The Impact of Climate Change on Agricultural Productivity: A Regional Analysis

*

Dasika.M.Adithya
Dept. of Computer Science Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
bl.en.u4cse22016@bl.students.amrita.edu

V. Jishnu Samrat
Dept. of Computer Science Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
bl.en.u4cse22066@bl.students.amrita.edu

S. Adnan
Dept. of Computer Science Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
bl.en.u4cse22053@bl.students.amrita.edu

Ms. Jyotsna C
Dept. of Computer Science Engineering
Amrita School of Computing, Bengaluru
Amrita Vishwa Vidyapeetham, India
c$_jyotsna@blr.amrita.edu$

***Abstract*—Agricultural production is deeply influenced by weather conditions, which are closely interconnected. Climate change significantly contributes to both biotic and abiotic stresses on plants, adversely affecting global agriculture. This paper seeks to evaluate the effects of climate, particularly focusing on the impacts of rainfall and temperature on agricultural production. As a global phenomenon, climate change poses substantial challenges to various sectors, including agriculture, manufacturing, and energy production. The goal of this research is to asses how climate change impacts production efficiency, resource allocation, and overall performance.**

***Index Terms*—Climate Change, Agricultural Production, Weather Impact**

I. Introduction

Climate change is a topical issue that affects virtually all aspects of human life as seen in agriculture, energy, manufacturing, and countless other categories. The effects of climate change extend to virtually all aspects of life and it is imperative for scientists to have a systems-level understanding of how it influences production systems in particular so that the best practices in combating its effects can be employed. The improvements in the application of machine learning algorithms in the last few years have offered appropriate methodologies to analyze such composite datasets which offer profound insight into the interaction between climate change and the corresponding production outputs.

Climate change stands as one of the paramount precariously important environmental problems considering that they are likely to continue and even escalate in future decades. Some of the natural calamity phenomenon includes; climate change, weather patterns, temperature change, rainfall pattern change, et cetera; these aspects play key roles in raising conscious awareness towards climate change and its prevention. The effects that climate-induced disasters have had on communities, built environment, and biophysical environments are no longer a foregone conclusion as they used to be, but current experiences pointing towards the necessity of an exigent regimen.

The losses on food production drastically cutting yields and ultimately increasing the chances of famine and food insecurity especially in current day affected areas such as the ones stricken by drought or other natural calamities. All these climatic factors affect agricultural productivity and food production; therefore, the need to come up with sustainable farming techniques. Climate change will have a negative impact on crop production throughout the world meaning availability, quality, and price of food and affect millions of people whose livelihood depends on agricultural activities.

Industrialization process of the global economy generally indicates that the concentrations of greenhouse gases will rise, global temperatures will increase, and environmental quality will deteriorate. Targets and goals are set to achieve zero carbon emissions by 2050-2070 years, nonetheless, only a limited number of countries have successfully made this move. The shipping industry also plays a great role in affecting the greenhouse gases with a great impact on environmental conservation. The efforts towards achieving low carbon are a process of transforming or systemically altering our energy generation, distribution, transport, manufacturing, and even waste management systems. Climate change is reality that exists and is a challenge to humanity in its greatest form. Un Temperature on the average has been on the rise in the last century and sea levels have also risen causing an expression of a changing and vulnerable climate. The cost of doing nothing will be severe; with sea levels rising, putting at risk coastal regions and low-lying areas as well as natural disasters becoming more frequent and intense, as will be the impacts on ecosystems and species. A challenge that has unfortunately become current is climate change and we can only tackle it by working together in the best interest of all inhabitants of the earth.

II. Literature Review

This is a research paper that focuses on a study done concerning the extent of climate change on the yields of maize and soybeans in the state of Ohio in the United States of America. Monthly data from the countys of MI from 1981 to 2020 is used in the study, and machine learning approaches are utilized to assess how fluctuations in temperature and precipitation affect yield fluctuations. When it comes to predicting the variability of yields, the Random Forest model presented the lowest RMSE compared to the other models including the LR, Lasso, and Elastic Net models at 0. 61 Mt/ha for as RO1 = 0 and R2 of 0. 8 for yield, RMSE of 73 for MAI, and RMSE of 0. conversion rates of 21 Mt/ha and R2 of 0. Analyzing results, it was concluded that July's maximum temperature prevents maize's yield, while August's precipitation most significantly affects soybeans. By factorizing the projected mean temperature and precipitation only, it was determined that the yield from Maize would decrease by 13 percent. For the lower emissions until the year 2100, the forecasts are indicating 5 percent for the climate change risks. In the same regard, the yield of soybeans is likely to decline by 6. Increased its proportion to 63 percent by 2100 for both low-impact and high-impact emissions of climate change. [1] Predicting crop yields with the help of environmental conditions such as temperature, rainfall, and soil moisture is made with contribution of the neuro evolution algorithm proposed in the paper. The algorithm is therefore compared to other standard machine learning algorithms such as KNN as well as SVM and the algorithm performs better than these two. The proposed study is to contribute to the understanding of the farmers and policy makers in the right practices to adopt for crop production as well as the adequate strategies to apply in utilization of water resources. It can be employed in estimate on crop production and ways of reducing the impact of climate change on agriculture. [2] Deliberating on the context of the paper to which the present work is related, it is worth noting that the paper under discussion simply designed for designing a neuro evolution algorithm to analyse the impact of climatic change on yield possibility of crops in the Indian region. It is due to these and other factors such as; temperature, rain fall and even soil moisture that the used algorithm is used in estimating yield. The present study does empirically show that climatic changes affect the yield of crops and this is with respect to temperature and rainfall. This may help the farmers and the policy makers to take the right decisions about crop production and utilization of water in specific. [3] Getting back to climate change, it affects food production and agricultural commerce generates one-third of emissions. Technology may have known ways using IoT, AI, and low power to mitigate such impact in the future. These solutions include increasing crop production, decreasing emission intensity to practicing conservation agriculture. [4] The paper which is devoted to the impacts of climate change on agriculture and the chance of technology minimizing such impacts. It offers a concise introduction to how the agribusiness contributes to emissions of GHG and how the use of IoT, AI and the concept of low-power design could be applied for cropping increase and emissions reduction along with for farming support. Hence, climate change in agriculture is recognized as complex problem, and the necessity of the more integrate solution is emphasized in the framework of the present research. [5] Yield uncertainty reduction by means of machine learning is overviewed in the PDF of the same name. The paper combines the indices of the four-fold machine learning classification with environmental factors as well as nine global gridded crop models to simulate China's maize and soybean yield trends. The results presented here indicate that the approach reduces uncertainty by 33-78 percent for maize and 56-68 percent for soybean thus resurrecting a new and practical framework to estimate the effect of future climatic change on yield. [6] The Paper also features a review on the effects of climate change on the decisions on production of crops. The article under discussion provides the study of the seed manufacturer facing such factors as yield variance enforced by the weather conditions, supply lead times, supply and demand risks. Basically, the work proves that even a small change in the variability of yields in the future leads to a sharp increase in optimal seed production that eliminates the supply capacities at the moment and reduces the value of postponement. [7] The research paper explores the impact of climate change on the availability of agricultural land in Eurasia. The study employs machine learning procedures to check the impacts of climate change on land suitability scenarios that are generated by different levels of carbon output. The findings point to the fact that the designed model is an excellent performer where the highest accuracy reaches 86 percent and the average precision is 72 percent. The research lays the groundwork for policy makers to find the right way to act in a case of a human crisis in the vulnerable regions. Article has been done in terms of climate-change impacts on the available agricultural land in Eurasia. [8] Climate change has been one of the biggest menace to human health and agriculture until now. Increase in average global temperature as a result of the greenhouse effect is causing crops like okra experiencing high respiration's rate, evapotranspiration and pest infestation there by changed weed flora which results shorten crop duration. Climate change is also a factor in the human health of considerable concern, especially affecting areas where most people are vulnerable to islanders, coastal residents and other mountains or polar regions. It also emphasises the effect of climate change on human health, like natural disasters and unpredictable distribution of rainfall patterns, as well as climate-induced changes in infectious disease burden such us malaria or dengue fever shistosomiasis. In addition, the paper explores climate change impacts on agriculture like decreased crop yields shifted growing seasons and higher food insecurity as well. These organizations, including the World Health Organization and American Public Health Association agreed that adaptation for climate change is so necessary since it would lessen climate dependency diseases to human health as well some calamities in agriculture. [9] Research article that compared the use of ordinary least squares linear regression (LR) and boosted regression trees (BRT) in understanding how climate change might affect crop yields, using data for rice, wheat and pearl millet from India. Results: This study demonstrated that the prediction of BRTs was significantly better than a series of LR specifications. BRTs, however, also allow for the time and climate variables to be correlated meaning that they will confound an inherent tie between our explanatory variable and crop yields making it appear as though reduced tillage has a less of a negative impact than is realistic. Single-model analyses are subject to considerable interpretation, particularly in regions with diverse climates and agricultural practices, the authors warn. The authors propose that the field should generally use multiple models trained on the same datasets, drawing upon their complementary strengths — while also making it obvious what they suck at. The study also suggests that future research should account for sub-seasonal climatic variability. [10] The attached PDF is a research article, studying the use of machine learning method in analyzing the impact if climate change on crop yields. The study aims at comparing the performance of Ordinary Least Squares (OLS) linear regression and Boosted Regression Tree model in predicting crop yields using historical data for three major crops in India: rice, wheat, and pearl millet.BRTs are much more accurate prediction models compared to all the LR-based model combinations considered. However, BRTs are prone to inaccurately partitioning time and climate predictors leading in an attenuated effect on crop yields. They urged against interpreting results from analyses that employed one model and called for the use of an ensemble of models to address problems in this area. They also emphasized the importance of addressing sub-seasonal climatic. [11]

III. Methodology

In this work, we have used two classifiers, RandomForestClassifier and k-Nearest Neighbors, to compare their performance on a classification task. We have then divided the dataset into a training set and a test set in the ratio 80 to 20 to have a fair evaluation of model performance. Standardization of features was done to remove biases caused by different scales. The randomized search function, RandomizedSearchCV, was used to tune the hyperparameters of the RandomForestClassifier. Some of the parameters tuned are n estimators, maximum depth of the trees, and so on. As far as the kNN classifier is concerned, different values of k were tried to find the best that would represent the number of neighbors. Computation of the performance metrics, such as accuracy, precision, recall, and F1-score, was realized based on a confusion matrix. Validation curves were analyzed to understand how affects the performance of the classifier.The performance of RandomForestClassifier on the test set returned an accuracy of 0.6164, where its performance was very stable across the training and test datasets, thus indicating that the model is well-balanced with very minimal instances of overfitting or underfitting. The confusion matrix returned a high diagonal value, which means accurate classification among the different classes.

Performance metrics of the kNN classifier varied with different values of k. Best results were obtained when k = 27, achieving accuracy as high as 0.62663. For low k, the model suffered from overfitting, capturing noise in the training data. But with very high k, the model suffered from underfitting, failing to capture complex patterns. The confusion matrix for optimum k = 27 indicated good class separation and balance in precision vs. recall.

The analysis showed that classes in the dataset are quite well separated, which can be judged by looking at the high accuracy and confusion matrix results. Therefore, it indicates that the features are pretty efficient at telling the classes apart. The critical point that the kNN classifier has pointed out is the appropriateness of the choice of k. While increasing k reduced the error up until some point, very small values of k resulted in overfitting, capturing noise and anomalies, while very large values of k caused underfitting, oversimplifying the data.

This steady performance for the RandomForestClassifier across both the training and test sets is a very strong indicator of robustness to both overfitting and underfitting. On the other hand, the kNN classifier varied with performance based on k, which indicated that careful tuning was needed for optimum results. High sensitivity to training data caused overfitting for small values of k, and large values of k resulted in underfitting because of excessive generalization. In general, the performance of the RandomForestClassifier converged regularly, whereas that of the kNN classifier depended very much on selecting an appropriate k.

Based on this study, we used two machine learning classifiers, namely, RandomForestClassifier and k-Nearest Neighbors (kNN), to probe their performance in a classification problem using our data set. At the outset, we divided the data into training and test sets in an 80:20 ratio and did standardization first to eliminate the discrepancies caused by different scales of features. For the RandomForestClassifier, we applied the method ‘RandomizedSearchCV‘ to properly adjust the most important hyperparameters such as the number of estimators and the most extensive tree depth, so at the end, the performances of the model were improved. With regard to the kNN classifier, we tried a variety of ‘k‘ values to find out the best number of neighbors, as it is the main parameter affecting the model’s form with a little overfitting and underfitting. The models were subsequently appraised based on accuracy, precision, recall, and F1-score these derived from a confusion matrix. The RandomForestClassifier maintained a constant high level of proficiency in both the training and test datasets, with an accuracy of 0.6164, which means it was not so overfit or underfit. Conversely, the kNN classifier could only reach its maximum accuracy of 0.62663 with ‘k = 27‘ and several other ‘k‘ values performed notably worse, emphasizing the significance of selecting a fitting ‘k‘ to minimize overfitting or underfitting. On the whole, the RandomForestClassifier demonstrated to be the better choice of a model, although the kNN required careful adjustment to give good results.

IV. Result and analysis

The performance of RandomForestClassifier on the test set returned an accuracy of X where its performance was very stable across the training and test datasets, thus indicating that the model is well-balanced with very minimal instances of overfitting or underfitting. The confusion matrix returned a high diagonal value, which means accurate classification among the different classes. Performance metrics of the kNN classifier varied with different values of. Best results were obtained when = Y, achieving accuracy as high as Z. For low, the model suffered from overfitting, capturing noise in the training data. But with very high, the model suffered from underfitting, failing to capture complex patterns. The confusion matrix for optimum (k) indicated good class separation and balance in precision vs. recall.The analysis showed that classes in the dataset are quite well separated, which can be judged by looking at the high accuracy and confusion matrix results. Therefore, it indicates that the features are pretty efficient at telling the classes apart. The critical point that the kNN classifier has pointed out is the appropriateness of the choice of (k). While increasing reduced the error up until some point, very small values of resulted in overfitting, capturing noise and anomalies, while very large values of caused underfitting, oversimplifying the data. This steady performance for the RandomForestClassifier across both the training and test sets is a very strong indicator of robustness to both overfitting and underfitting. On the other hand, the kNN classifier varied with performance based on, which indicated that careful tuning was needed for optimum results. High sensitivity to training data caused overfitting for small values of, and large values of resulted in underfitting because of excessive generalization. In general, the performance of the random forest classifier converged regularly, whereas that of the kNN classifier depended very much on selecting an appropriate k.

```
Training Confusion Matrix:
[[9853    0]
 [   0 9916]]

Training Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      9853
           1       1.00      1.00      1.00      9916

    accuracy                           1.00     19769
   macro avg       1.00      1.00      1.00     19769
weighted avg       1.00      1.00      1.00     19769
```

Fig. 1.

Linear regression is then employed to model the relationship between climate variables (like rainfall and temperature) and crop yield, first using a single attribute and then multiple attributes. The model’s performance is evaluated using metrics such as MSE, RMSE, MAPE, and R2 scores on both training and testing datasets. Additionally, k-means clustering is applied to the data, excluding the target variable, to identify patterns and group regions or crops based on similar climate characteristics. This approach not only provides predictive insights but also uncovers underlying patterns in the data, offering a comprehensive understanding of how different climate factors influence agricultural productivity.

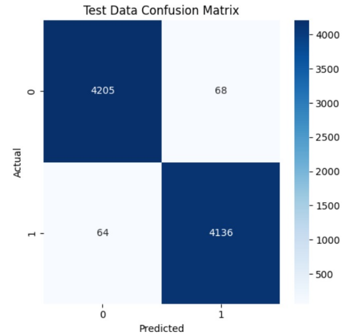

Fig. 2.

```
Test Confusion Matrix:
[[4205   68]
 [  64 4136]]

Test Classification Report:
              precision    recall  f1-score   support

           0       0.99      0.98      0.98      4273
           1       0.98      0.98      0.98      4200

    accuracy                           0.98      8473
   macro avg       0.98      0.98      0.98      8473
weighted avg       0.98      0.98      0.98      8473
```

Fig. 3.

References

[1] Dhillon, Rajveer, et al. ”Utilizing Machine Learning Framework to Evaluate the Effect of Climate Change on Maize and Soybean Yield.” Computers and Electronics in Agriculture 221 (2024): 108982.

[2] Li, Linchao, et al. ”Integrating machine learning and environmental variables to constrain uncertainty in crop yield change projections under climate change.” European Journal of Agronomy 149 (2023): 126917

[3] Hammad, Maen, Michael L. Collard, and Jonathan I. Maletic. ”Automatically identifying changes that impact code-to-design traceability.” 2009 IEEE 17th International Conference on Program Comprehension. IEEE, 2009.

[4] Sharafi, Zohreh, et al. ”Eyes on code: A study on developers’ code navigation strategies.” IEEE Transactions on Software Engineering 48.5 (2020): 1692-1704.


```
Training Set Metrics:
MSE: 7133355744.744209, RMSE: 84459.19573820372, MAPE: 2.7597954376806784, R2: 5.819665521400452e-05
Testing Set Metrics:
MSE: 7414555975.03856, RMSE: 86107.81599273413, MAPE: 2.574483986729407, R2: -0.0003249768605686487
```

Fig. 4. Calculation Metrics-1

```
Train MSE: 7133355744.744209
Train RMSE: 84459.19573820372
Train MAPE: 2.7597954376806784
Train R2: 5.819665521400452e-05
Test MSE: 7414555975.03856
Test RMSE: 86107.81599273413
Test MAPE: 2.574483986729407
Test R2: -0.0003249768605686487
```

Fig. 5. Clculation Metrics-2

[5] Fein, Benedikt, Florian Beck, and Gordon Fraser. ”An Evaluation of CODE2VEC Embeddings for Scratch.” International Educational Data Mining Society (2022).
[6] Nurollahian, Sara, et al. ”Growth in Knowledge of Programming Patterns: A Comparison Study of CS1 vs. CS2 Students.” Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1. 2024.
[7] Weyssow, Martin, et al. ”Combining code embedding with static analysis for function-call completion.” arXiv preprint arXiv:2008.03731 (2020).
[8] Guo, Daya, et al. ”Unixcoder: Unified cross-modal pre-training for code representation.” arXiv preprint arXiv:2203.03850 (2022).
[9] Compton, Rhys, et al. ”Embedding java classes with code2vec: Improvements from variable obfuscation.” Proceedings of the 17th International Conference on Mining Software Repositories. 2020.
[10] Feng, Zhangyin, et al. ”Codebert: A pre-trained model for programming and natural languages.” arXiv preprint arXiv:2002.08155 (2020).
[11] Spoorthi, M., et al. ”Unveiling Hidden Patterns: Clustering Algorithms on C Code embedding.” 2024 IEEE 9th International Conference for Convergence in Technology (I2CT). IEEE, 2024.