

Stochastic Methodology for Coffee Crop Yield Prediction based on Abiotic Factors

Santhosh C S¹ and Umesh K K²

¹Department of Computer Applications, JSS Science and Technology University, Mysuru, Karnataka, India
Email: sancs84@jssstuniv.in

²Department of Information Science & Engineering, JSS Science and Technology University, Mysuru, Karnataka, India
Email: umeshkatte@jssstuniv.in

Abstract—Predicting coffee crop yields improves agricultural decisions, food security, and environmental effect. In agricultural system models, climatic and pedagogical factors affect coffee yield estimates. India produces numerous Arabica and Robusta coffees, two of 103 types. Few countries like Africa, Zimbabwe, Mexico, and Vietnam have studied plantation crop coffee despite its importance. Stochastic machine learning regression models predict coffee output and quantify great abiotic influences using multivariate feature selection technique and feature importance methods which selects the best features based on multivariate statistical tests. Our research used Karnataka's Central Coffee Research Institute (CCRI) coffee research station's 2004–2022 abiotic factors: year, rainfall, temperature, sunlight, RH, vapour, dew points, and yield.

Abiotic predictor variables were used in the present two-model research on elastic net and extra tree models. The features have been grouped into group-1, 2, 3, 4, and 5 based on the multivariate feature selection technique and feature importance methods. Group-1 features had the highest constant of determination for the 70:30 splitting proportion, using Year, Rainfall, Temperature, Sunshine, Relative Humidity, Vapour, and Dew point as predictor variables (R-square = 0.72 and Root Mean Square Error = 76.54 kg per ha) for Elastic net model and coffee yield R-square = 0.53 and Root Mean Square Error = 96.75 kg per ha was the best coefficient of determination for the extra tree regression model with group-3 parameters (Year, Year + Relative Humidity Minimum and Maximum + Rainfall + Minimum and Maximum Temperature). This research found the best coffee production weather parameter, defeating extra tree model.

Keywords— Coffee, Stochastic approach, machine learning (ML), yield prediction, abiotic variables.

I. INTRODUCTION

Coffee stimulates and delights millions worldwide. Coffee is second only to gasoline in world commerce. India grows coffee predominantly in hilly areas like Karnataka, Kerala, Tamil Nadu, and the southern states, as well as in non-traditional locations like Andhra Pradesh, Orissa, West Bengal, Maharashtra, and the north and east [1]. C. Arabica and C. Robusta dominate Karnataka's Chikamagaluru and Coorg coffee production. Agroecological conditions in these places are lowering production, and forecasts are that by 2050, it will drop by 10–20% [2][3].

D R Mehta et. al., (2002) and Lima Deleon Martins et. al., (2018) Rajkot, India, Central Dry Husbandry Research Station, Targhadia district, and Brazilian southeast have provided weekly rainfall data from 1958 to 1996. Additionally, seasonal precipitation patterns, regional and temporal fluctuations in climatic aquatic equilibrium, elevation, and C have been studied for vulnerability Robusta fields [4][5]. Richard G Pearson et. al., (2003) have explored bio climatic envelope model critics' assertions and how they impact modeling methods [6]. Monica Matoso Campanh et. al., (2004) Compared agroforestry with monoculture C. Arabica plants based on rainfall data. Given these facts, researchers have focused on how rainfall affects coffee C. Arabica type for higher yield [7]. Eduardo Delgado Assad et. al., (2004) and Emerson a de Silva et. al., (2005) have evaluated scenarios' effects on C. Arabica coffees agroclimatic zoning, Brazil's principal planting regions [8][9]. Reshma Chengappa et. al., (2012) have examined resource utilization and water energy, the plant's principal energy source for crop production. The research indicated that a Robusta coffee plantation in Coorg uses 11630 more MJ to increase crop yields and gross income of Rs.18, 882 per acre [10]. Giselly Aparecida et. al., (2012) examined Paraná's C. Robusta coffee agroclimatic zoning and confirmed global warming's 100-year trends [11]. Abel Chemura et. al., (2014) used rainfall data to compare monoculture and agroforestry coffee plants like C. Arabica. Based on these data, experts say rainfall affects C. Arabica coffee more for higher output [12]. Veenadhari, S. et. al., (2014) a user-friendly website called "Crop Advi-sor" evaluates how climate affects crop productivity in this study. Climate affects agricultural production, as seen in this software [13]. Wang, N et. al., (2015) this study monitored yields on 254 plots in each of the five major coffee-growing regions: Northern, Eastern, Central, South-western, and Northwest. Boundary line analysis might quantify local yield disparities and assess how production parameters affect coffee output proportionally [14]. Chengappa, P. G et. al., (2016) study reveals climate change issues for India's small-holder-dominated coffee business. Climate variability is assessed using 33 years of meteorological data from Kodagu, India's largest coffee-growing region. Variables should impact coffee production [15]. Natarajan, R et. al., (2016) FCM-DDNHL-GA classifies sugarcane production using FCM, DDNHL, and GA algorithms. FCMs' hybrid learning technique is compared to several common sugarcane production monitoring machine learning algorithms for classification and inference [16]. To estimate crop yields for the Coffee Research Station (CCRI) in Chikamgaluru, Karnataka, India, we explored several machine learning methods that account for abiotic data. A number of specialists believe data mining and AI are ready to create reliable fore-casting models. We recommended and tested machine learning strategies to predict coffee production based on abiotic parameters. It covered data sets and techniques in detail. Re-sults are explained in detail. Coffee production forecasting using six regression models was presented.

II. DATASET AND METHODS

A. Area of Study and Dataset

The time period covered by the data, 2004–2022 (total 6570 Rows/Days), was collected from the Central Coffee Research Institute (CCRI) in the Balehonnur, Karnataka, district of the Karnataka State of India. The dataset for the model employed in this investigation con-sists of abiotic components, totaling 10 input characteristics from seventeen coffee growing blocks in coffee research station, balehonnur. Table-1 depicts the datasets and provides a detailed description of ten abiotic and a yield parameters.

TABLE I: ABIOTIC AND YIELD PARAMETERS

Parameter Name	Description/Ranges
Year	2004-2022
Temperature – Minimum in Degrees	12.4 to 22.8
Temperature Maximum in Degrees	20.6 to 29.8
Sunshine – Minimum in Hours	1 to 6
Sunshine – Maxi mum in Hours	4.5 to 8
Rainfall in Centimetres	186.1cm to 329.5cm
Relative Humidity – Minimum in Percentage (%)	32% to 74%
Relative Humidity – Maximum in Percentage (%)	83% to 100%
Vapour Percentage (VP) in Percentage (%)	16.7% to 25.63%
Dew Point in Degrees	14.8 to 21.0
Yield in Kg/Acre	250kg to 640kg

B. Proposed Methodology

We provide stochastic machine learning approaches in this part to produce and forecast coffee output based on abiotic standards. Figure-1 depicts the recommended method in square form.

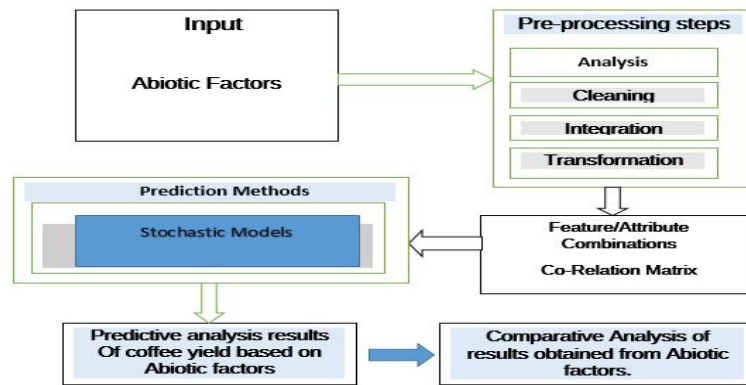


Fig-1: Block diagram of the proposed approach for predicting coffee harvests based on abiotic factors

Pre-processing took just a few stages in our methodology. We first observed few gaps in the datasets. Statisticians utilize arithmetic mean, median, and binning techniques to recreate missing data. Use the CCRI's standard critical values for climate-related parameters [2]. We box plotted abiotic parameters to remove outliers [17]. Box plots showed abiotic characteristics and deleted outliers. Normality tests guarantee a skewness- and outlier-free distribution [18]. Finally, a co-relation matrix shows input variable correlations and feature relevance for coffee yield prediction.

C. Methods

Stochastic machine learning methods were discussed here.

- Elastic Net (ENET) Regression: Elastic-net regression (ENET) was designed to solve complaints that the variable selection in Lasso regression was excessively data-dependent and unpredictable. Mixing Ridge and Lasso regression penalties yielded the best results. Ridge and Lasso algorithms provide a convex mixture of regressions in ENR [19].
- Extra Tree Regression: This new RF model uses "Extra Tree Regression". ETR creates unpruned judgments or regression trees. Bagging and bootstrapping are used in RF regression [20].

The instructions below show how to use the Extra Tree Regression Algorithm for numerical characteristics [20]:

Step-1: Dividing a node (A)

The learning subset A for the neighboring node that has to be divided is the input.

A node split [x xy] or a zero split is the output.

Return 0 if Stop split (A) equals.

Anyhow, from all non-Consistent (in A) Applicant characteristics, choose

B attributes (c1... cb);

Describe the locations of the B divides e1...eb. ei = choose a random number

Splice (A, xi), v j = 1, and then B;

To ensure that Count (d*, A) = maxi=1... B Count (ei, A), return a split e*.

Step-2: Take a haphazard split. (A,x)

Inputs include an attribute x and x subclass A.

Results: x split

Let xAmax and xAmin denote the maximum and insignificant estimates of m in A, respectively.

Illustrate any boundary ac in accordance with [xAmin, xAmax].

Send the split ([x xy]) back.

Step-3: Reverse split (A)

Enter: x subclass Abinary output for xreturn TRUE if |A| xmin;Return TRUE if A's properties are all consistent.

Return TRUE if the result is in accordance with A; FALSE otherwise.

The steps above describe the Extra-Trees splitting approach for numerical features.

D. Predictive Model Development

Two models—ETR and Enet —were created on Windows 11 using Collab and Jupyter on an Intel® core i7 laptop. To predict Robusta and Arabica Yield(Y) in Central Coffee Research Institute (CCRI), coffee research station at Chikamagaluru, the BRR model used patterns embedded in the K (=10) lots of Abiotic factors data matrix from Table-1 and its relationship with the objective variable, Y. Compared to all four models. This paper

uses a cross correlation analysis between X_k (Input Parameters like Group-1, Group-2, Group-3, Group-4, and Group-5) and Y to examine the links between each component and coffee yield data.

For model creation, evaluation/selection, and testing, measured data was independently split into two sets: training () and testing (). Table 2 lays out the descriptive statistics of the Abiotic factors model's input parameters for the training and testing phases. The most affecting features/predictable variables for yield were determined using univariate feature selection technique, which selects the best features based on univariate statistical tests. For Pre-processing of an estimator we have used selectkbest to eliminate all except the K highest scoring characteristics for prediction in univariate analysis.

A total of 1024 distinct combinations of input features and total six models were employed in this study, with varying R² values for each group of feature combinations. Since each of the prescribed abiotic parameters was anticipated to contribute to the predict-ed overall coffee yield, a total of Five optimal BRR models were developed which are la-belled as group-1, group-2 and group-3 taking into account the K feature values as subsets as 1, 2, 3, 4,..., 10. The below table-2 gives the detail information about selected input sub-sets for the prediction.

TABLE II: FIVE OPTIMAL MODELS SUBSETS OF FEATURES SELECTED FOR PREDICTION

Models/Groups	Features Selected	Number of Features
1	Year + Rainfall + Temperature Minimum & Maximum, Sunshine Minimum & Maximum + Relative Humidity Minimum & Maximum + Vapour + Dew Point Vs. Yield,	10
2	Year + Rainfall + Temperature Minimum & Maximum, Sunshine Minimum & Maximum Vs. Yield	06
3	Year + Relative Humidity Minimum and Maximum + Rainfall + Minimum and Maximum Temperature Vs. Yield	06
4	Year + Rainfall + Temperature Minimum & Maximum + Vapour Vs. Yield	05
5	Year + Rainfall + Temperature Minimum & Maximum + Dew Point Vs. Yield	05

The ETR model used an n estimators (number of trees=1000) strategy to build an en-semble of decision trees. It regressed the exploratory and response correlations between the testing phase agro ecology factor data and the coffee yield data, acting as an objective com-parison to the BRR model.

By iteratively trying out ensembles with numbers ranging from 100 to 600, in one-fold increments, we were able to maximize n_estimator, a critical parameter.

Here, we found that 500 trees, with leaf=1 and foots=1, produced the best ETR model.

To further compare Enet models were created for the same set of predictors (group-1, group-2, 3, 4 and group-5). Here n was the number of training/validation/testing data points in each predictor matrix. Based on the objective criterion applied, the R² and RMSE was monitored in each trial and each models were evaluated on Testing Datasets. Table-4 shows the four model design parameters used in this study.

Table-3: Model Design parameters considered based on Group-1, 2, 3, 4 and 5 subsets of features (Table-2) for Extra Tree Regression (ETR) and Elastic Net Regression (Enet) models.

TABLE III

Model	Design Parameters	State of parameters in study
ETR	n_estimators (Number of Trees) Min_sample_split Min_sample_leaf Max_features Random_state	100,200,300,400,500,600,700,800 (Optimal value = 600) 5 , Int, Default = 2 Int, Default=1 Int, Group-1,2,3 subsets , Default =1.0 Int, Default = none.
Enet	Alpha L1_ratio fit_intercept precompute max_iter tol warm_start positive random_state selection	float, default=1.0 float, default=0.5 bool, default=True bool or array-like of shape (n_features, n_features),default=False int, default=1000 float, default=1e-4 bool, default=False bool, default=False int,RandomState instance, default=None {'cyclic', 'random'}, default='cyclic'

E. Model Performance Evaluation Metrics

The current study compared measured yield data with anticipated yield data from the test phase to evaluate six regression models for coffee yield prediction. Analyzed R2, mean absolute error, mean square error, root MSE. The core performance measures in Table-4 forecast production:

TABLE IV: STANDARD PERFORMANCE METRICS

Sl. No	Performance Metrics	Formula
1	R-Squared (R2)	$R^2 \text{ Squared} = 1 - \frac{SSr}{SSm}$ SSr – Squared regression line sum error SSm - Squared mean line sum error.
2	Mean Absolute Error (MAE)	$MAE = \frac{1}{N} \sum_{i=1}^{i=N} (y_i - \hat{y}_i)$ There are N anticipated values..The ith data's real true value is represented by yi. $\hat{y}_i$ is the i-th data's anticipated value.
3	MSE, or Mean Squared Error	$MSE = 1/N \sum (y_i - \hat{y}_i)^2$
4	RMSE, or Root Mean Square Error	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{i=N} (y_i - \hat{y}_i)^2}$

III. RESULTS

During testing, scatterplots (fig. 2) were utilized to visually analyze the degree of agreement between actual yield and projected yield data to simplify all characteristics to a single scale without altering the stochastic models' range of values. In this study, we employed ordinary scalar fine tuning. Additionally, we have included five separate sets of parameters when analyzing and building the model: Group-1 parameters: Year + Rainfall + Temperature Minimum & Maximum, Sunshine Minimum & Maximum + Relative Humidity Minimum & Maximum + Vapour + Dew Point Vs. Yield, Group-2 parameters: Year + Rainfall + Temperature Minimum & Maximum, Sunshine Minimum & Maximum Vs. Yield, Group-3 parameters: Year + Relative Humidity Minimum and Maximum + Rainfall + Minimum and Maximum Temperature Vs. Yield, Group-4 parameters: Year + Rainfall + Temperature Minimum & Maximum + Vapour Vs. Yield, Group-5 parameters: Year + Rainfall + Temperature Minimum & Maximum + Dew Point Vs. Yield. With the use of scatterplots and Figure-2 displays performance measurements R-Square (R²) and RMS Error (RMSE), the actual and expected yield, and model error rate.

Elastic Net Regression: We tested elastic net regression's consistency compared to other probabilistic models with a 10Kfold increase. Table-7 shows overall split consistency, with notable variances among parameter groups.

Table-5: Elastic Net Regression was used to forecast coffee yield with variable abiotic factors for five different groups, and the outcomes of the testing phase assessment for a different split ratios and performance indicators for MAE, MSE, and RMSE were examined in terms of R². The ideal values are shown in boldface.

TABLE V

Group-1 Elastic Net Model: Year + Rainfall + Temp Min Max + Sunshine Min Max + Relative Humidity Min Max + Vapour + Dew Point vs. Yield				
Quantity Shared	R-Square	Mean Absolute Error kg per ha	Mean Square Error kg per ha	Root Mean Square Error kg per ha
90:10	0.71	91.91	8454.73	91.95
80:20	0.78	67.68	4989.60	70.64
70:30	0.72	71.09	5858.16	76.54
60:40	0.48	84.36	8365.08	91.46
Elastic Net Model Constructed Using Group-2 : Year + Rainfall + Temp Min Max + Sunshine Min Max Vs. Yield				
Quantity Shared	R-Square	Mean Absolute Error kg per ha	Mean Square Error kg per ha	Root Mean Square Error kg per ha
90:10	0.65	104.07	10852.35	104.17
80:20	0.73	68.89	6050.81	77.79
70:30	0.64	73.47	7038.19	83.89
60:40	0.44	87.21	8919.88	94.45
Elastic Net Model Constructed Using Group-3 : Year + Rainfall + Temp Min Max + Relative Humidity Min Max Vs. Yield				

Quantity Shared	R-Square	Mean Absolute Error kg per ha	Mean Square Error kg per ha	Root Mean Square Error kg per ha
90:10	0.73	91.59	8396.96	91.63
80:20	0.78	67.26	4914.64	70.10
70:30	0.69	70.67	5772.07	75.97
60:40	0.48	83.97	8287.57	91.04
Elastic Net Model Constructed Using Group-4 : Year + Rainfall + Temp Min Max + Vapour Vs. Yield				
Quantity Shared	R-Square	Mean Absolute Error kg per ha	Mean Square Error kg per ha	Root Mean Square Error kg per ha
90:10	0.64	104.80	11005.62	104.91
80:20	0.73	69.85	6228.52	78.92
70:30	0.64	74.06	7161.53	84.63
60:40	0.43	87.72	9030.26	95.03
Elastic Net Model Constructed Using Group-5 : Year + Rainfall + Temp Min Max + Dew Point Vs. Yield				
Quantity Shared	R-Square	Mean Absolute Error kg per ha	Mean Square Error kg per ha	Root Mean Square Error kg per ha
90:10	0.65	104.07	10852.35	104.17
80:20	0.73	68.89	6050.81	77.79
70:30	0.65	73.41	7017.06	83.77
60:40	0.44	87.19	8907.59	94.38

Visually comparing testing procedures. Group 1's elastic net regression model (All parameters vs. Yield) had the highest constant of determination for the 70:30 splitting proportion, using Year, Rainfall, Temperature, Sunshine, Relative Humidity, Vapour, and Dew point as predictor variables (R-square = 0.72 kg per ha and Root Mean Square Error = 76.54 kg per ha) (fig Performance measurements showed that the Elastic Net Regression model failed for group-2, 3, 4, and 5 parameters. Other inputs helped the Elastic Net regression model forecast coffee yield, indicating that parameter data can't predict. Table-5 compares testing efficiency for several splitting models using R², MAE, MSE, and RMSE.

Extra Tree Regression: Extremely Randomized Trees (Extra Trees) is a regression like Random Forest. Feature splits are random. Mixing the test and train datasets, the Extra tree regression model was created using 100 DTs from diverse informative categories.

Table-6: Extra Tree Regression model was used to forecast coffee yield with variable abiotic factors for five different groups, and the outcomes of the testing phase assessment for a different split ratios we looked at the ratios of several performance indicators including R², MAE, MSE, and RMSE. Boldface indicates the optimal values.

TABLE VI

Group-1 Extra Tree Model : All Parameters(Year + Rainfall + Temp Min Max + Sunshine Min Max + Relative Humidity Min Max + Vapour + Dew Point) vs. Yield				
Quantity Shared	R-Square	Mean Absolute Error kg per ha	Mean Square Error kg per ha	Root Mean Square Error kg per ha
90:10	0.78	59.79	6860.77	82.83
80:20	0.46	94.86	12221.13	110.55
70:30	0.47	93.29	10520.19	102.57
60:40	0.34	91.67	10518.58	102.56
Extra Tree Model Constructed Using Group-2 : Year + Rainfall + Temp Min Max + Sunshine Min Max vs. Yield				
Quantity Shared	R-Square	Mean Absolute Error kg per ha	Mean Square Error kg per ha	Root Mean Square Error kg per ha
90:10	0.79	57.19	6275.86	79.22
80:20	0.38	103.37	14183.45	119.09
70:30	0.44	88.62	11097.62	105.35
60:40	0.27	95.50	11631.80	107.85
Extra Tree Model Constructed Using Group-3 : Year + Rainfall + Temp Min Max + Relative Humidity Min Max vs. Yield				
Quantity Shared	R-Square	Mean Absolute Error kg per ha	Mean Square Error kg per ha	Root Mean Square Error kg per ha
90:10	0.78	58.80	6635.44	81.46
80:20	0.43	99.66	12998.17	114.01
70:30	0.53	88.25	9360.95	96.75
60:40	0.37	92.00	10020.85	100.10
Extra Tree Model Constructed Using Group-4 : Year + Rainfall + Temp Min Max + Vapour vs. Yield				
Quantity Shared	R-Square	Mean Absolute Error kg per ha	Mean Square Error kg per ha	Root Mean Square Error kg per ha
90:10	0.80	56.42	6107.72	78.15

80:20	0.41	99.94	13339.82	115.50
70:30	0.48	86.75	10198.86	100.99
60:40	0.32	94.32	10801.73	103.93
Extra Tree Model Constructed Using Group-5 : Year + Rainfall + Temp Min Max + Dew Point vs. Yield				
Quantity Shared	R-Square	Mean Absolute Error kg per ha	Mean Square Error kg per ha	Root Mean Square Error kg per ha
90:10	0.80	56.95	6223.36	78.89
80:20	0.44	98.23	12820.74	113.23
70:30	0.51	83.17	9638.42	98.18
60:40	0.32	93.99	10827.54	104.06

The variables in Group 3 are: Season + Average + Extreme Conditions (Temperatures and RH) yields five groupings, with minimum and maximum being most relevant. The additional tree regression model with group-3 parameters (Year, Year + Relative Humidity Minimum and Maximum + Rainfall + Minimum and Maximum Temperature) had the best coefficient of determination (R-square = 0.53 kg per ha and Root Mean Square Error = 96.75 kg per ha). Comparing measured yield to expected yield for 70:30 split ratio showed this. Performance indicators showed the extra tree regression model performed badly for groups 1, 2, 4, and 5 parameters. Using the same inputs and group-3 parameters, the Bayesian Ridge regression model and Lasso regression model fail to extract predictive features from multi-parameter data of coffee yield. Table-6 displays R^2 , MAE, MSE, and RMSE results for several splitting models during testing.

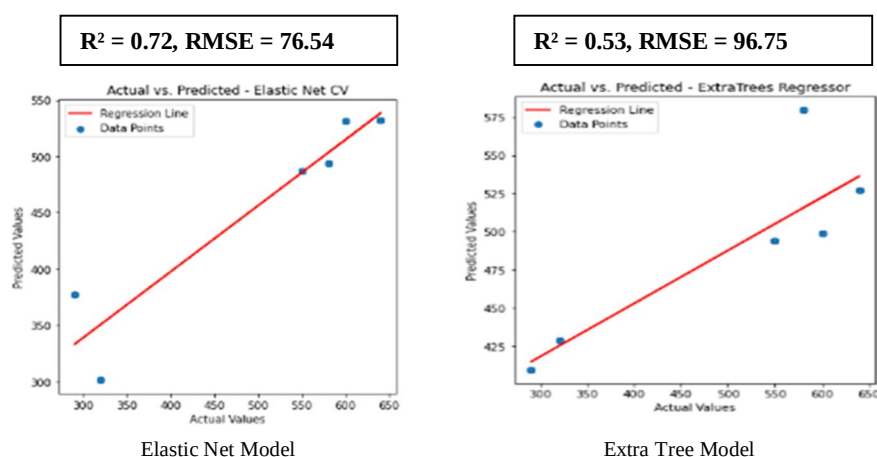


Figure.2: Group 1 & 3 served as input parameters for the following models shown in above scatterplots based on stochastic models based on the elastic net and extra tree algorithms. Above are scatterplots depicting the highest expected and actual coffee yields during the experimentation period (70:30 Splits).

IV. DISCUSSIONS

Coffee output depends on the growing region, rainfall, temperature, sunshine hours, vapor, dew points, and relative humidity. These factors affect each month's coffee crop. Planning production to balance supply and demand is difficult. This research used four stochastic models like Random forest regression, Gradient boosting regression, Elastic net regression and Extra Tree Regressions models to predict coffee crop output using 2004–2022 abiotic factor data. Among the four stochastic models Extra tree model and Elastic net model performed well with an accuracy of $R^2 = 72\%$ and $R^2 = 53\%$ in comparison with Random forest with $R^2 = 52\%$ and Gradient boosting with $R^2 = 46\%$ accuracy. Parameterizing crop development simulation models is difficult, however statistical analytics in a data intelligent system can predict the greatest yielding abiotic elements. Even if non-selected abiotic factors impact coffee growth, our findings may guide future research on smallholder coffee farm decision-making. Other elements that may affect coffee output that were not explored here Pests, pests, soil fertility, and fertilizers matter. These traits may increase coffee production prediction model performance. Abiotic data-based algorithms can predict coffee output and its key components, according to our research. Other studies indicated that soil parameters were the biggest yield confounders. Work done thus far shows that Tables 5, 6 under the model created for group 1,2,3,4 & 5 parameters are valid (fig. 2). R-squared compares model fit to horizontal straight line. Throughout our analysis, we found no instances where the selected model fit the data worse than a horizontal line.

As per the literature survey reported, no work has been presented for coffee yield prediction using abiotic factors. Also, the other works reported considered the datasets belongs to other parts of the globe like Vietnam, Mexico, Uganda, Zimbabwe etc. The comparison of propose work with other models reported is not drawn as there is a change of datasets and regions. So in this work we exclusively considered Indian origin datasets with our own dataset which are collected and prepared from central coffee research institute (CCRI), balehonnur, Karnataka.

This research found that abiotic components value connections above individual meanings. Now that land management, soil, and environmental conditions need various chemical, biological, and physical indicators, studies must concentrate on more than one or two environmental parameters. We divided each parameter's relevance by its greatest value to get its index weight. Therefore, the model with the highest R^2 and lowest RMSE is best.

V. CONCLUSION

Abiotic parameters from the Central Coffee Research Institute Station in Chikamagaluru, Karnataka, are used to compare actual and expected coffee yield. The reliability of stochastic models as a data-driven strategy for examining predictive features in abiotic variables data was tested to improve coffee crop productivity at the Chikamagaluru coffee research station. The goal variable was coffee yield (Y), and the 10 abiotic characteristics were divided into 5 groups to create stochastic models, a unique machine learning technique for complex and ill-defined situations. Bayesian Ridge regression models extracted features between abiotic factors and used a variety of variables to predict coffee yields more accurately than Lasso, Elastic Net, Extra Tree, Random Forest, and Gradient Boost Regression models. This study demonstrated the potential benefits of integrating biophysical crop models with artificial intelligence algorithms in precision agriculture decision-support systems by consciously increasing productivity in smallholder farms using adequately screened abiotic characteristics. Studies of machine learning methods help enhance analytical and prediction models.

ACKNOWLEDGEMENT

We appreciate the aid we received historical coffee records from the Central Coffee Research Institute (CCRI) in Balehonnur, Chikamagaluru District, Karnataka, India, through JSS Science and Technology University, Mysuru in our quest to anticipate coffee harvest yields and also records were crucial in helping us understand the many factors affecting the coffee production.

REFERENCES

- [1] Coffee board research department – 70th annual report 2016-17, Central coffee research institute, coffee research station, Government of India.(2017)
- [2] Coffee guide book-a manual of coffeeCultivation,Central coffee research Institute (ministry of commerce and Industry, Govt. of india). (2014)
- [3] K M Nair K S Anil Kumar - Fertility of coffee growing soils of south india, Central Coffee Research Institute, Coffee Board. (2017)
- [4] D.R. Mehta, A D. Kalola, D.A Saradava and A S. Yusufzai “rainfall variability analysis and its impact on crop productivity- a case study “indian j. agrlc. res., 36 (1): 29 – 33. (2002)
- [5] Lima Deleon Martins, Fernando Coelho Eugenio, Wagner Nunes Rodrigues “Adaptation to Long-Term Rainfall Variability for Robusta Coffee Cultivation in Brazilian Southeast “American Journal of Climate Change, 7, 487-504. (2018).
- [6] Richard G. Pearson and Terence P. Dawson “Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? “Global Ecology & Biogeography, 2003, 12, 361–371. (2003)
- [7] Mônica Matoso Campanhã, Ricardo Henrique Silva Santos, Gilberto Bernardo de Freitas, Hermínia Emília Prieto Martinez, Silvana Lages Ribeiro Garcia and Fernando Luiz Finger “Growth and yield of coffee plants in agroforestry and monoculture systems in Minas Gerais, Brazil “Agroforestry Systems. 63: 75–82. (2004)
- [8] Eduardo Delgado Assad, Hilton Silveira Pinto, Jurandir Zullo Junior e Ana Maria Helminsk Ávila “Climatic changes impact in agroclimatic zoning of coffee in Brazil “Pesq. ag-ropec. bras. Brasília, v.39, n.11, p.1057-1064. (2004).
- [9] Emerson A. de Silva, Paulo Mazzafera, Orivaldo Brunini, Emilio Sakai, “the influence of water management environmental conditions on the chemical composition and beverage quality of coffee beans “Braz. J. Plant Physiology, 17(2):229-238.(2005)
- [10] Reshma Chengappa “Seasonal Scarcity of Water and The Change In Energy Use Pattern In Robusta Coffee Plantation – A Study In Kodagu District Of Karnataka In India” International Journal of Emerging Technology and Advanced Engineering, November, ISSN 2250-2459, Volume 2, Issue 11. (2012)

- [11] Giselly Aparecida Andrade Wilian da Silva Ricce Paulo Henrique Caramori Geovanna Cristina Zaro Cristiane de Conti Medina “ Agroclimatic zoning of robusta coffee in the State of Paraná and impacts of climate change “ *Semina: Ciências Agrárias, Londrina*, v. 33, n. 4, p. 1381-1390, jul./ago. (2012)
- [12] Abel Chemura “The growth response of coffee (*Coffea arabica* L) plants to organic manure, inorganic fertilizers and integrated soil fertility management under different irrigation water supply levels “*International Journal Recycling Organic Waste Agriculture*, May , 3:59. (2014)
- [13] Veenadhari, S., Misra, B., & Singh, C. D... Machine learning approach for forecasting crop yield based on climatic parameters. In *2014 International Conference on Computer Communication and Informatics*, Jan, (pp. 1-5). IEEE. (2014)
- [14] Wang, N., Jassogne, L., van Asten, P. J., Mukasa, D., Wanyama, I., Kagezi, G., & Giller, K. E. , Evaluating coffee yield gaps and important biotic, abiotic, and management factors limiting coffee production in Uganda. *European Journal of Agronomy*, 63, 1-11. (2014)
- [15] Chengappa, P. G., & Devika, C. M., Climate variability concerns for the future of coffee in India: an exploratory study. *International Journal of Environment, Agriculture and Bio-technology*, 1(4), 238602. (2016)
- [16] Natarajan, R., Subramanian, J., & Papa Georgiou, E. I., Hybrid learning of fuzzy cognitive maps for sugarcane yield classification. *Computers and Electronics in Agriculture*, 127, 147-157. (2016)
- [17] Schwartzman, N. C., Owens, M. A., and Adnan, R., A simple more general boxplot method for identifying outliers. *Computational statistics and data analysis*, 47 (1), 165-174. (2006)
- [18] Joanes, D. N., and Gill, C. A., Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1), 183-189. (1998)
- [19] Garcia-Nieto, P. J., Garcia-Gonzalo, E., & Paredes-Sánchez, J. P., Prediction of the critical temperature of a superconductor by using the WOA/MARS, Ridge, Lasso and Elastic-net machine learning techniques. *Neural Computing and Applications*, 2021, 33, 17131-17145. (2021).
- [20] Geurts, P., Ernst, D., & Wehenkel, L., Extremely randomized trees. *Machine learning*, 63, 3-42. (2006).