A Neutrosophic Machine Learning-Based Intelligent Ensemble Model for Sustainable Tea Yield Prediction Under Climatic Variability

Maitraya Dey[1], Pushpita Roy[2], Shubhendu Banerjee[3]*, Amrut Ranjan Jena[4],
Rakesh Naskar[5], Suparna Dasgupta[6], Soumyabrata Saha[7], Sudarshan Nath[8], Bikash Mondal[9]
Department of CSE, Narula Institute of Technology, Kolkata, India[1, 2, 3]
Dr. Sudhir Chandra Sur Institute of Technology and Sports Complex, Kolkata, India[4]
Department of CSE, Gargi Memorial Institute of Technology, Kolkata, India[5]
Department of Information Technology, JIS College of Engineering, Kalyani, Nadia, India[6, 7, 8]
Department of ECS, Narula Institute of Technology, Kolkata, India[9]

Abstract—**For efficient agricultural planning, resource management, and enhancing farmer livelihoods in significant tea-producing regions, tea production prediction is essential. However, climate variability—including temperature, rainfall, humidity, and sunlight duration—has a significant impact on tea output, making precise forecasting difficult. Using meteorological data from 2015 to 2025, this study suggests a hybrid machine learning approach for predicting tea production. Initially, four models are created as separate predictors: Random Forest, XG Boost, Light GBM, and Cat Boost. Three ensemble models are shown to increase prediction accuracy: a Neutrosophic Ensemble Model, a Fuzzy Logic Weighted Ensemble, and an optimized weighted ensemble utilizing Sequential Least Squares Programming (SLSQP). According to experimental data, the optimized ensemble outperforms individual and alternative ensemble models, achieving the best performance with an R^2 value of 0.86, an RMSE value of 130.89, and an MAE value of 103.96. The suggested methodology improves the accuracy of the tea yield forecast while managing climate variability.**

Keywords—Random Forest; XG Boost; Light GBM; Cat Boost; SLSQP; Fuzzy; Neutrosophic

I. Introduction

Tea is a cash crop used for plantation in the world economy and plays a very important role in the agricultural economy of many developing nations. Tea crop (Camellia sinensis) is highly sensitive to climatic parameters, such as temperature, rainfall, humidity and solar irradiance. The alteration of climatic parameters hampers the productivity of the crop, as well as the sustainability of crop cultivation in the long run. Over the past few decades, climate variability and extreme weather have impacted the sustainability of tea crop production. Research has demonstrated the effect of climatic variables on the growth patterns of tea crop production, underscoring the importance of yield forecasting for proper management of tea crop production resources [1-2].

Recently, optimized weighted ensemble models for climate-based tea yield forecasting have attracted considerable attention within agriculture analytics. That is because these models can combine many different individual prediction models to produce an improved prediction accuracy of forecasted yield. Generally, statistical models frequently used for yield forecast have historically struggled with the non-linear relationship between climate variables and tea yield. However, machine learning algorithms such as Random Forest and XG Boost have demonstrated superior performance when working with larger data sets as well as managing the non-linear relationship of climate variables to tea yield than traditional regression-style forecasting models [3-4].

The climate greatly affects tea cultivation performance. This includes things like how hot or cold the weather is, how much rain falls in a specific region, and when the various growing periods occur within the season. There has been extensive research done in regard to how climatic conditions impact the amount of tea (i.e., tea yield) produced [5-6]. Climate change is also seen to be a key driving force that influences tea yields across many countries, including India, Kenya, Sri Lanka and China, as well as across other regions [7-8]. For instance, climatic conditions in regions where tea is grown, like Assam, play an important role in determining how easy or difficult it will be to produce large amounts of quality tea (i.e., the trend of tea yields), therefore requiring the development of sophisticated modelling techniques to assess what the potential yields will be for future growing seasons within those regions [2]. Likewise, variations in the way that rainfall patterns and temperature affect how much tea will be produced in each area where tea is produced are also dramatic.

Fuzzy set theory and Neutrosophic logic are two soft computing methods designed to help deal with the uncertain and imprecise nature of agricultural data. These methods can be used to consider incomplete, inconsistent, and uncertain data that frequently occur in agricultural and climatic datasets [9-10], thus enabling the use of these methods in predictive models, which not only strengthen the robustness of these models, but also improve their reliability when used in an applied setting. The advent of large-scale agricultural and climatic datasets has facilitated the development of data-driven forecasting models. There are significant historical datasets in the International Statistical Database that may serve as useful tools for the analysis of crop yield forecasting and for the development of

*Corresponding author

climatic forecasting models based primarily on sophisticated machine learning techniques [11-12].

To explore future challenges and provide opportunities in terms of establishing an optimized ensemble method using weights for climate-based tea yield prediction models, this study intends to find solutions. The authors have developed a prototype combination of various machine learning models while assigning optimal weights in order to create the best-performing model of its type. The purpose of this model is to examine how climatic influences combined with ensemble approaches can produce the most accurate estimates when predicting tea yields. It is hoped that such an approach will assist agriculture policymakers, tea planters and researchers with the decision-making process related to the management of tea as it relates to climatic impacts and sustainability.

II. Literature Review

The production of tea can be affected by several climatic and environmental components, which various modeling techniques have examined. Raj et al. [13] analyzed the long-term variation of tea production in some India tea growing areas using data from 1981 to 2015. Different statistical and time series techniques (like multiple regression, SARIMAX and artificial neural networks [ANNs]) were utilized and the authors concluded that the ANN and vector autoregressive model would be best to show Nonlinear interactions between climate input and tea yield, as well as identifying temperature variation as the primary driver for tea yield.

Batool et al. [14] developed a hybrid method for estimating tea yield through a combination of the FAO AquaCrop simulation model and machine learning models. Using weather, soil, crop, and management data from Pakistan, they demonstrated that predictive errors with machine learning (especially with the XG Boost algorithm) were lower than those associated with the simulation method, thus showing great promise for the prediction of tea yield using data-driven models.

In a study by Xiang et al. [15], ecological variables were controlled to determine the effect of temperature, humidity, light intensity, and substrate moisture on physiological function in tea plants. The study results indicate that some combinations of ecological variables exert a significant synergistic influence over shoot development, yield, and total catechin content of tea foliage, supporting a correlation between tea yield and quality. Raghav et al. [16] employed Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) networks for forecasting tea area, production and yield based on historical data from 1918 to 2023. Results of the study showed that the CNN model had lower error values than the LSTM model, which indicates that the CNN model has superior predictive capability relative to the LSTM model. This study illustrates that deep learning models are capable of accurately forecasting tea yield. Premkumar et al. [17] examined climate variability as it relates to tea production in Tamil Nadu using an ARDL error correction model. The authors concluded that lower temperatures and higher relative humidity promote tea production; however, rainfall negatively affects tea production.

The tea yield is known to be highly sensitive to both climate and environment, and recent research indicates that the traditional statistical models often do not adequately capture the complex or nonlinear relationship between climate and tea yield. While some advanced methods have demonstrated better success, such as machine learning, hybrid simulation/machine learning models, and controlled ecological experiments, there remain limitations concerning the applicability of these methods both regionally and operationally. In addition, temperature, rainfall and humidity - all identified as climatic variables relevant to tea cultivation - vary widely throughout the world; therefore, area-specific studies must be conducted. In light of these issues, this study will focus on developing an accurate and robust modelling method for predicting tea yields using climatic variables and advanced machine learning techniques.

III. Methodology

Fig. 1 represents an illustration of the proposed methodology's overview. In this example, we show how a machine learning-based approach provides predictions on tea production within India. First, a comprehensive dataset that is publicly available and encompasses a ten-year period and five major Indian states that are predominantly responsible for producing tea has been compiled. The study incorporates datasets collected across diverse temporal and environmental conditions, which inherently capture varying climatic scenarios and thus indirectly validate the model's robustness to climatic variability. The dataset will be split into model evaluation and model training datasets for two separate analyses. Model evaluation will be done by utilizing individual models against each data split during the model-building phase. The models that will be employed to develop the prediction model will consist of Random Forest (RF), XG Boost (XGB), Light GBM (LGBM), and Cat Boost models, which will be trained respectively on the newly created evaluation datasets. Further, an ensemble will be developed based on the combination of all four models trained to yield a better prediction for tea yield. The average weight applied to each of the models will be statistically determined based on the degree to which its prediction was accurate compared to the actual tea production on an annual basis, throughout the combined evaluation. The weight assigned to each model will then serve as the basis for developing the predictions for the current model based on this methodology. Algorithm 1 identifies the steps included in the proposed methodology.

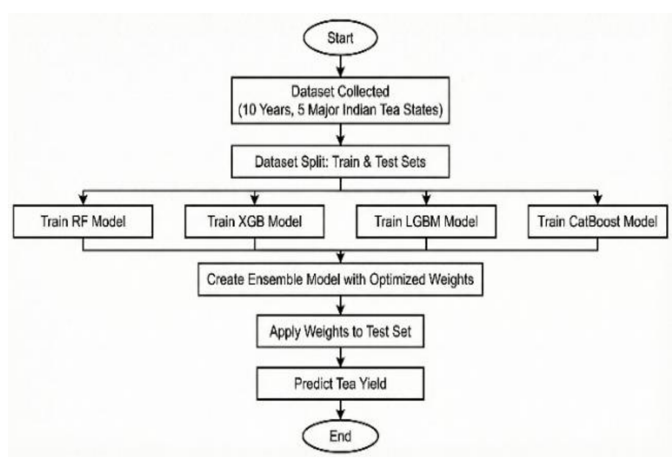

Fig. 1. Flow chart of proposed methodology.

Algorithm 1: SLSQP-based Optimized Weighted Ensemble Model

Require: Dataset D_{total}, Minimum weight threshold δ

Ensure: Final ensemble prediction $\hat{Y}_{final}$

Import training and test data

Split dataset into feature matrix X and target vector y

Perform one-hot encoding on categorical variables (State, District)

Split data into D_{train} and D_{val}

Train regression models: RF, XGB, LGBM, CAT on D_{train}

For each model $m_k \in \{RF, XGB, LGBM, CAT\}$ do

$$\hat{y}_{val}^{(k)} = m_k(X_{val})$$

End for

Form prediction matrix $P \in \mathbb{R}^{N_{val} \times 4}$:

$$P = \left[\hat{y}_{val}^{(1)}, \hat{y}_{val}^{(2)}, \hat{y}_{val}^{(3)}, \hat{y}_{val}^{(4)}\right]$$

For each model m_k do

$$RMSE_k = \sqrt{\frac{1}{N_{val}} \sum_{i=1}^{N_{val}} \left(y_i - \hat{y}_i^{(k)}\right)^2}$$

End for

Initialize weights using inverse RMSE:

$$w_k^{(0)} = \frac{1/(RMSE_k)}{\sum_{j=1}^{4} 1/(RMSE_j)}$$

Ensemble prediction:

$$\hat{y}_i = \sum_{k=1}^{4} w_k P_{i,k}$$

Objective function:

$$\min_w J(w) = \sqrt{\frac{1}{N_{val}} \sum_{i=1}^{N_{val}} (y_i - \hat{y}_i)^2}$$

Subject to equality constraint:

$$\sum_{k=1}^{4} w_k = 1, \qquad w_k \geq \delta$$

Apply Sequential Least Squares Programming (SLSQP)

Iteratively update weight vector w using quadratic approximation while satisfying constraints

Obtain optimal weight vector w^*

For each model m_k do

Retrain m_k on full dataset D_{total}

$$\hat{y}_{test}^{(k)} = m_k(X_{test})$$

End for

$$\hat{Y}_{final} = \sum_{k=1}^{4} w_k^* \hat{y}_{test}^{(k)}$$

Return $\hat{Y}_{final}$

Algorithm 1 describes the overall structure of the traditional optimized weighted ensemble paradigm, wherein multiple machine learning models are trained and the optimal weights are determined via constrained optimization techniques based on the validation data set. Each individual model is trained separately on the same training data and produces a corresponding prediction $\hat{y}_i$ for a given input. Initial weights are derived using the inverse Root Mean Square Error (RMSE) of each base model:

$$w_i^{(0)} = \frac{1/(RMSE_i)}{\sum_{j=1}^{M} 1/(RMSE_j)} \tag{1}$$

The ensemble weights are optimized by minimizing the RMSE of the ensemble on the validation set using the Sequential Least Squares Programming (SLSQP) optimization algorithm. The objective function is defined as:

$$\min_w \sqrt{\frac{1}{n} \sum_{k=1}^{n} \left(y_k - \sum_{i=1}^{M} w_i \hat{y}_{i,k}\right)^2} \tag{2}$$

The final ensemble prediction is computed as a weighted linear combination of the base model predictions, expressed as:

$$\hat{y} = \sum_{i=1}^{M} w_i \hat{y}_i \tag{3}$$

where, M denotes the number of base models and w_i represents the weight assigned to the i^{th} model. To ensure a valid convex combination, the weights are constrained to satisfy:

$$\sum_{i=1}^{M} w_i = 1, w_i \geq 0 \tag{4}$$

Each of the base learners is trained separately on the same training dataset and generates predicted values for the tea yield. The prediction error of each learner is calculated on the validation dataset using the Root Mean Square Error (RMSE) metric. The RMSE values are used as crisp inputs to the fuzzy logic system.

Fuzzification Using Gaussian Membership Function: Each RMSE value is converted to a fuzzy membership degree, which represents the level of trust in the respective model. A Gaussian membership function is used:

$$\mu_i = exp\left(-\frac{RMSE_i^2}{2\sigma^2}\right) \tag{5}$$

where, σ is the adaptive spread parameter, which is calculated as the mean RMSE of all base models:

$$\sigma = \frac{1}{4} \sum_{i=1}^{4} RMSE_i \tag{6}$$

The membership degree increases with lower RMSE values, representing higher confidence in the model.

Defuzzification: The fuzzy membership degrees are normalized to get the final ensemble weights:

$$w_i = \frac{\mu_i}{\sum_{j=1}^{4} \mu_j} \tag{7}$$

The final fuzzy ensemble prediction for the kth sample is calculated as a weighted sum of the individual predictions:

$$\hat{y}_k = \sum_{i=1}^{4} w_i \hat{y}_{i,k} \tag{8}$$

These values confirm that the fuzzy logic-based weighting approach successfully combines the predictions of the individual base models, providing accurate and robust predictions of tea yield even in uncertain conditions.

Extending this traditional methodology further, Algorithm 2 introduces the concept of fuzzy logic to determine the adaptive weights of the machine learning models based on the performance on the validation dataset.

Algorithm 2: Proposed Fuzzy Logic Weighted Ensemble (FWE) Algorithm

Require: Validation dataset $D_{val} = \{(x_i, y_i)\}_{i=1}^{n}$, Base models $M = \{m_1, m_2, \dots, m_k\}$

Ensure: Final ensemble prediction Y_{ens}

For each model $m_j \in M$ do

  $\hat{y}_{i,j} = m_j(x_i)$ for all $(x_i, y_i) \in D_{val}$

  Compute RMSE:

$$RMSE_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_{i,j}\right)^2}$$

End for

Compute adaptive spread parameter:

$$\sigma = \frac{1}{k}\sum_{j=1}^{k} RMSE_j$$

For each model $m_j \in M$ do

  Compute Gaussian membership degree:

$$\mu_j = exp\left(-\frac{RMSE_j^2}{2\sigma^2}\right)$$

End for

For each model $m_j \in M$ do

$$W_j = \frac{\mu_j}{\sum_{l=1}^{k} \mu_l}$$

End for

For a new input x_{new} do

$$Y_{ens}(x_{new}) = \sum_{j=1}^{k} W_j \cdot m_j(x_{new})$$

End for

Return Y_{ens}

Algorithm 2 shows a Fuzzy Logic Weighted Ensemble (FWE) method in which the weights of models are computed using Gaussian membership functions, which are functions of the corresponding RMSE values. In this approach, the contribution of each base learner in the ensemble output is adaptively controlled in proportion to the accuracy of the base learner, so that models with higher accuracy contribute more to the ensemble output. However, fuzzy logic essentially represents membership functions without considering any uncertainty or indeterminacy in the system output. Therefore, to improve the robustness of the fuzzy-based approach, Algorithm 3 proposes a Neutrosophic Weighted Ensemble (NWE) method, in which a fuzzy-based approach is generalized to consider all three components of a Neutrosophic set, namely, truth (measured in terms of R^2 values), indeterminacy, and falseness (measured in terms of normalized RMSE values), to obtain a more robust and reliable ensemble output, considering all the uncertainty in the system output, to obtain a more informative ensemble output compared to the fuzzy-based approach.

The Neutrosophic approach defines uncertainty using three parts: Truth (T), Indeterminacy (I), and Falsity (F). The values of T, I, and F are calculated for each base model using the validation performance metrics.

Truth Membership (T): Truth reflects the predictive reliability of a model and is defined using the coefficient of determination.

$$T_i = R_i^2 \tag{9}$$

Indeterminacy Membership (I): Indeterminacy captures model uncertainty.

$$I_i = 1 - R_i^2 \tag{10}$$

Falsity Membership (F): Falsity represents prediction error and is derived from the normalized RMSE.

$$F_i = \frac{RMSE_i - min\,(RMSE)}{max(RMSE) - min\,(RMSE)} \tag{11}$$

A single Neutrosophic score is computed through deneutrosophication to balance accuracy, uncertainty, and error.

$$S_i = \frac{2 + T_i - I_i - F_i}{3} \tag{12}$$

The final ensemble weight for each model is obtained by normalizing these scores.

$$w_i = \frac{S_i}{\sum_{j=1}^{4} S_j} \tag{13}$$

The Neutrosophic ensemble prediction:

$$\hat{y}_k = \sum_{i=1}^{4} w_i \hat{y}_{i,k} \tag{14}$$

The performance metrics show that the combination of Neutrosophic logic helps to strike a balance between the accuracy of the model and the uncertainty of the environment, resulting in a well-performing ensemble model for tea yield prediction. These three algorithms, in total, form a progressive and robust ensemble learning methodology. Algorithm 1 provides a solid foundation by optimally combining multiple machine learning models through constrained optimization, thus achieving balanced weight allocation mathematically. Algorithm 2 extends this methodology by incorporating fuzzy logic, thus achieving adaptive and performance-sensitive weighting according to validation errors, thus increasing the flexibility of the methodology. Algorithm 3 extends this methodology by incorporating Neutrosophic logic, thus achieving more efficient management of uncertainty and variability in the performance of models, thus achieving a more reliable, intelligent, and robust ensemble prediction system, moving from a deterministic approach to uncertainty management to intelligent weighting.

Algorithm 3: Proposed Neutrosophic Weighted Ensemble (NWE) Algorithm

Require: Validation dataset $D_{val} = \{(x_i, y_i)\}_{i=1}^{n}$, Base models $M = \{m_1, m_2, \dots, m_k\}$

Ensure: Final ensemble prediction Y_{ens}

For each model $m_j \in M$ do

  Compute validation metrics: R_j^2 and $RMSE_j$ on D_{val}

End for

$E_{min} = min(RMSE_j)$
$E_{max} = max(RMSE_j)$
For each model $m_j \in M$ do
 Truth membership: $T_j = R_j^2$
 Indeterminacy membership: $I_j = 1 - R_j^2$
 Falsity membership:
$$F_j = \frac{RMSE_j - E_{min}}{E_{max} - E_{min}}$$
End for
For each model $m_j \in M$ do
 Compute neutrosophic score:
$$S_j = \frac{2 + T_j - I_j - F_j}{3}$$
End for
Normalize scores to obtain weights:
$$W_j = \frac{S_j}{\sum_{l=1}^{k} S_l}$$
For a new input instance x_{new} do
$$Y_{ens}(x_{new}) = \sum_{j=1}^{k} W_j \cdot m_j(x_{new})$$
End for
Return Y_{ens}

IV. Results

A. Dataset Description

The empirical data are in numerical form, representing observations of tea production (kg/ha) and other agro-meteorological factors of major tea-yielding states like Assam, Himachal Pradesh, Kerala, Tamil Nadu, West Bengal, recorded at the district level during the period 2015 to 2025. The dataset is taken from [12]. Table I shows the description of the dataset that includes temporal factors and other climatic variables such as temperature, rainfall, relative humidity, and sunshine hours. The data is presented in arithmetic mean values, along with the minimum and maximum ranges, thus reflecting the variability of the environmental factors that affected tea production during the period. The data is given as a bar chart in Fig. 2.

TABLE I. Dataset Description

Variables	Mean (min-max)
Tea Yield (kg/ha)	1750.27 (1021.0–2923.4)
Maximum Temperature (°C)	29.09 (19.2–37.8)
Minimum Temperature (°C)	19.25 (2.6–26.5)
Rainfall (mm)	119.48 (0.0–485.6)
Relative Humidity – Morning (%)	85.24 (65.0–100.0)
Relative Humidity – Afternoon (%)	66.91 (34.7–96.7)
Sunshine Hours (hours)	5.60 (0.2–11.5)

The input of these regression algorithms is meteorological factors, and the target value is the observed tea yield. Data is split into a ratio of 80% and 20%. We have utilized 80% of the data for training the models, and the rest of 20% is used for testing.

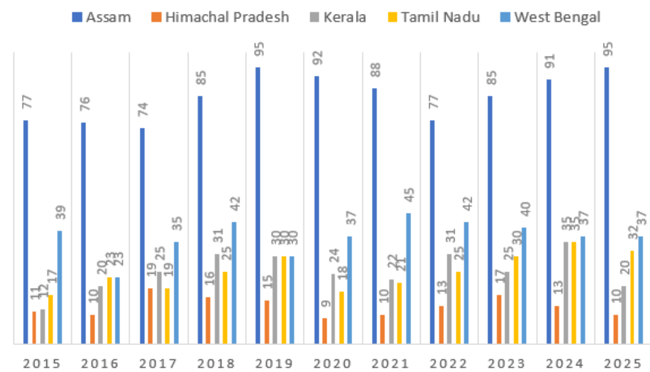

Fig. 2. Year-wise data collection

B. Performance of the Machine Learning Techniques for Tea Yield Prediction

Here we use four initial machine learning models, Random Forest, XG Boost, LGBM, Cat Boost for implementation, and also use a weighted ensemble model for better results. The performance metrics are as follows:

Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2} \tag{15}$$

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum(y - \hat{y})^2} \tag{16}$$

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum|y - \hat{y}| \tag{17}$$

In the above equations, y is the observed yield and $\hat{y}$ is the predicted yield, and n is the total number of samples. The performance of all the machine learning models and the ensemble model is given in Table II and Fig. 3.

TABLE II. Performance of Machine Learning Techniques

Model	R^2	RMSE	MAE
Random Forest	0.84	139.69	111.07
XG Boost	0.83	141.92	112.49
Light GBM	0.83	144.26	116.09
Cat Boost	0.85	131.94	104.03
SLSQP-based Ensemble Model	0.86	130.89	103.96

Table II shows the comparative assessment of the models, with each model's performance compared and evaluated using R^2, RMSE, and MAE. From the table, it is clear that the Cat Boost model outperforms the Random Forest, XG Boost, and Light GBM models, as it has the highest R^2 and lower error values. However, the Ensemble Model, which is based on Sequential Least Squares Programming, has the highest R^2 value of 0.86 and the least error values of 130.89 and 103.96, respectively. This shows that the ensemble model is more reliable compared to the other models. The ensemble model is more effective compared to the other models. Optimized weights are: Random Forest: 0.26, XG Boost: 0.01, Light GBM: 0.02, Cat Boost: 0.69.

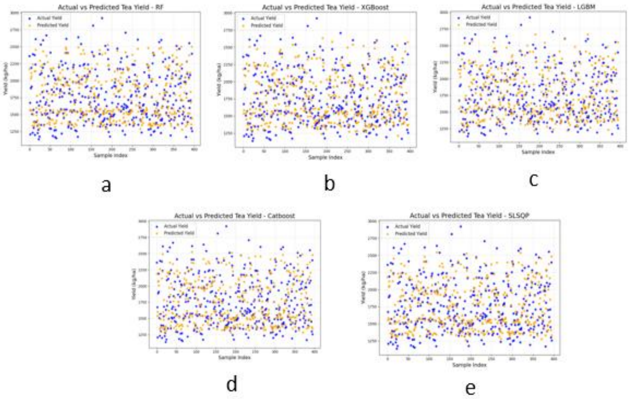

Fig. 3. Graphical representation of machine learning models: a) Random Forest, b) XG Boost, c) Light GBM, d) Cat Boost, e) Weighted ensemble model.

C. Performance for Fuzzy Logic for Tea Yield Prediction

To overcome the uncertainty associated with the prediction of tea yield and to apply dynamic weights to the learners, an ensemble framework using fuzzy logic is developed. The proposed ensemble framework combines four tree-based regression learners: Random Forest (RF), XG Boost (XGB), Light GBM (LGBM), and Cat Boost (CAT). Each of the base learners is trained separately on the same training dataset and generates predicted values for the tea yield. The prediction error of each learner is calculated on the validation dataset using the Root Mean Square Error (RMSE) metric. The RMSE values are used as crisp inputs to the fuzzy logic system. Table III and Fig. 4 show the performance of the Fuzzy-Based Ensemble Model.

TABLE III. PERFORMANCE OF FUZZY ENSEMBLE MODEL TECHNIQUES

Model	R^2	RMSE	MAE
Fuzzy-based Ensemble Model	0.84	136.32	108.77

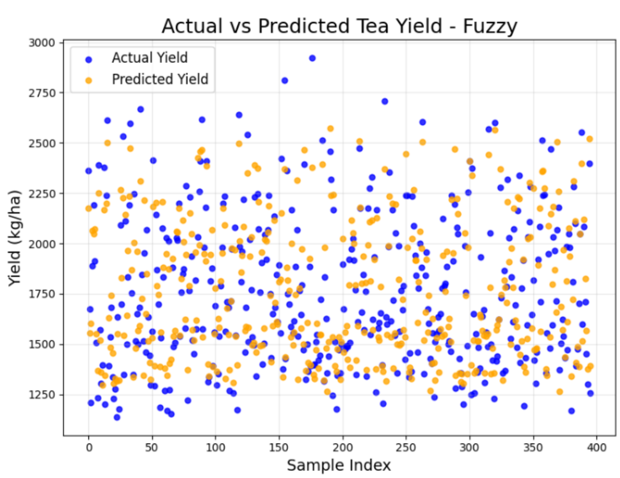

Fig. 4. Graphical representation of fuzzy logic-based tea yield prediction.

The model provides an R^2 value of 0.84979, which shows satisfactory accuracy for the model's predictions. The RMSE value of 136.3297 and the MAE value of 108.7715 show that the errors associated with the predictions are relatively low. Overall, the fuzzy ensemble model shows good performance and improves the accuracy of predictions by optimally assigning the weights according to the performance of the models. Optimized weights are: Random Forest: 0.2512, XG Boost: 0.2472, Light GBM: 0.2431, Cat Boost: 0.2585.

D. Performance for Neutrosophic Logic for Tea Yield Prediction

To enhance the accuracy of tea yield prediction in an uncertain environment, a Neutrosophic ensemble learning approach is proposed by combining four tree-based regression models: Random Forest (RF), XG Boost (XGB), Light GBM (LGBM), and Cat Boost (CAT). Each base model is trained separately on the same training data and produces predicted values of the tea yield, represented as $\hat{y}_i$.

The Neutrosophic approach defines uncertainty using three parts: Truth (T), Indeterminacy (I), and Falsity (F). The values of T, I, and F are calculated for each base model using the validation performance metrics, and the metrics are shown in Table IV.

TABLE IV. NEUTROSOPHIC METRICS

Model	T	I	F
RF	0.8423	0.1577	0.4725
XG Boost	0.8372	0.1628	0.7301
LGBM	0.8318	0.1682	1.0000
Cat Boost	0.8514	0.1486	0.0000

A single Neutrosophic score is computed through deneutrosophication to balance accuracy, uncertainty, and error.

$$S_i = \frac{2+T_i-I_i-F_i}{3} \tag{18}$$

The final ensemble weight for each model is obtained by normalizing these scores.

$$w_i = \frac{S_i}{\sum_{j=1}^{4} S_j} \tag{19}$$

The performance of the Neutrosophic fuzzy model is shown in Table V and Fig. 5.

TABLE V. PERFORMANCE OF NEUTROSOPHIC ENSEMBLE MODEL TECHNIQUES

Model	R^2	RMSE	MAE
Neutrosophic Ensemble Model	0.85	135.94	108.32

Table V shows the overall performance of the Neutrosophic Ensemble Model. From the table, it can be concluded that the model has an R^2 value of 0.85064, which shows high predictive accuracy. At the same time, the model has an RMSE value of 135.9409 and an MAE value of 108.3259, which shows that the errors obtained from the model are quite low. Therefore, the Neutrosophic ensemble model has shown good and reliable performance while dealing with the uncertainty that occurs while evaluating the model and computing the weights. Optimized weights are: Random Forest: 0.2596, XG Boost: 0.2281, Light GBM: 0.1952, Cat Boost: 0.3171.

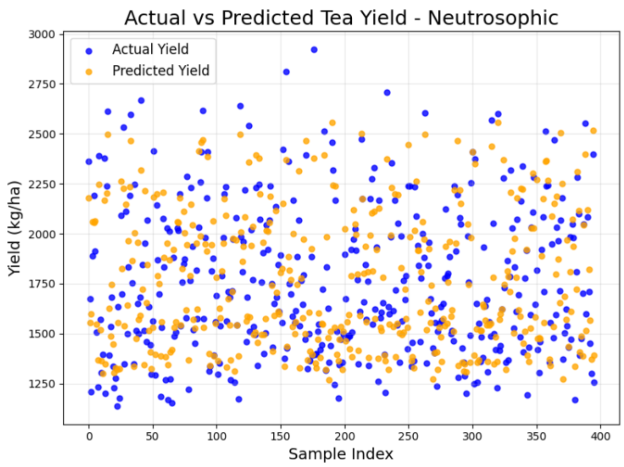

Fig. 5. Graphical representation of Neutrosophic-based tea yield prediction.

Performance of our proposed model (Sequential Least Squares Programming (SLSQP) algorithm) is improved and the results are found to be the best, with the R^2 value being high (0.8629) and the error values being low. The Fuzzy Ensemble Model and the Neutrosophic Ensemble Model are found to have high performance. The performance of the Neutrosophic Ensemble model is slightly better than the Fuzzy Ensemble model. This is evident by the R^2 value being high and the error values being low. Overall, the results show that the performance of the Ensemble model is high compared to the individual algorithms. The Ensemble model is found to have the most precise and reliable performance. The comparative analysis of our proposed ensemble model with the methods in the earlier literature is given in Table VI.

TABLE VI. COMPARISON STUDY WITH EARLIER LITERATURE

References	R^2	RMSE	MAE
Raj et al. [13]	0.612	--	--
Batool et al. [14]	--	0.154	0.123
Xiang et al. [15]	0.70	--	--
Raghav et al. [16]	--	1259.1	1086.4
Premkumar et al. [17]	0.61	--	--
Proposed Method	**0.86**	**130.89**	**103.96**

V. CONCLUSION

This study proposes various machine learning and ensemble techniques for predicting tea yield using climatic data. The proposed methodology, the optimized weighted ensemble method based on Sequential Least Squares Programming (SLSQP) is found to provide the highest prediction accuracy compared to other machine learning models. However, the models. Fuzzy and Neutrosophic ensemble models are proposed to improve the reliability of prediction techniques based on uncertainty handling capabilities. The Neutrosophic model is found to perform slightly better compared to the fuzzy model for handling uncertainty conditions.

Hence, our proposed methodology provides better prediction accuracy than Fuzzy and Neutrosophic-based models. This study can be extended by incorporating climatic data from various tea-producing states of world-wide to improve the proposed prediction techniques using machine learning models. Deep learning techniques can also be explored to improve the prediction accuracy of the proposed prediction techniques.

ACKNOWLEDGMENT

The authors sincerely acknowledge the support of JIS Group for providing financial assistance to carry out this research work. The authors are also grateful for the institutional support, infrastructure, and research environment that facilitated the successful completion of this study. Special thanks are extended to all individuals and collaborators who contributed their time and expertise toward this work.

REFERENCES

[1] S. Ahmed, J. R. Stepp, C. Orians, T. Griffin, C. Matyas, A. Robbat, S. Cash, E. Kennelly, "Effects of extreme climate events on tea (Camellia sinensis) functional quality validate indigenous farmer knowledge and sensory preferences in Tropical China", PLoS One, Vol. 9, No. 10, 2014, Article e109126.

[2] J. M. A. Duncan, S. D. Saikia, N. Gupta, E. M. Biggs, "Observing climate impacts on tea yield in Assam, India", Applied Geography, Vol. 77, 2016, pp. 64–71.

[3] L. Breiman, "Random forests", Machine Learning, Vol. 45, No. 1, 2001, pp. 5–32.

[4] T. Chen, C. Guestrin, "XGBoost: A scalable tree boosting system", Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.

[5] M. K. V. Carr, "The climatic requirements of the tea plant: A review", Experimental Agriculture, Vol. 8, 1972, pp. 1–14.

[6] M. K. V. Carr, W. Stephens, "Climate, weather and the yield of tea", in Tea, Dordrecht: Springer Netherlands, 1992, pp. 87–135.

[7] M. Wijeratne, A. Anandacoomaraswamy, M. Amarathunga, J. Ratnasiri, B. Basnayake, N. Kalra, "Assessment of impact of climate change on productivity of tea (Camellia sinensis L.) plantations in Sri Lanka", Journal of the National Science Foundation of Sri Lanka, Vol. 35, 2007, pp. 119–126.

[8] J. Sitienei, S. G. Juma, E. Opere, "On the use of regression models to predict tea crop yield responses to climate change: A case of Nandi East, sub-county of Nandi county, Kenya", Climate, Vol. 5, 2017, Article 54.

[9] L. A. Zadeh, "Fuzzy sets", Information and Control, Vol. 8, No. 3, 1965, pp. 338–353.

[10] F. Smarandache, "Neutrosophic logic and its applications", American Research Press, 2010.

[11] FAO, "FAOSTAT statistical database", Food and Agriculture Organization of the United Nations, 2022.

[12] Biswas, "All Over India TEA YIELD dataset", Kaggle Dataset, Available: https://www.kaggle.com/datasets/arpanbiswas001/all-over-india-tea-yield

[13] E. E. Raj, K. V. Ramesh, R. Rajkumar, "Modelling the impact of agrometeorological variables on regional tea yield variability in South Indian tea-growing regions: 1981–2015", Cogent Food & Agriculture, Vol. 5, No. 1, 2019, Article 1581457.

[14] D. Batool, M. Shahbaz, H. S. Asif, K. Shaukat, T. M. Alam, I. A. Hameed, Z. Ramzan, A. Waheed, H. Aljuaid, S. Luo, "A hybrid approach to tea crop yield prediction using simulation models and machine learning", Plants, Vol. 11, No. 15, 2022, Article 1925.

[15] P. Xiang, Q. Zhu, M. Tukhvatshin, B. Cheng, M. Tan, J. Liu, J. Lin, "Response of shoot growth to ecological factors highlights a synergistic relationship between yield and catechin accumulation in tea plant (Camellia sinensis L.)", Horticulturae, Vol. 11, No. 6, 2025, Article 624.

[16] Y. S. Raghav, O. A. Alqasem, N. Mishra, B. Kumar, G. Chandel, P. Mishra, K. M. Alakdari, "Forecast using LSTM-CNN model (area, production and yield rate) of tea in India", Journal of Animal & Plant Sciences, Vol. 35, No. 3, 2025, pp. 769–779.

[17] Premkumar, R. Kishan, D. Kalaiarasi, "Repercussions of climatic variabilities on tea production in Nilgiris district of Tamil Nadu, India", Journal of Agrometeorology, Vol. 27, No. 2, 2025, pp. 173–176.