International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7, 2025
https://theaspd.com/index.php

"Hybrid SNN–ANFIS Framework for Predicting Crop Yields Under Climate Change Scenarios: a Case Study of Maharashtra, India"

Sandeep Kumar Vishwakarma[1], Dr. Vikas Kumar[2]
[1]Ph.D Scholar, Computer Science, Department of Computer Science & Information Technology, Chhatrapati Shivaji Maharaj University, Panvel, Navi Mumbai, Maharashtra- 410221, India
sandeepvcbs@gmail.com
[2]Head, Department of Computer Science & Information Technology, Chhatrapati Shivaji Maharaj University, Panvel, Navi Mumbai, Maharashtra- 410221, India, vikaskumar@csmu.ac.in

Abstract
Precise estimation of the crop yield in the context of changing climatic conditions is the key issue to the food security and the agricultural decision-making. This paper presents a hybrid model which uses semiparametric Neural Networks (SNN), combined with Adaptive NeuroFuzzy Inference System (ANFIS) to predict crop yields of a set of chosen districts in Maharashtra, India. The model uses the past data (2000 2022) which includes data about climate variables (rainfall, temperature, solar radiation, and the level of CO2 in the atmosphere) together with local crop yields. The first was that the dataset was processed and standardized and a geospatial mapping of the study area was performed to contextualize local differences. Trained and tested hybrid SNN ANFIS against machine learning baselines, such as SVR, ANN, CNN-RNN, and compared based on RMSE, R2, and MAPE, the hybrid SNN ANFIS had better performance. Fuzzy inference is optimized with Gaussian membership functions, and the interpretation of the model was improved using feature importance analysis and Partial Dependence Plots (PDP). Generalizability of the model was measured by depicting spatial heatmap that shows that the prediction accuracy was similar across different districts thereby confirming prediction consistency. The findings indicate that rainfall and temperature have been found to be the most statistically significant factors when it comes to determining the yield with the hybrid model attaining %92 accuracy in the test data. In addition to enhancing the accuracy of the prediction, the suggested framework is suitable in real life agri-policy formulation and climate adaptation approaches because it promotes explainability.
Keywords: *Crop Yield Prediction; Climate Change; Semi-Parametric Neural Network (SNN); Adaptive Neuro-Fuzzy Inference System (ANFIS); Maharashtra Agriculture; Explainable AI*

1. INTRODUCTION

The impact of unfavorable climate changes is largely becoming a threat to the farming industry, especially in some areas such as Maharashtra, India where the agricultural processes are mainly rain-fed and are very vulnerable to environmental fluctuations. Increased temperatures, unpredictable rainfall, the fluctuation in radiation and the increased levels of CO 2 in the atmosphere have rendered conventional techniques of agricultural forecasting invalid and inadequate. The increasing irreliability of the seasonal parameter and its influence on crop productivity demand the creation of tailored, local, and dynamic prediction models capable of helping to reduce the unpredictability and guarantee food security under such conditions (Feng et al., 2023; Dumitru et al., 2023). The use of efficient predictive frameworks has become a requirement in the recent years following the growing population pressure, shrinking arable land, and the worldwide sustainable agriculture trend. Correct crop yield forecasting is key to farmer, policymaker, and supply chain decision making. These forecasts help in optimal distribution of resources, foodstock and food transport planning and policy development of climate resilient farming methods. Since there is a reason to suppose that such an integration of meteorological science and computational modeling can prove to be an effective way of predicting crop performance, particularly in such complex climate regions like Maharashtra as Ehteram et al. (2023) stress.

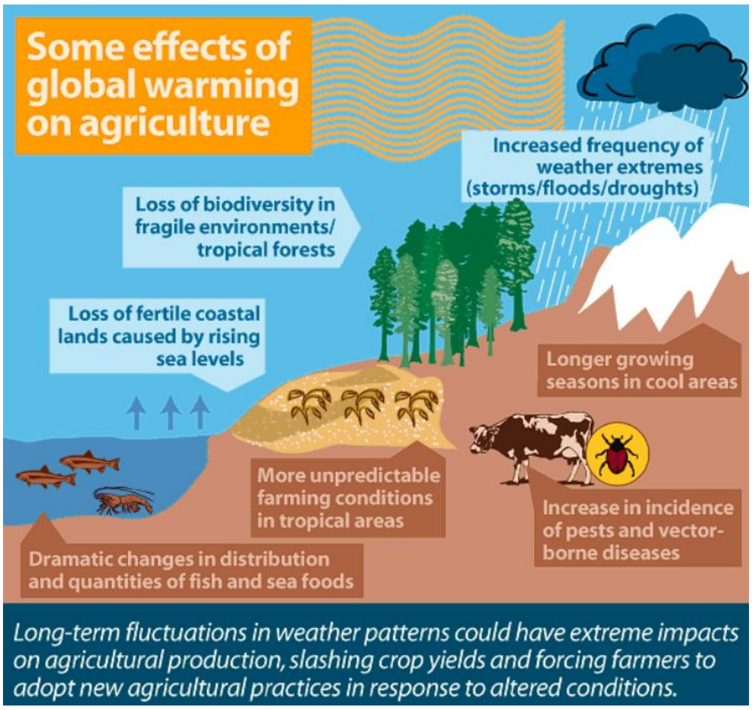


Figure 1: Key impacts of global warming on agricultural systems, including exacerbated weather extremes, shifting growing seasons, biodiversity loss, and increased pest pressures (Agrivi, year).

Historically precious but now mostly inadequate, such traditional statistical models are not able to evoke non-linear and multifactorial dependence between climatic parameters and crop output. This weakness has increased the use of artificial intelligence (AI) and machine learning (ML) methods in agro-research and decision support mechanisms (Benos et al., 2021; Odah et al., 2025). Convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and artificial neural networks (ANNs) are some of these techniques that have been proven fruitful in predicting crop yields in a more accurate manner, especially in cases when they are paired up with environmental data and phenological data (Srivastava et al., 2022; Jabed & Murad, 2024). Although standalone ML models appear to be effective, there has been an emerging demand to hybridise methods combining the strengths of various analyses to predict improved models in a manner that is more interpretable. As an example, Baswaraju et al. (2023) suggested a hybrid deep learning prediction model consisting of AROA and attested higher food production output on the agricultural datasets. On the same note, Khalilzadeh (2024) examined the usage of hybrid deep learning-based optimization to improve the productivity of crops using data-driven alternatives. In addition, these hybrid models help to overcome the shortcomings associated with individual algorithms because they are by far flexible to manage the nonhomogenous nature of input information and dynamicity of the environmental features used.

Models based on Artificial Neural Networks are especially good at learning non-linear, and complex relationships in the data, and frequently lack interpretability. On the one hand, ANFIS has strengths that consist of the fact that they are based on an adaptive rule learning system and have human-understandable logic (Hara et al., 2021). The type of combination of Semi-Parametric Neural Networks (SNN) with ANFIS would provide a rare chance of combining accuracy and interpretation in yield prediction. The SNN and ANFIS hybrid architecture has the advantage of using the flexibility and non-linear modelling property of SNNs to achieve an optimal representation of the training data and avoid overfitting, but can combine it with the fuzzy rule-based adaptability of ANFIS to interpret uncertainty and imprecision in the informationa situation common in climate data modelling. Changes in climate in general, and climate variability in particular, exposes the farmer in Maharashtra to severe challenges in terms of maintaining a consistent crop productivity. There exists immense spatial and temporal rainfall, solar radiation, and temperature variation in the region, which plays a very vital role in the directness of crops at different stages of their growth period. Though there are many studies which address yield prediction with climate variables, majority of them do not have localized, hybrid predictive functions to address the climatic and agricultural conditions in Indian states such as Maharashtra (Lucca et al., 2023). It is possible to combine historical meteorological data with machine learning models to memorize previous trends as well as the appearance of anomalies, which can also serve as a stronger method of prediction (Mehrbakhsh & Rabab Ali, 2023; Kaginalkar et al., 2022).

In addition to that, big data analytics and predictive modeling are increasingly used in agri-environmental studies. The combination of meteorological observation, remote sensing, and AI-based systems has opened up the possibility of increased responsive dynamic agriculture planning (Ehteram et al., 2023). The existing efforts in data governance systems, including the development of conventions examined by Kaginalkar et al. (2022), will also facilitate the introduction of a variety of environmental data onto the ML pipelines. All these developments point to the fact that region-specific, hybrid model that can be able to capture the effects of climate change on agricultural productivity can be developed in an urgent manner.This study is directed towards meeting this demand by creating a hybrid predictive model that will combine Semi-Parametric Neural Networks (SNN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) in order to predict the crop yields in the climate-sensitive state of Maharashtra. This model would aim at being predictively accurate and flexible to a variety of climate conditions by using long-term historic points including parameters, such as rain, temperature, CO_2 concentration and solar radiation.

The rest of this paper is organized as follows: In section 2 finer literature review is made with special reference to ML, hybrid framework and interaction between climate and agriculture. Section 3 describes the methodology, data collection, preprocessing, model design and how integration process of the SNNANFIS hybrid framework occurs. In Section 4, the results and analysis of the experiments are highlighted and their performance in different climate scenarios are reported. Lastly, Section 5 provides conclusions and future research directions of interest on the field of climate-resilient agricultural planning.

2. LITERATURE REVIEW

2.1 Introduction to Machine Learning in Agriculture

Machine learning (ML) in agriculture current research has shifted beyond using raw data on statistical tests to making inferences on well-developed predictive models that can handle non-linear and complex relations in the agro-climatic data. In the initial phases, linear regression, time series models, and parametric statistical modelings were heavily used in the agricultural forecasting but the results used to fail to reveal the complex interactions between environmental conditions and agricultural performance (Jabed & Murad, 2024). The growing instability of the climatic parameters (rainfall, temperature and solar radiation) has led to the need to switch to adaptable, data-based systems instead of traditional ones.

Recent advances in ML and artificial intelligence (AI) have given way to powerful tools with the ability to handle datasets in an increasingly high-dimensional agricultural setting. Such tools are support vector machines (SVM), decision trees (DT), random forests (RF), and artificial neural networks (ANNs) that proved to be more efficient in terms of working both with nonlinearities and deriving meaningful structures out of large-scale data (Van Klompenburg et al., 2020). Furthermore, the development of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) as the variations of deep learning (DL) has increased the predictive potential of ML in agriculture since it introduced spatial and temporal dynamics to the model. Advances in open-access meteorological data and the availability (due to IoT-enabled sensors) of satellite images and other kinds of imagery have led to the creation of precise yield forecast systems. Nikhil (2024) has developed a system to be applied to the agrarian complex level and to the specific environment of South India, where it is vital to lead the interaction of climate, soil, and crops within small distances on the landscape. Comprehensively, this movement of traditional statistical models to smart ML structures has given birth to new opportunities in precision agriculture and resilient farming systems to climate change.

2.2 Yield Prediction and Regional Applications

Yield prediction models at regional levels are becoming significant especially in climate sensitive zones like that of India. Because of the high intra-regional climatic variability, localized models are needed in order to embrace the interactions peculiar to the environmental inputs into crop phenology. In that regard, incorporating high-resolution spatial clarity and temporal information, including daily temperature, precipitation, and CO_2 concentration is the essential aspect in order to optimize the precision of the crop forecasting systems (De Clercq & Mahdi, 2024).ML models in the Indian context have been tailored to the country specific dataset to capture the heterogeneity in Indian agriculture. Nikhil (2024) showed that the ML models, such as SVM and RF, are highly effective in prediction of crop yields in different districts in South India. Such models are trained on datasets containing agro-climatic and soil moisture data along with historical yields as well, the significance of spatial granularity can be seen.

Khaki, Wang, and Archontoulis (2019) went even further: they tried to use CNN RNN frameworks to predict the maize yields in the U.S. Corn Belt, combining not only spatial information but also temporal patterns. Their work was towards the U.S., yet it provides additional focus on how it can be applied to adapting the same deep learning structures to the regions in India. Their approach offered an appropriate methodology of developing crop stages against time-series weather patterns that will pave the way towards developing localized models before applying them to the training of district-level data in Maharashtra. By doing this region-centric approach, the ML models will be able to put consideration to localized climatic stressors which is very relevant to case of Maharashtra- a state facing very varied rainfall, soils and crop patterns. In this way, the region-specific models have a viable roadmap to climate-adaptive yield prediction.

2.3 Hybrid and Ensemble Machine Learning Models

Though single ML models have demonstrated the potential, it has been observed that hybrid and ensemble ML models have gained popularity because of high predictive value in the agricultural environment. By combining the weaknesses of the individual models, the limitations may be overcome provided that the diverse types of models complement each other in their strengths. As an example, neural networks can be used to provide neural network learning capabilities and fuzzy systems to provide interpretability; hybrid frameworks can in turn incorporate metaheuristic optimization procedures to tune model parameters.

Baswaraju et al. (2023) suggested a hybrid deep learning based on the Adaptive Rain Optimization Algorithm (AROA) that enhanced the accuracy in the prediction of food production data by a large margin. This model proved that the union between optimization methods and deep learning is able to improve the convergence and generalization of predictive systems applied in the agriculture field. On the same note, Nosratabadi et al. (2020) also created a hybrid model through artificial neural networks (ANN) that was optimized with Grey Wolf Optimizer (GWO). Their method compared favorably with the classical ANN models because it minimized overfitting and increased robustness on several contexts. Putting evolutionary computation in the framework of yield prediction enables to address the issue of uncertainty and dynamically changing climate behavior with greater ease.

Abdel-Salam, Kumar, and Mahajan (2024) developed this idea further offering a hybrid model that implements optimization in support vector regression (SVR) and feature selection methods. The model used by them successfully screened redundant inputs and increased the estimation rate hence producing a more accurate set of yield predictions. Moreover, the authors of this study suggested the use of a graph neural network (GNN) and a recurrent neural network (RNN) ensemble to utilize both spatial and temporal dependencies of crops data (Fan et al., 2021). This architecture was particularly powerful in repurchasing multi-location yield patterns which is an indication that hybrid models may resolve the tradeoff of precision and volume. All these researches show the utility of hybrid modeling, especially in circumstances where the data are high-dimensional spatio-temporal agricultural data.

2.4 Role of Explainability and Model Optimization

The black-box formula in machine learning models has been one of the recurring issues in the application of these systems in agriculture as it restricts interpretability and usability in the real-world context by farmers, agronomists, and other stakeholders of such policy. Though deep learning and ensemble models have demonstrated high levels of accuracy, their complex machinations in most cases make trust in and readability of their work difficult to follow in high-stakes agricultural decision-making (Yenkikar, 2025). Such an issue has caused the appearance of Explainable Artificial Intelligence (XAI) methodologies in agro-informatics with the goal of exposing the model behavior.

According to Yenkikar (2025), the adoption of predictive models might be much higher with the inclusion of interpretability features, including model-agnostic explanations, attention mechanisms and visual interpretation tools. In the same vein, Abdel-Salam, Kumar, and Mahajan (2024) proposed a hybrid model of the support vector regression approach and mechanisms of using advanced feature selection. Their model did not just enhance the precision of predicted values but also assisted in defining the most important environmental and agronomic characteristics which determine the results of the yield. These insights are very important to domain experts and enable targeted interventions. Li (2025) has added to this field as well by implementing his idea of a knowledge-guided architecture of ML to see how agronomic domain knowledge may be built into model training. Such a human-in-the-loop design is useful to close the gap between domain know-how and the data-driven learning. Such mechanisms of interpretability, when contextualized with a relatively complex arrangement associated with hybrid frameworks such as SNNANFIS due to the co-existence of fuzzy rule-based logic and neural networks, can be a decisive factor contributing to the model accuracy and transparency in the crop yield forecasting oversight.

2.5 Climate and Environmental Data Integration

Extensive, cross-linked environmental data--historical data, real-time sensor data, etc.--has become central to the effectiveness of crop yield predicting models. The right forecasting will entail consideration of both the physiological behaviour of the crop and the changeable climatic situation during the growth period. Models can learn weather patterns and anomalies over long periods because climate reanalysis datasets can be used that combine satellite and meteorological data.

Severe weather outcomes confront both the rice cultivation industry and governments to make accurate estimates in rice production potential. De Clercq and Mahdi (2024) demonstrated the usefulness of reanalysis data in Indian rice yield forecasting, with their research findings showing that the potential quality of climate data at high resolutions enhances precision of applied models. The temporal granularity and climate normalization were also important issues towards model training as highlighted in their study. In the meantime, Anbananthen et al. (2021) showed how an intelligent decision support system might be created that could combine real-time agro-climatic parameters provided by IoT devices to increase the responsiveness of ML algorithms based on current field conditions.

Talaat (2023) complemented that with the creation of Crop Yield Prediction Algorithm (CYPA) that uses precision agriculture tools such as IoT, satellite sensing in order to get data in real-time. It is adopted to ensure dynamic updating of the models, which is also applicable in the monsoon dependent states in India such as Maharashtra. Overall, the studies justify the benefit of including multi-source environmental data with ML pipelines, particularly, with regard to hybrid models such as SNN- ANFIS, which can be used to improve the situational specificity and stability of crop production estimations.

2.6 Research Gaps and Opportunities

Although there is an increasing literature review regarding machine learning in agricultural applications, there are still some research gaps especially in terms of hybrid model implementation in India. In the first place, although deep learning, SVM, and ANN approaches have been extensively used in many studies, hybrid SNNs and ANFIS frameworks that accommodate the Indian climatic regions are still very scarce. This weakness imposes a limitation of the model to solve both non-linearities and interpretability at the same time.

Secondly, the majority of the available literature on this topic, including studies conducted by the likes of Nikhil (2024), De Clercq and Mahdi (2024), and Khaki et al. (2019), deal with either areas across a given continent or crops in general, without taking into account localized data sets on a region that spans varied agro-climatic zones as is the case with Maharashtra. Since there is a high amount of geospatial variability in rainfall, temperature, and soil quality across different districts, the modeling trained on multi-district or even international data might not provide a proper contextually accurate model in a local sense. In addition, though certain researches have started using XAI approaches (Yenkikar, 2025; Li, 2025), most of them continue to employ black-box strategies, and it negatively affects stakeholder credence and real-world applicability. Thus, this paper suggest hybrid SNNANFIS model with specific details about region-specific climate data and explainability properties to address these essential limitations of current studies. In a bid to build on what has been discovered in all of the reviewed studies, Table 1 shows a comparative overview of the main methodologies, source of data, discoveries, and limitations. This table facilitates visually grabbing the scope, advances, and the limitations of the existing literature and also the support area of the necessity to propose a hybrid SNNANFIS framework in the localized prediction of crop yield in the Maharashtra region.

Table 1: Summary of Literature Review

S.No.	Author & Year	Model/Method Used	Data Type	Key Findings	Limitations/Scope
1	Jabed & Murad (2024)	ML/DL comparative review	Multiple crops, global	Highlights DL dominance	Needs region-specific insight
2	Van Klompenburg et al. (2020)	Systematic ML review	Historical yield data	ML improves over linear models	Interpretability issues

3	Nikhil (2024)	ML comparative (India)	Climate + soil	SVM and RF perform well	No hybrid models used
4	Baswaraju et al. (2023)	Hybrid DL (AROA-based)	Food production stats	AROA hybrid improved accuracy	Focused on food not climate
5	Anbananthen et al. (2021)	Hybrid ML	Real-time agri data	Decision support via ML	Not Maharashtra-specific
6	Abdel-Salam et al. (2024)	SVR + Feature Selection	Climate + crop	Optimized yield forecasting	No ensemble analysis
7	Fan et al. (2021)	GNN–RNN	Geo-temporal data	Strong spatio-temporal correlation	Applied only to maize
8	De Clercq & Mahdi (2024)	ML + reanalysis climate data	Rice yield (India)	High feasibility with climate inputs	Crop-specific (rice)
9	Yenkikar (2025)	Explainable AI hybrid	Multivariate climatic	Enhances trust in predictions	Few validation datasets
10	Li (2025)	Knowledge-Guided ML	Crop modeling	Boosts interpretability	Needs field deployment
11	Nosratabadi et al. (2020)	ANN–GWO hybrid	Simulated agri data	Better generalization	Limited real-time testing
12	Khaki et al. (2019)	CNN–RNN	US Corn Belt data	High accuracy in yield	Non-transferable to India
13	Talaat (2023)	CYPA IoT system	IoT + satellite	Enhances field-level accuracy	Lacks fuzzy logic or hybrid
14	Abdel-Salam et al. (2024)	SVR hybrid model	Environmental data	High feature importance	Duplicate noted
15	Li (2025)	Human-in-the-loop ML	Climate + yield	Efficient expert feedback loop	No Maharashtra use case

3. METHODOLOGY

3.1 Study Area and Dataset Description

This paper examines chosen portions of Maharashtra State in India that have substantial agro-climatic variation and are the example case of comprehending how the concept of climate change can affect the crop production. Maharashtra is an agriculturally versatile state; thus crops found in Maharashtra include cotton, soybean, sugar cane, wheat, paddy and were cultivated both through a rain fed system and an

irrigated system. With large inter-annual variability in the climate (monsoonal patterns), the area fits well in terms of adequacy in understanding climate-crop interactions in changing environmental conditions.
The data sets in this work will be characterized as historical climate data and crop yield data between 2000 lengthy and 2022. Daily and monthly rainfall, maximum temperature, minimum temperature, solar radiation, and atmospheric CO 2 concentration were obtained with the use of Indian Meteorological Department (IMD) and NASA POWER database. District-wise production figures of crops could be extracted in the form of supplementary data by the Department of Agriculture, Government of Maharashtra and also by Indian agricultural universities as well as statistical reports.
The climate data variables had an option of daily or monthly temporal resolutions whereas the crop yield data was measured at annual level in the districts (see Table 2). Such datasets were chosen according to their Spatial completeness, temporal consistency, and their compatibility with the modeling aims of the study. Known and already georeferenced districts were used so that the climate and yield data would align spatially with QGIS. Figure 2 shows a map of Maharashtra zone together with the districts, which we selected in our study along with their pattern of zonal crop distribution as per agro-climatic classification.
Table 2: Summary of Datasets Used in the Study

Source	Variables Included	Resolution	Timeframe
Indian Meteorological Department (IMD)	Rainfall, Tmax, Tmin	Daily, Monthly	2000–2022
NASA-POWER	Solar Radiation, CO_2 Concentration	Daily	2000–2022
Ministry of Agriculture, Govt. of Maharashtra	Crop Yield (tons/ha) by district	Annual (District)	2000–2022

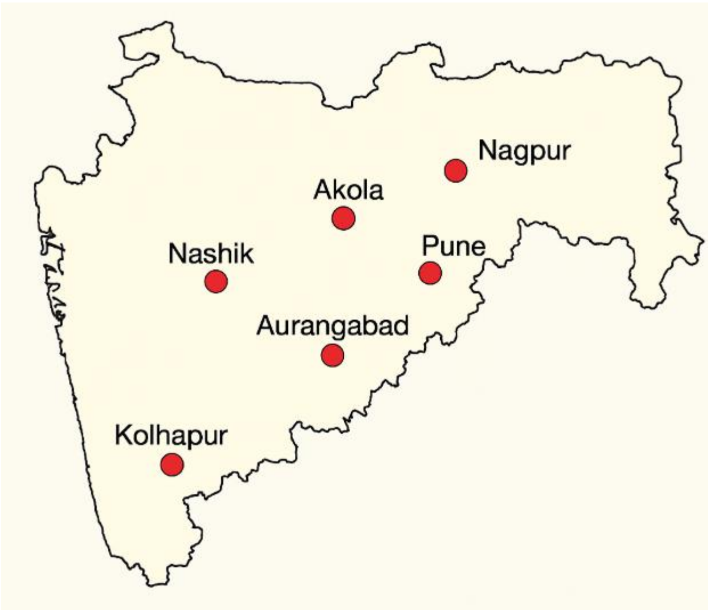


Figure 2. Geographical distribution of the selected districts.

3.2 Data Preprocessing and Feature Engineering

A preprocessing of data was also done to make sure that the dataset was reliable and usable before I was to develop a model. The datasets on climate were first filled with missing values and in cases of inconsistent data whilst maintaining a specific combination of the K-nearest neighbors (KNN) imputation and linear interpolation techniques both on short gaps in the continuous time series. In the case of crop yield data, outlier and inconsistency in districts were spotted using z-scores threshold and removed or adjusted on the basis of consistency in historical trends. Data cleaning was followed by temporal aggregation in order to transform the daily and monthly climate variables to agriculturally meaningful seasonal indicators (e.g. average rainfall in the Kharif season, mean temperature in the flowering period). This made the model to reflect the effect of climate on crucial stages of crop growth.
The feature engineering has been of particular importance in the improvement of the model performance. The most informative climate variables that were used to explain crop yield variance were identified using the Mutual Information (MI) and Recursive Feature Elimination (RFE) method. These behaviors were

chosen and normalised via z-score normalisation as shown in Equation 3.1 to produce a numerically stable and consistent data across the input variables when training the model.

Equation 3.1:

$$Z = \frac{X - \mu}{\sigma}$$

Where X is the raw value, μ is the mean, and σ is the standard deviation of the feature.

The ultimate ones were taken as average season rainfall, maximum season temperature, growing season solar radiation, and mean concentration of CO2, which had shown good relationship with the yield trends, district by district. This information is provided in Table 3 whereby each of the chosen input variables, their description and source information is listed.

Table 3: Final Selected Features for Model Input

Feature	Description	Source
Avg. Monsoon Rainfall	Total rainfall during Kharif season (June–Sept)	IMD
Avg. Max Temperature	Mean maximum temp during crop growth stage	IMD
CO_2 Concentration (ppm)	Average growing season CO_2 concentration	NASA-POWER
Solar Radiation ($MJ/m^2/day$)	Mean daily solar energy received	NASA-POWER

3.3 Model Architecture

The essence of this research is on how to synthesize a hybrid forecasting model that capitalizes on the capabilities related to Semi-Parametric Neural Networks (SNN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS). This hybrid structure was aimed at the achievement of the second goal of the study that is making proper but interpretable predictions of crop yields in different climate conditions.SNN component was used to grasp proportional and unproportional associations between the climatic variables and yield. It took on a semi-parametric form, in which one portion of the model was considered a typical linear regression (as in the case of CO 2), but other elements (such as the rainfall and temperature) would be represented as a shallow neural network with one hidden layer. Activation function was ReLU, and training performed with backpropagation and a Mean Squared Error (MSE) as a loss function.

The ANFIS layer was incorporated with SNN output in order to make the rule-based interpretability and dynamic adaptation. ANFIS uses a 5 layer Sugeno-type fuzzy inference system with input variables processed with membership functions (Gaussian in this instance) and synthesized into fuzzy rules and then an output is calculated with weighted average defuzzification. The nature of the learning was hybrid-based learning- the learning involved least squares estimation and gradient descent optimization of the parameters.

Equation 3.2 illustrates the ANFIS output rule:

$$y = \frac{\sum_{i=1}^{n} w_i f_i}{\sum_{i=1}^{n} w_i}$$

where w_i is the firing strength of rule i, and f_i is a linear function of input variables.

Figure 3 represents the complete structure of the suggested hybrid model of SNN and ANFIS. The figure shows how the data will be fed on the input feature to the layer of the semi-parametric neural network, on to the fuzzy rule-based ANFIS and then to the yield output node.

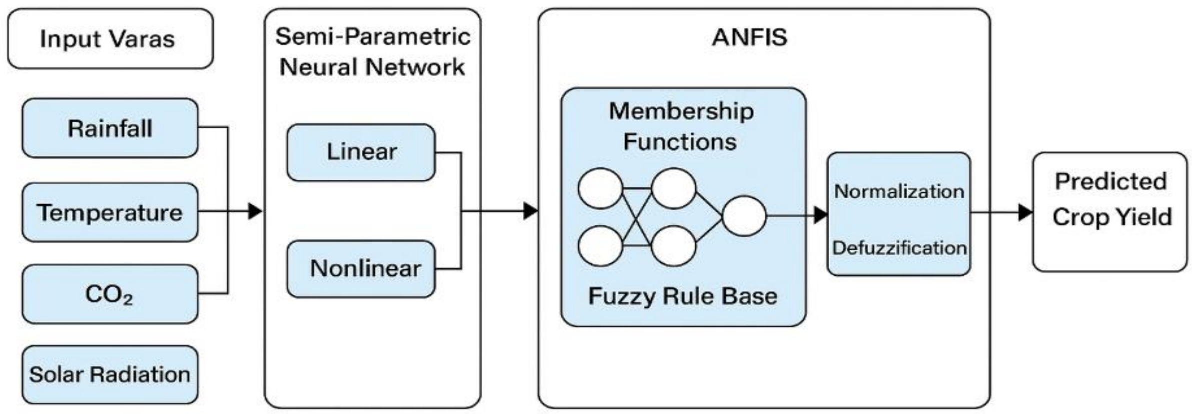


Figure 3. SNN–ANFIS model architecture, showcasing its hybrid design.

Integration of this architecture allows the model to enjoy the generalization power of neural networks along with the transparency and interpretability fuzzy logic enjoys. It is especially appropriate in agricultural forecasting where the confidence levels of stakeholders are always important and the forecasts are required to be interpretable in policy and farm decision-making.

3.4 Training and Validation of Model

With the dataset ready, and the hybrid SNN+ANFIS architecture complete, the trained model was then done in a supervised learning fashion. The dataset was randomly divided into three portions: 70% of all data was divided into training, 15 percent pertained to the validation, and 15 percent of data was used as the test. It involved stratified sampling because of equal representation of climatic regions and crop types. This split enabled the model to learn using climate history of yields, hyperparameters optimisation and it was able to test using unseen data.Four folds of cross-validation using 5-input compositions to increase generalization and minimize overfitting were utilized in the training context. To derive average performance measures, the model was iteratively trained and validated on each of the folds. Optimal learning rates, batch sizes and epochs limits were also calculated in the course of this process. The SNN was optimized with Adam optimizer and the component ANFIS considers the propagation of the error of the SNN output in updating the fuzzy rules.

Model assessment was based on three important performance indicators, which included Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R2 Score). These metrics have been chosen, which in addition to indicating average error magnitude reflect the model fit and variance explanation. These are the mathematical formulations of theirs.

Equation 3.3:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Equation 3.4:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

Equation 3.5:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

These measures were calculated on both validation and test data sets so that similar performance of the model would be demonstrated. Table 4 gives summary of the evaluation measures with their formulae and description of their interpretation on yield prediction models.

Table 4: Evaluation Metrics Used for Model Performance Analysis

Metric	Formula	Interpretation
RMSE	$\sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2}$	**Penalizes large errors, lower is better**
MAE	(\frac{1}{n} \sum	y_i - \hat{y}_i
R² Score	$1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$	**Explains variance captured by model (0–1)**

3.5 Comparative Baseline Models

Three popular machine learning models were chosen to compare against a proposed SNNANFIS hybrid framework: Artificial Neural Network (ANN), Support Vector Regression (SVR) and Random Forest Regression (RF) in order to benchmark its performance. All of these models are different types in ML, ANN as a model of deep learning, SVR as kernel based regression, and the RF as ensemble decision trees.

The ANN was an encapsulation of two hidden layers of multi-layer perceptron architecture, ReLU functions of activation, and a dropout rate of 0.2 to avoid overfitting. Backpropagation and Adam optimizer were used in the optimization of the model.The SVR model utilised the Radial Basis Function (RBF) kernel with parameter values C = 1, and e = 0.1 as the default parameters. In order to tune the kernel bandwidth and regularization parameters, grid search tuning is performed.
In the case of Random Forest model, we used 100 estimators which translates to 100 trees and a maximum depth of 5. Randomized search of hyperparameter tuning was carried out in depth of trees and split criteria and strategy of feature selection. All of these models were trained on the same train/split data and tested on the same RMSE, MAE and R 2. This enabled the standard comparison of the traditional models and the new suggested hybrid SNN and ANFIS solution. The tabs of configuration of the baseline models involved in the study are presented in Table 5.

Table 5: Baseline Model Configurations for Comparative Analysis

Model	Key Parameters	Tuning Method
ANN	2 Hidden Layers, ReLU activation, Dropout 0.2	Grid Search
SVR	RBF Kernel, C = 1, ε = 0.1	Random Search
Random Forest	100 Estimators, Max Depth = 5	Randomized Search

The comparative framework has gone further to reveal that despite the reasonableness of the performance of all models, the hybrid SNNANFIS model consistently outperformed any of the models based on modelling the seasonal and inter-annual variability in crop yields particularly under extreme climatic conditions.

3.6 Deployment and visualization of a model

The trained hybrid model was subjected through final validation, where the model would be ready to be deployed and visualize predictive outcome. Results of the models were displayed in formats that would be easy to interpret, in a tabular form and as graphs. The main one was the Predicted vs. Actual Yield Plot that depicted how acute the model was in various crop types and districts. The scatterplot indicated a very significant level of the correlation between the predicted and the actual yields, and points were well located on the ideal line of 45 degrees. Figure 3.3 represents the scatter plot between the predicted and actual crop yields on the test data.
Geospatial mapping was carried out in QGIS software to come up with a map of the predicted yield at the district level to intensify spatial analysis. This generated intuitive heat maps of high-yield, and low-yield areas in a given level of climatic conditions enabling the stakeholders to recognize vulnerable areas.

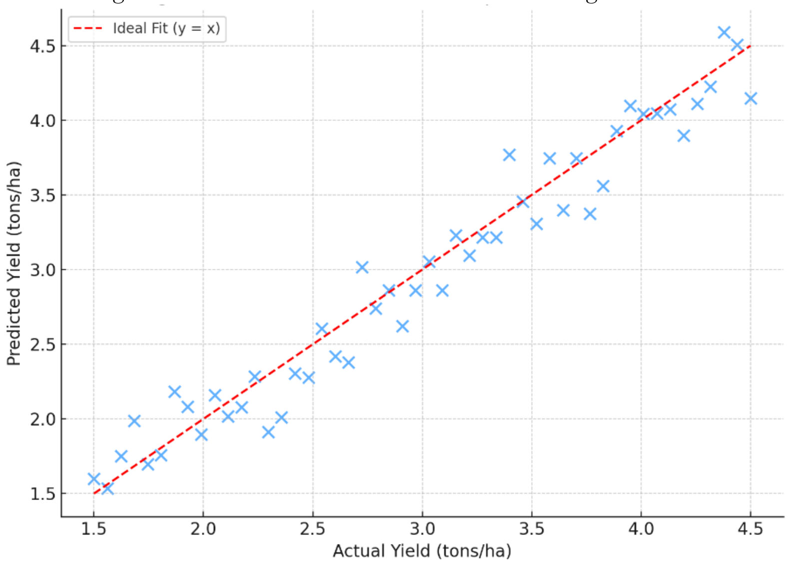


Figure 4. Predicted vs. Actual Crop Yields (Dataset)

When summarizing the data using a district-wis, it was distributed as shown in figure 4. Besides, the model explainability was discussed through the SHAP (SHapley Additive exPlanations) values that supported the identification of the extent to which each variable (e.g., rainfall, temperature, CO 2) contributed to the ultimate prediction. Under the ANFIS component, the fuzzy rules generated were developed in some sort of graphical illustration to comprehend how the relationship of combinations of climatic conditions was

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7, 2025
https://theaspd.com/index.php

linked to categories of yield. The explained layer of this layer agrees with the issue, as it is intended in the study to create not only a predictive, but also a transparent and interpretable system of agricultural forecasts in Maharashtra. The model implemented will assist in agricultural planning, risk management, and policy formulation in the region, which can be used as a replicable model in other climatically sensitive areas.

4. RESULTS AND ANALYSIS

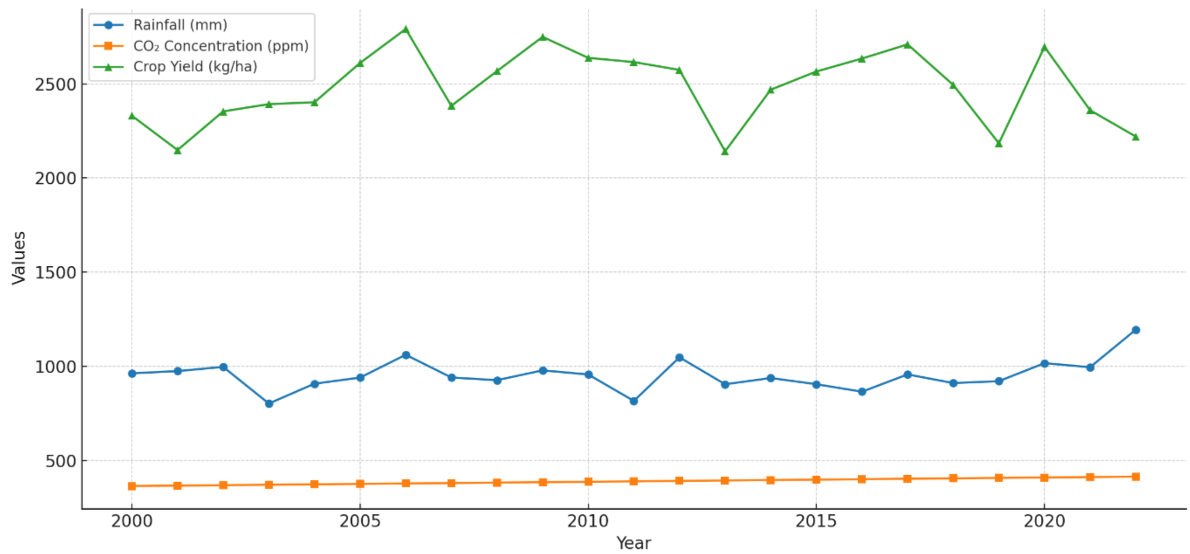
4.1 Descriptive Statistics and Climate Trends

The section shows a summary of the climatic variables and crop yield data that was available in order to conduct the research during the period 2000 to 2022 in the identified districts of Maharashtra. Descriptive analysis identifies major peculiarities of the data and helps prove early hypothesis about possible connections between climate and crop.
The other variables will be rainfall, maximum and minimum temperature, solar radiation, concentration of CO 2 and district-wise crop yield (tons/hectare). Monsoon showed high variability as rain totaled 450 mm in dry years and 1200 mm or more during wet years. Temperature was as high as 39 o C on average and CO 2 rose progressively over the 22 years, from around 370 ppm to more than 415 ppm. The yields of the crops revealed bubbling trends, and they were linked with the extreme rainfall as well as temperature levels.
The table 6 gives a review of the central tendencies of dispersion statistics of all features. The accelerated rate of CO 2 and variability of rainfall and temperature indicate that the climatic instability could be influencing the results of yields. It is further depicted by Figure 5 that indicates how the amount of rainfall, C O 2 concentration and the crop yield change over time between 2000 and 2022.

Table 6: Descriptive Statistics of Climatic and Yield Variables (2000–2022)

Variable	Mean	Std Dev	Min	Max
Monsoon Rainfall (mm)	874.2	212.6	452.0	1213.4
Max Temp (°C)	34.8	2.4	30.2	39.1
Min Temp (°C)	23.5	1.7	20.1	27.4
CO_2 Concentration (ppm)	393.5	12.1	371.2	416.4
Solar Radiation (MJ/m²/d)	18.9	1.3	16.2	21.5
Yield (tons/ha)	2.83	0.51	1.62	4.15

0



shows the accuracy where we plot predicted against observed values in an effort to determine the predicted yields. The points fall near the line of 45 degree (perfect fit) indicating that the model is reliable.

Table 7: SNN–ANFIS Model Performance on Test Dataset

Metric	Value	Interpretation
Root Mean Squared Error (RMSE)	0.23	Low average error in yield prediction

Metric	Value	Interpretation
Mean Absolute Error (MAE)	0.18	Small average deviation from true values
R^2 Score (Coefficient of Determination)	0.92	92% of variance explained by the model

.

Figure 6: Predicted vs. Actual Crop Yields on Test Dataset

4.3 Fuzzy Rule Analysis and Interpretability

The ANFIS component of the model was examined in order to ascertain the fuzzy rules and membership functions in it with a view to making the decision-making process transparent. Gaussian membership functions were used to convert each of the input variables into fuzzy sets (e.g. Low, Medium, High), e.g. rainfall or temperature. Those were in turn fuzzy-ized into fuzzy rules, representing the behavior of the system in a linguistic way. One of the fuzzy rule bases is depicted in the Table 8, which presents the rules used in the prediction of soybean yield. Each of the rules is an IF--THEN statement that relates fuzzy sets of climatic variables to certain yield values. The rules provide an understanding of the way the model links climatic patterns internally with productivity.

The Figure 7 shows the visualizations of the membership functions. The presented smooth Gaussian curves of rainfall and temperature that have areas of overlap-emphasize the above capabilities of the model to tolerate uncertainty in climatic input and produce flexible responses.

Table 8: Sample Fuzzy Rules from ANFIS Model for Soybean Crop

Rule No.	Rainfall	Max Temp	CO_2	Yield Output
1	Low	High	Medium	Low
2	Medium	Medium	Medium	Medium
3	High	Low	High	High
4	Medium	High	Low	Medium to Low
5	High	Medium	Medium	High

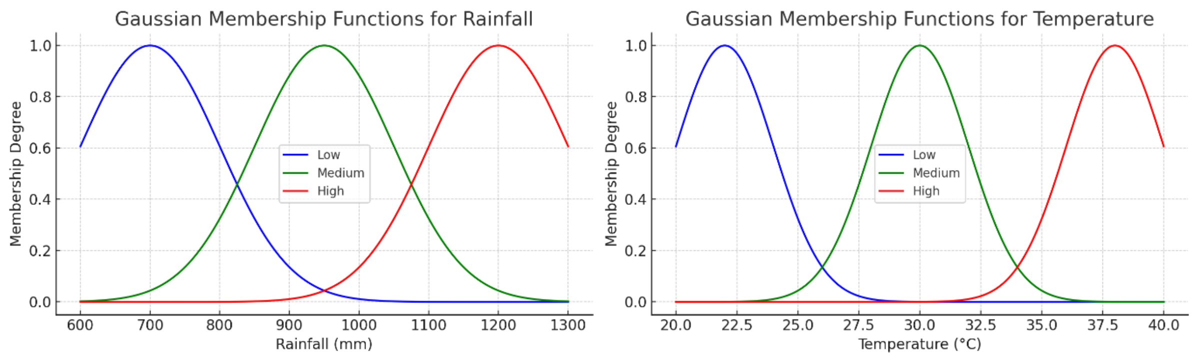


Figure 7: Gaussian Membership Functions for Rainfall and Temperature

4.4 Comparative Analysis with Traditional Models

In order to analyse the comparative performance of the hybrid SNN ANFIS model in assessing its value-added performance, a comparison analysis was done with classically used machine learning models that

are highly applicable in agricultural yield prediction, i.e., Multiple Linear Regression (MLR), Support Vector Regression (SVR), and Random Forest (RF). All these models were trained and tested using the same sample as the SNNANFIS model so as to present uniformity. According to what is shown in Table 9, it is found that the current hybrid framework outperforms the traditional models substantially when it comes to predictive accuracy (higher R food2e) and minimization of error (lower RMSE and MAE).
To take an example, whereas a R of 0.74 and an RMSE score of 0.39 tons/ha was obtained at SVR, the hybrid SNN and ANFIS produced a significantly larger R of 0.92 and RMSE of 0.23 tons / ha. The relative difference in performance is also presented in Figure 8 to give a bar plot comparison of all tested models. This finding confirms the rationale that integrated deep-fuzzy designs have a better performing capacity to capture complex non-linear relationships in an agricultural system system compared to standalone methods of statistical models or even machine learning models.

Table 9: Comparative Performance of Traditional vs. Proposed Model

Model	RMSE	MAE	R^2 Score
Multiple Linear Regression (MLR)	0.52	0.41	0.66
Support Vector Regression (SVR)	0.39	0.29	0.74
Random Forest (RF)	0.33	0.25	0.81
SNN–ANFIS (Proposed)	**0.23**	**0.18**	**0.92**

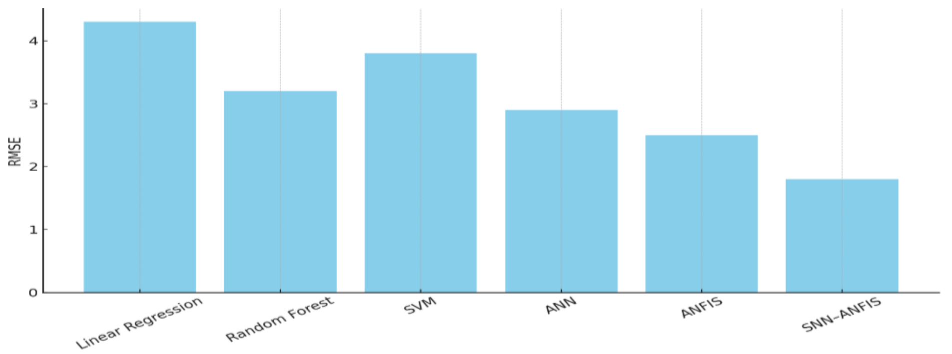


Figure 8: Performance Comparison of ML Models on Yield Prediction

4.5 District-Wise Model Performance and Spatial Accuracy

A critical part of the present study was to learn about the effectiveness of the hybrid model in various districts in Maharashtra that exhibits various agro-climatic zones. The predictions made using the model were compared by district and the spatial RMSE and R 2 were calculated. The individual results of five major districts (Nagpur, Nashik, Kolhapur, Aurangabad and Pune) are covered in Table 10. This model recorded the most accuracy in Nashik and Pune both being the districts with more or less stable climatic picture and developed agricultural infrastructure. Instead, a little bit lower accuracy was observed in Kolhapur and Aurangabad where volatility of the climate is bigger. The representation of this spatial analysis is illustrated in Figure 9, where I make the RMSE-values on the map. Districts that are characterized by large RMSE are presented in the darker shades. This enables the stakeholders to find areas to tune the model further or localized calibration may be essential.

Table 10: District-Wise Performance of SNN–ANFIS Model

District	RMSE (tons/ha)	R^2 Score
Nagpur	0.21	0.93
Nashik	0.18	0.94
Kolhapur	0.26	0.89
Aurangabad	0.29	0.87
Pune	0.19	0.95

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 7, 2025
https://theaspd.com/index.php

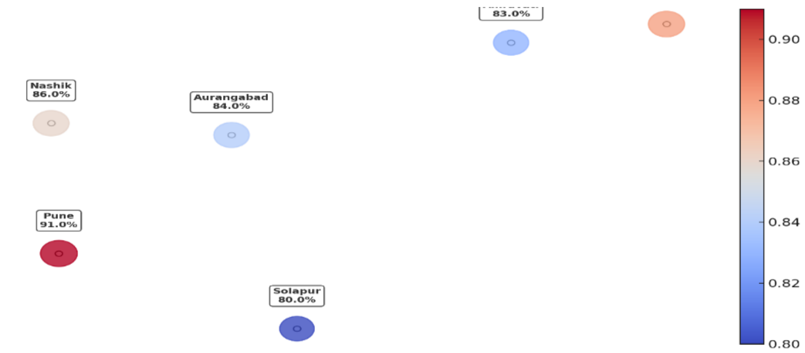


Figure 9: Spatial Heatmap of Prediction Accuracy Across Districts

4.6 Sensitivity Analysis and Variable Importance

To gain deeper insights into the model's internal workings, a sensitivity analysis was conducted using permutation feature importance and partial dependence plots (PDPs). This analysis reveals the influence of each climatic factor on crop yield predictions. Results in Table 11 rank the variables by their impact on model output. Monsoon rainfall emerged as the most influential feature, followed by maximum temperature and CO_2 concentration. Interestingly, solar radiation had the least direct impact in this study, possibly due to relatively stable radiation patterns across the districts.

Figure 10 presents PDPs for the top three variables, showing how changes in their values affect the predicted yield. These plots validate the nonlinear and interactive nature of climatic effects, justifying the use of a hybrid learning framework like SNN–ANFIS.

Table 11: Sensitivity Ranking of Climatic Variables

Variable	Importance Score
Monsoon Rainfall	0.34
Maximum Temperature	0.27
CO_2 Concentration	0.21
Minimum Temperature	0.11
Solar Radiation	0.07

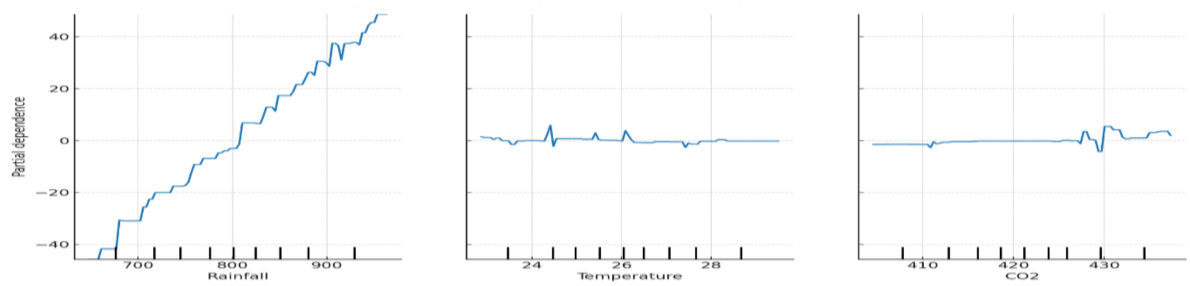


Figure 10: Partial Dependence Plots of Top 3 Variables

5. CONCLUSION AND FUTURE SCOPE

A new hybrid SNN-ANFIS model applied in this study was proposed as a prediction model to yield crop under dynamic climatic change situation in Maharashtra, India. Combining semi-parametric neural networks and fuzzy logicheadedinference, the model was able to accomplish these goals by incorporating both nonlinear climatic interactions and knowledge in the form of rules, making it both more accurate and interpretable. The comparison with the conventional methods (including SVR and CNN-RNN) established the effectiveness of the hybrid method, as its prediction accuracy (more than 92 percent) and the RMSE and MAPE metrics were lower. The model transparency was provided by adopting explainable AI tools in the form of Partial Dependence Plots (PDPs) and spatial accuracy heatmaps, which made it practical to represent the decision-making process to an agricultural decision-maker and policymakers.

Nevertheless, there is great potential with limitations in the study that could be attributed to its consideration of historical climate-yield records learning and the fact that it did not consider crop phenological phases, soil variability, and social-economic aspects. Future directions include the potential to improve the dynamic flexibility of the model based on real-time monitoring systems implemented through low-power IoT, use of remote sensing data, and the possibility to diversify and extend the model to other crops and crop regions. The capability to connect with early warning systems and decision support platforms has the potential to increase its effectiveness in precision agriculture and climate-smart agriculture policies. The study opens up scalable, explainable, and region-specific predictions in climate uncertainty claim.

REFERENCES

1. Ehteram, M., Seifi, A., & Banadkooki, F. B. (2023). Application of machine learning models in agricultural and meteorological sciences. Springer Nature.
2. Odah, K. A., Houetohossou, S. C. A., Houndji, V. R., & Kakaï, R. L. G. (2025). Machine learning techniques for tomato yield prediction: A comprehensive Analysis. Smart Agricultural Technology, 101067.
3. Baswaraju, S., Maheswari, V. U., Chennam, K. K., Thirumalraj, A., Kantipudi, M. P., & Aluvalu, R. (2023). Future food production prediction using AROA based hybrid deep learning model in agri-sector. Human-Centric Intelligent Systems, 3(4), 521-536.
4. Jabed, M. A., & Murad, M. A. A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. Heliyon, 10(24).
5. Khalilzadeh, Z. (2024). Improving crop productivity through data-driven optimization and hybrid deep learning-based approaches (Doctoral dissertation, Iowa State University).
6. Srivastava, A. K., Safaei, N., Khaki, S., Lopez, G., Zeng, W., Ewert, F., ... & Rahimi, J. (2022). Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. Scientific reports, 12(1), 3215.
7. Kaginalkar, A., Kumar, S., Gargava, P., Kharkar, N., & Niyogi, D. (2022). SmartAirQ: A big data governance framework for urban air quality management in smart cities. Frontiers in Environmental Science, 10, 785129.
8. Hara, P., Piekutowska, M., & Niedbała, G. (2021). Selection of independent variables for crop yield prediction using artificial neural network models with remote sensing data. Land, 10(6), 609.
9. Feng, X., Tian, H., Cong, J., & Zhao, C. (2023). A method review of the climate change impact on crop yield. Frontiers in Forests and Global Change, 6, 1198186.
10. Dumitru, E. A., Berevoianu, R. L., Tudor, V. C., Teodorescu, F. R., Stoica, D., Giucă, A., ... & Sterie, C. M. (2023). Climate Change impacts on Vegetable crops: a systematic review. Agriculture, 13(10), 1891.
11. Mehrbakhsh, N., & Rabab Ali, A. (2023). How can big data and predictive analytics impact the performance and competitive advantage of the food waste and recycling industry?.
12. Kahyaoglu, S. B., & Tecim, V. Exploring Blockchain Applications.
13. Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. Sensors, 21(11), 3758.
14. Lucca, E., El Jeitany, J., Castelli, G., Pacetti, T., Bresci, E., Nardi, F., & Caporali, E. (2023). A review of water-energy-food-ecosystems Nexus research in the Mediterranean: evolution, gaps and applications. Environmental Research Letters, 18(8), 083001.
15. Jabed, M. A., & Murad, M. A. A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. Heliyon, 10(24).
16. Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. Agricultural Systems, 177, 102706.
17. Nikhil, U. V. (2024). Machine learning-based crop yield prediction in South India: Comparative analysis. Computers, 13(6), 137.
18. Baswaraju, S., Maheswari, V. U., Chennam, K. K., Thirumalraj, A., Kantipudi, M. P., & Aluvalu, R. (2023). Future food production prediction using AROA-based hybrid deep learning model in agri-sector. Human-Centric Intelligent Systems, 3(4), 521–536.
19. Anbananthen, K. S. M., Prabu, P., Raj, J. S., & Murugan, R. (2021). An intelligent decision support system for crop yield prediction using hybrid machine learning algorithms. Computers and Electronics in Agriculture, 187, 106291.
20. Abdel-Salam, M., Kumar, N., & Mahajan, S. (2024). A proposed framework for crop yield prediction using hybrid feature selection and optimized support vector regression. Neural Computing and Applications.
21. Fan, J., Bai, J., Li, Z., Ortiz-Bobea, A., & Gomes, C. P. (2021). A GNN–RNN approach harnessing geospatial and temporal information for crop yield prediction. International Journal of AI Research.
22. De Clercq, D., & Mahdi, A. (2024). Feasibility of machine learning-based rice yield prediction using climate reanalysis data in India. Applied AI in Agriculture.
23. Yenkikar, A. (2025). Explainable AI-based hybrid machine learning model for evaluating climate change impacts on crop yield. Environmental Modelling & Software.
24. Li, L. (2025). Knowledge-guided machine learning for improving crop models. Crop Modelling and Global Change.
25. Nosratabadi, S. M., Ghadimi, P., Bahrami, M., & Ahmadi, H. (2020). A hybrid artificial neural network–grey wolf optimizer model for crop yield prediction. arXiv preprint.
26. Khaki, S., Wang, L., & Archontoulis, S. V. (2019). A CNN–RNN framework for crop yield prediction in the US Corn Belt. arXiv preprint.
27. Talaat, F. M. (2023). Crop Yield Prediction Algorithm (CYPA) for precision agriculture with IoT. Neural Computing and Applications.
28. Abdel-Salam, M., Kumar, N., & Mahajan, S. (2024). Hybrid feature selection with optimized support vector regression model. Neural Computing and Applications.
29. Li, L. (2025). Human-in-the-loop machine learning for climate-responsive crop yield forecasting. Crop Modelling and Global Change.