2011).
62. Anita, S. In classification Cherry's Coffee using k-Nearest Neighbor (KNN) and Artificial Neural Network (ANN). In *2020 International Conference on Information Technology Systems and Innovation (ICITSI)*. 117–122 (IEEE). <https://doi.org/10.1109/icitsi.2020.9264925> (2020).
63. Ahmad, A. S. *et al.* A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renew. Sustain. Energy Rev.* **33**, 102–109. <https://doi.org/10.1016/j.rser.2014.01.069> (2014).
64. Leme, D. S., da Silva, S. A., Barbosa, B. H. G., Borém, F. M. & Pereira, R. G. F. A. Recognition of coffee roasting degree using a computer vision system. *Comput. Electron. Agric.* **156**, 312–317. <https://doi.org/10.1016/j.compag.2018.11.029> (2019).
65. Van Klompenburg, T., Kassahun, A. & Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* **177**, 105709. <https://doi.org/10.1016/j.compag.2020.105709> (2020).

66. Torkashvand, A. M., Ahmadi, A. & Nikravesh, N. L. Prediction of kiwifruit firmness using fruit mineral nutrient concentration by artificial neural network (ANN) and multiple linear regressions (MLR). *J. Integr. Agric.* **7**, 1634–1644. [https://doi.org/10.1016/s2095-3119\(16\)61546-0](https://doi.org/10.1016/s2095-3119(16)61546-0) (2017).
67. Dhyani, S. *Predicting Rainfall for Agriculture in India Using Regression* (Dublin Business School, 2020). <https://esource.dbs.ie/handle/10788/4230>.
68. Etmian, A. *et al.* Determining the best drought tolerance indices using artificial neural network (ANN): Insight into application of intelligent agriculture in agronomy and plant breeding. *Cereal Res. Commun.* **47**, 170–181. <https://doi.org/10.1556/0806.46.2018.057> (2019). Accessed 30 March 2022.
69. Afan, H. A. *et al.* ANN based sediment prediction model utilizing different input scenarios. *Water Resour. Manag.* **4**, 1231–1245. <https://doi.org/10.1007/s11269-014-0870-1> (2015).
70. Bouktif, S., Fiaz, A., Ouni, A. & Serhani, M. A. Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies* **7**, 1636. <https://doi.org/10.3390/en11071636> (2018).
71. Landwehr, N., Hall, M. & Frank, E. Logistic model trees. *Mach. Learn. Mastery.* **1**, 161–205. <https://doi.org/10.1007/s10994-005-0466-3> (2005).
72. Wilkinson, L. Statistical methods in psychology journals: Guidelines and explanations. *Am. Psychol.* **8**, 594. <https://doi.org/10.1037/0003-066x.54.8.594> (1999).
73. Kath, J., Byrareddy, V. M., Mushtaq, S., Craparo, A. & Porcel, M. Temperature and rainfall impacts on robusta coffee bean characteristics. *Clim. Risk Manag.* **32**, 100281. <https://doi.org/10.1016/j.crm.2021.100281> (2021).
74. Jayakumar, M., Rajavel, M. & Surendran, U. Climate-based statistical regression models for crop yield forecasting of coffee in humid tropical Kerala, India. *Int. J. Biometeorol.* **12**, 1943–1952. <https://doi.org/10.1007/s00484-016-1181-4> (2016).
75. Ilaboya, I. R. & Igbinedion, O. E. Performance of multiple linear regression (MLR) and artificial neural network (ANN) for the prediction of monthly maximum rainfall in Benin City, Nigeria. *Int. J. Eng. Sci. Appl.* **1**, 21–37 (2019).
76. Katongtung, T., Onsee, T. & Tippayawong, N. Machine learning prediction of biocrude yields and higher heating values from hydrothermal liquefaction of wet biomass and wastes. *Bioresour. Technol.* **344**, 126278. <https://doi.org/10.1016/j.biortech.2021.126278> (2022).
77. Wongchai, W., Onsee, T., Sukkam, N., Promwungkwa, A. & Tippayawong, N. Machine learning models for estimating above ground biomass of fast growing trees. *Expert Syst. Appl.* **199**, 117186. <https://doi.org/10.1016/j.eswa.2022.117186> (2022).
78. Bayat, M., Ghorbanpour, M., Zare, R., Jaafari, A. & Pham, B. T. Application of artificial neural networks for predicting tree survival and mortality in the Hyrcanian forest of Iran. *Comput. Electron. Agric.* **164**, 104929. <https://doi.org/10.1016/j.compag.2019.104929> (2019).
79. Ustaoglu, B., Cigizoglu, H. & Karaca, M. Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. *Meteorol. Appl.* **15**, 431–445. <https://doi.org/10.1002/met.83> (2008).
80. El-Shafie, A., Mukhlisin, M., Najah, A. A. & Taha, M. R. Performance of artificial neural network and regression techniques for rainfall-runoff prediction. *Int. J. Phys. Sci.* **6**, 1997–2003. <https://doi.org/10.54302/mausam.v6i2i1.4711> (2011).
81. Matsumura, K., Gaitan, C. F., Sugimoto, K., Cannon, A. J. & Hsieh, W. W. Maize yield forecasting by linear regression and artificial neural networks in Jilin, China. *J. Agric. Sci.* **153**, 399–410. <https://doi.org/10.1017/s0021859614000392> (2015).
82. Patle, G., Chettri, M. & Jhajharia, D. Monthly pan evaporation modelling using multiple linear regression and artificial neural network techniques. *Water Supply* **20**, 800–808. <https://doi.org/10.2166/ws.2019.189> (2020).
83. Kisi, O., Tombul, M. & Kermani, M. Z. Modeling soil temperatures at different depths by using three different neural computing techniques. *Theor. Appl. Climatol.* **121**, 377–387. <https://doi.org/10.1007/s00704-014-1232-x> (2015).
84. Yasar, A., Simsek, E., Bilgili, M., Yucel, A. & Ilhan, I. Estimation of relative humidity based on artificial neural network approach in the Aegean Region of Turkey. *Meteorol. Atmos. Phys.* **115**, 81–87. <https://doi.org/10.1007/s00703-011-0168-2> (2012).
85. Li, Y. *et al.* Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the US. *Field Crops Res.* **234**, 55–65. <https://doi.org/10.1016/j.fcr.2019.02.005> (2019).
86. Han, X. *et al.* Crop evapotranspiration prediction by considering dynamic change of crop coefficient and the precipitation effect in back-propagation neural network model. *J. Hydrol.* **596**, 126104. <https://doi.org/10.1016/j.jhydrol.2021.126104> (2021).
87. Adisa, O. M. *et al.* Application of artificial neural network for predicting maize production in South Africa. *Sustainability.* **4**, 1145. <https://doi.org/10.3390/su11041145> (2019).
88. Gonzalez-Sanchez, A., Frausto-Solis, J. & Ojeda-Bustamante, W. Attribute selection impact on linear and nonlinear regression models for crop yield prediction. *Sci. World J.* **2014**, 1–10. <https://doi.org/10.1155/2014/509429> (2014).
89. Balakrishnan, N. & Muthukumarasamy, G. Crop production-ensemble machine learning model for prediction. *Int. J. Comput. Sci. Softw. Eng.* **7**, 148 (2016).
90. Leng, J. *et al.* A loosely-coupled deep reinforcement learning approach for order acceptance decision of mass-individualized printed circuit board manufacturing in industry 4.0. *J. Clean. Prod.* **280**, 124405. <https://doi.org/10.1016/j.jclepro.2020.124405> (2021).
91. Leng, J., Chen, Q., Mao, N. & Jiang, P. Combining granular computing technique with deep learning for service planning under social manufacturing contexts. *Knowl. Based Syst.* **143**, 295–306. <https://doi.org/10.1016/j.knsys.2017.07.023> (2018).
92. Phromphithak, S., Onsee, T. & Tippayawong, N. Machine learning prediction of cellulose-rich materials from biomass pretreatment with ionic liquid solvents. *Bioresour. Technol.* **323**, 124642. <https://doi.org/10.1016/j.biortech.2020.124642> (2021).
93. Leng, J. *et al.* Bi-level artificial intelligence model for risk classification of acute respiratory diseases based on Chinese clinical data. *Appl. Intell.* <https://doi.org/10.1007/s10489-022-03222-y> (2022).
94. Leng, J. & Jiang, P. Granular computing-based development of service process reference models in social manufacturing contexts. *Concurr. Eng.* **25**, 95–107. <https://doi.org/10.1177/1063293x16666312> (2017).
95. Bu, F. & Wang, X. A smart agriculture IoT system based on deep reinforcement learning. *Future Gener. Comput. Syst.* **99**, 500–507. <https://doi.org/10.1016/j.future.2019.04.041> (2019).
96. Hsu, K. L., Gupta, H. V. & Sorooshian, S. Artificial neural network modeling of the rainfall-runoff process. *Water Resour. Res.* **10**, 2517–2530. <https://doi.org/10.1029/95wr01955> (1995).

Acknowledgements

This work was supported by Chiang Mai University. One of the authors (Y.K.) wishes to acknowledge the CMU Graduate School for Research Assistant Scholarship. We also wish to thank the Department of Industrial Engineering, Excellence Center in Logistics and Supply Chain Management (E-LSCM), Chiang Mai University for providing research facilities. This research is part of the project “Industry 4.0 for SMEs” from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant agreement No 734713.

Author contributions

Conceptualization, K.Y.T. and N.T.; methodology, Y.K. and K.Y.T.; validation, Y.K. and K.Y.T.; formal analysis, Y.K. and K.Y.T.; investigation, Y.K.; writing—original draft preparation, Y.K.; writing—review and editing, N.T.;

supervision, K.Y.T.; funding acquisition, K.Y.T. All authors have read and agreed to the published version of the manuscript.

Funding

Chiang Mai University.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.K. or K.Y.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022