International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

A Comprehensive Dataset for Crop Yield Analysis in Telangana Region

P. Sowmya[1], Dr. A. V. Krishna Prasad[2]

[1]Telangana Mahila viswavidyalayam, Koti, Hyd,
Email: *soumya.padam[at]gmail.com*

[2]Maturi Venkata Subba Rao Engineering college, Nadergul, Hyd.
Email: *kpvambati[at]gmail.com*

Abstract: ***This paper presents a meticulously curated and extensive dataset tailored for in-depth analysis of crop yields in the Telangana region, India. Spanning a period of [1966-2020], the dataset comprises 496 records, each meticulously compiled from a variety of reputable sources. The dataset encapsulates a diverse array of agricultural and meteorological features, including Dist. Code, Year, State Code, Dist. Name, Rice Irrigated Area (1000 ha), Total Consumption (tons), Annual Rainfall (Millimeters), Precipitation Avg, Max Temp, Min Temp, Wind Speed, Wind Direction, Surface Soil Wetness, Profile Soil Moisture, and Root Zone Soil Wetness. The importance of this dataset lies in its potential to shed light on critical determinants of crop yield in this agriculturally vibrant region. By providing a comprehensive record of environmental conditions, agronomic practices, and yield outcomes, researchers, agronomists, and policymakers gain a valuable tool for data-driven decision-making in agricultural planning and resource allocation. The dataset's temporal and geographic coverage facilitates not only cross-sectional analyses but also the exploration of temporal trends and variations in crop yield. As climate change increasingly impacts agricultural productivity, this dataset offers a vital resource for studying the resilience and adaptability of crops to evolving environmental conditions. With its rich features and robust data collection methodology, this dataset serves as a foundational resource for developing predictive models, optimizing resource allocation, and informing policies aimed at enhancing agricultural productivity and sustainability in the Telangana region.***

Keywords: crop yield, climate, features, dataset.

1. Introduction

The Telangana region of India boasts a dynamic agricultural landscape characterized by diverse agro-climatic zones, a wide array of crops, and varying farming practices. Understanding the intricate interplay between environmental factors, agronomic practices, and crop yields is of paramount importance for sustainable agricultural development in this region. To this end, this paper introduces a comprehensive dataset meticulously curated to facilitate a nuanced analysis of crop yields in Telangana.

1.1 Significance of Crop Yield Analysis

Crop yield analysis serves as the cornerstone of effective agricultural planning and resource allocation. It enables policymakers, agronomists, and researchers to make informed decisions regarding crop selection, irrigation management, and input allocation. Furthermore, in the face of rapidly changing environmental conditions and evolving climate patterns, an in-depth understanding of the determinants of crop yield becomes imperative for ensuring food security and economic stability in the region.

Crop yield analysis is the cornerstone of informed decision-making in agriculture, playing a pivotal role in ensuring food security, economic stability, and sustainable resource management. Understanding the factors that influence crop productivity is essential for several key reasons:

1.1.1 Food Security and Economic Stability:
Accurate predictions of crop yields are essential for ensuring a stable and sufficient food supply for growing populations. By understanding the expected output of key crops, policymakers and agricultural stakeholders can implement measures to prevent food shortages and stabilize prices in the market.

1.1.2 Resource Allocation and Optimization:
Effective resource allocation is crucial in agriculture, where inputs such as water, fertilizers, and labor have a direct impact on crop yields. Through rigorous analysis of historical yield data and environmental conditions, stakeholders can optimize resource allocation to maximize productivity while minimizing waste and environmental impact.

1.1.3 Climate Resilience and Adaptation:
In the face of climate change, understanding how different crops respond to varying environmental conditions is imperative. Crop yield analysis provides insights into the resilience of different varieties to extreme weather events, changing temperature patterns, and shifts in precipitation. This information is critical for adapting agricultural practices to a changing climate.

1.1.4 Policy Formulation and Agricultural Planning:
Governments and policymakers rely on accurate crop yield forecasts to formulate effective agricultural policies. These policies may include incentives for specific crops, water management strategies, and investment in infrastructure to support agricultural development.

1.1.5 Sustainable Agriculture Practices:
Crop yield analysis forms the basis for sustainable agriculture practices. By identifying the most efficient and

environmentally friendly techniques for maximizing yield, stakeholders can work towards long-term agricultural sustainability, preserving both natural resources and the economic viability of farming communities.

1.1.6 Research and Innovation:
Crop yield data serves as a foundation for ongoing agricultural research and innovation. Researchers can use this data to develop new varieties of crops, improve cultivation techniques, and create more efficient and sustainable agricultural technologies.

1.2 Features of the Dataset

This dataset, comprising 496 records, offers a wealth of information crucial for unraveling the complexities of crop yield dynamics. Among the key features included are Dist Code, Year, State Code, and Dist. Name, providing the spatial and temporal context for each record. Additionally, the dataset encompasses a comprehensive set of agricultural and meteorological parameters, including RICE IRRIGATED AREA (1000 ha), TOTAL CONSUMPTION (tons), ANNUAL RAINFALL (Millimeters), Precipitation Avg, Max Temp, Min Temp, Wind Speed, Wind Direction, Surface Soil Wetness, Profile Soil Moisture, and Root Zone Soil Wetness.

1.3 Research Objectives

The primary objective of this paper is to introduce and provide an in-depth description of this dataset, highlighting its potential applications in the realm of agricultural research. By offering a detailed record of environmental conditions and agronomic practices, this dataset facilitates a granular analysis of the factors influencing crop yield variations across the Telangana region.

1.4 Structure of the Paper

The remainder of this paper is organized as follows: Section 2 provides a detailed account of the data collection methodology, elucidating the sources, features, and preprocessing steps. Section 3 offers a comprehensive description of the dataset, including its size, scope, geographic coverage, and temporal span. Section 4 discusses the potential applications of this dataset in informing agricultural research, policymaking, and resource allocation. Finally, Section 5 concludes the paper by reiterating the significance of the dataset and its potential impact on agricultural development in the Telangana region.

2. Data Collection Methodology

Information is gathered from various government sources, and the features are consolidated, considering the year, state name, and district name. The data collection and preparation process encompassed extracting information from various government websites, utilizing Excel for collaborative efforts. After merging the datasets, additional information was incorporated to enrich the dataset comprehensively. To address missing values, a meticulous approach was taken, involving the imputation of missing values using statistical techniques.

To address missing values, a meticulous approach was taken, involving the imputation of missing values using the mean imputation method. This adaptive strategy aimed to ensure the dataset's integrity and optimize its suitability for subsequent analyses.

Sample dataset:

Dist Code	Year	State Name	Dist Name	RICE IRRIGA	TOTAL CONS	ANNUAL RAI	Precipitation	Max Temp	Min Temp	Wind_Speed	Wind Directi	Surface_soil	Profile_soil	Root_Zone_S	RICE YIELD (Kg per ha)
58	2010	Telangana	Mahabubnag	192.71	143591	755.1	2.45	43.42	9.74	0.06	235.56	0.53	0.64	0.65	2774.63
58	2011	Telangana	Mahabubnag	168.32	136767	494.4	1.26	42.69	11	0.09	254.5	0.43	0.56	0.58	2304.64
58	2012	Telangana	Mahabubnag	137.36	106717	632.7	1.72	44.33	10.56	0.02	266.19	0.41	0.57	0.59	2679.01
58	2013	Telangana	Mahabubnag	164.19	144946	910.8	2.26	44.17	11.73	0.05	246.12	0.52	0.61	0.62	2838.76
58	2014	Telangana	Mahabubnag	157.2	124485	601.1	1.98	42.78	12.7	0.05	206.38	0.48	0.57	0.61	2600.61
58	2015	Telangana	Mahabubnag	92.58	149617	471.9	1.44	44.09	10.93	0.02	215.69	0.4	0.54	0.58	2233.28
58	2016	Telangana	Mahabubnag	139.23	132783	567	1.77	43.75	12.22	0.08	265.38	0.45	0.58	0.6	2696.59
58	2017	Telangana	Mahabubnag	168.46	117612	486	1.89	43.36	12.51	0.06	271.19	0.45	0.58	0.61	2459.12
58	2018	Telangana	Mahabubnag	154.34	116783	831	1.16	44.16	11.31	0.05	266.62	0.38	0.54	0.57	2420
58	2019	Telangana	Mahabubnag	163.67	127658	756	1.57	44.48	10.64	0.04	219.06	0.43	0.55	0.59	2490
58	2020	Telangana	Mahabubnag	145.28	135647	894	3.08	43.56	11.29	0.03	190.12	0.58	0.66	0.69	3380

3. Comprehensive description of the dataset

The dataset encompasses a diverse set of agricultural and meteorological parameters, each playing a crucial role in determining crop yields. Understanding these features is essential for comprehending the complexities of crop production dynamics in the Telangana region.

3.1 District Code (Dist. Code):

The District Code serves as a unique identifier for each administrative district within Telangana. This feature provides the spatial context for each record, allowing for analysis at the district level. It enables researchers to explore variations in crop yields across different districts, which may be influenced by factors such as soil type, climate, and agricultural practices.

3.2 Year

The Year feature denotes the specific calendar year in which the data was recorded. This temporal dimension is crucial for studying the temporal trends and variations in crop yields. It enables researchers to assess the impact of changing environmental conditions, technological advancements, and policy interventions on agricultural productivity over time.

3.3 State Code (State Code)

The State Code provides a unique identifier for the state of Telangana. While this may appear redundant, it serves as an additional layer of verification, ensuring data integrity and consistency in multi-state datasets.

3.4 District Name (Dist Name)

The District Name feature offers a human-readable representation of the administrative district corresponding to the District Code. This provides clarity and context for researchers, especially when interpreting and visualizing the data.

3.5 Agricultural and Meteorological Parameters:

3.5.1 Rice Irrigated Area (1000 ha):
This parameter quantifies the extent of irrigated land dedicated to rice cultivation, measured in thousands of hectares. It is a critical factor influencing rice yield, as it reflects the availability of water resources for irrigation, a crucial aspect of rice farming.

3.5.2 Total Consumption (tons):
Total consumption in this context pertains to the cumulative quantity of N (Nitrogen), P (Phosphorus), and K (Potassium) fertilizers utilized, measured in tons. This metric is crucial for analyzing consumption patterns, aiding in the estimation of fertilizer demand, and ensuring a secure and sufficient food supply for the region.

3.5.3 Annual Rainfall (Millimeters):
Annual rainfall represents the total amount of precipitation received in millimeters over the course of a year. This meteorological parameter is a key determinant of crop growth and productivity, especially in rain-fed agriculture.

3.5.4 Precipitation Avg:
This feature denotes the average precipitation levels, providing insights into the typical rainfall patterns experienced in the Telangana region. It offers valuable information for assessing the suitability of different crops to the prevailing climate conditions.

3.5.5 Max Temp (Maximum Temperature):
Maximum temperature is a critical climatic factor influencing crop development. It impacts factors such as evaporation rates, water requirements, and growth stages, all of which contribute to overall crop yield.

3.5.6 Min Temp (Minimum Temperature):
Minimum temperature, like maximum temperature, plays a significant role in shaping crop growth. It affects factors such as dormancy, germination, and photosynthesis, influencing the overall yield potential.

3.5.7 Wind Speed:
Wind speed is a meteorological parameter that can impact crops in various ways, including affecting transpiration rates, pollen dispersal, and the susceptibility of plants to lodging (bending or breaking under wind pressure).

3.5.8 Wind Direction:
Wind direction provides additional context to wind speed, indicating the prevailing direction from which winds originate. This information is valuable for understanding the local climate and its potential impact on crop growth.

3.5.9 Surface Soil Wetness:
Surface soil wetness measures the moisture content of the top layer of soil. It is a crucial factor influencing root development, nutrient absorption, and overall plant health.

3.5.10 Profile Soil Moisture:
Profile soil moisture provides information on moisture levels at various depths below the soil surface. This feature is vital for understanding the availability of water to plant roots and can help guide irrigation practices.

3.5.11 Root Zone Soil Wetness:
Root zone soil wetness indicates the moisture content in the region of the soil where plant roots are actively absorbing water and nutrients. This parameter directly influences plant water uptake and, consequently, crop yield.

4. Potential applications of the dataset

The rich and comprehensive dataset presented in this study holds immense potential for a range of applications in agricultural research, policymaking, and resource management within the Telangana region. The following sections outline in detail the specific areas where this dataset can be leveraged:

4.1 Crop Yield Prediction Modeling

The dataset forms the cornerstone for constructing robust predictive models to estimate crop yields. By harnessing a wide array of agricultural and meteorological variables, researchers can apply machine learning algorithms and statistical methodologies to craft precise and dependable models. These models, once developed, offer the capability to predict crop yields. This empowers both farmers and policymakers with valuable insights for making judicious choices regarding planting timetables, resource distribution, and strategic decisions in the market.

4.2 Climate Impact Studies

Given the dataset's detailed records of meteorological parameters, it can be instrumental in conducting climate impact studies on agricultural production. Researchers can analyze the relationships between temperature, precipitation, and crop yields to understand how changing climate patterns may influence agricultural productivity. This information is invaluable for developing strategies to mitigate the effects of climate change on crop production.

4.3 Optimization of Resource Allocation

The dataset offers a treasure trove of insights into the determinants of crop yields, encompassing variables like irrigation, temperature, and soil moisture. Through the application of data-centric methodologies, stakeholders can fine-tune the allocation of resources. This ensures that vital inputs such as water, fertilizers, and labor are utilized judiciously to achieve the dual objectives of maximizing yield and minimizing both costs and environmental footprint.

4.4 Crop Selection and Rotation Strategies

This dataset equips farmers and agricultural advisors with the knowledge needed to make well-informed choices concerning crop selection and rotation strategies. Through a meticulous analysis of past yield data alongside prevailing environmental conditions, they can pinpoint the optimal crops for specific regions and seasons. Furthermore, the dataset serves as a

valuable guide for determining crop rotation schedules, ultimately promoting soil health and mitigating challenges related to pests and diseases.

4.5 Policy Formulation and Agricultural Planning

Policymakers and government agencies can leverage the dataset to formulate evidence-based agricultural policies. The insights gained from the analysis of historical yield data and environmental parameters can inform policies related to water management, subsidy programs, and infrastructure development. This, in turn, can contribute to the overall growth and sustainability of the agricultural sector in the Telangana region.

4.6 Research and Innovation

The dataset serves as a valuable resource for ongoing agricultural research and innovation. Researchers can use this data to investigate emerging trends, develop new crop varieties, and explore innovative cultivation techniques. Additionally, the dataset can be a catalyst for interdisciplinary research, fostering collaborations between agronomists, climatologists, and data scientists.

4.7 Education and Capacity Building

The dataset can be a valuable educational tool for training future generations of agricultural scientists and practitioners. It provides real-world, region-specific data for hands-on learning experiences. Additionally, it can be used to develop case studies and research projects that enhance students' understanding of agricultural systems and data analysis techniques.

4.8 Decision Support Systems

The dataset can be integrated into decision support systems (DSS) aimed at assisting farmers and agricultural advisors in making timely and informed decisions. By providing real-time or near-real-time insights based on historical data trends, the DSS can offer personalized recommendations for optimizing crop management practices.

5. Significance of the Dataset and its Potential impact on Agricultural Development in the Telangana Region

The dataset presented in this study stands as a testament to the meticulous effort invested in collecting and curating a comprehensive record of agricultural and meteorological parameters in the Telangana region. Its diverse features, ranging from district-specific identifiers to critical environmental conditions, form a valuable resource for advancing agricultural research, policy formulation, and practical farming practices.

5.1 Significance of the Dataset

The significance of this dataset cannot be overstated. It encapsulates not only the complex interplay between environmental factors and agronomic practices but also the profound implications of these interactions on crop yields. By offering a granular view of the agricultural landscape, this dataset empowers stakeholders with the knowledge needed to make informed decisions in a rapidly evolving agricultural sector.

5.2 Potential Impact on Agricultural Development

The potential impact of this dataset on agricultural development in the Telangana region is far-reaching. It holds the promise of revolutionizing the way farmers, researchers, policymakers, and advisors' approach agricultural practices. Through the applications outlined in Section 4, the dataset can:

5.2.1 Enhance Productivity and Food Security

Accurate crop yield predictions and optimized resource allocation strategies made possible by this dataset have the potential to substantially increase agricultural productivity. This, in turn, can contribute to food security by ensuring a stable and sufficient food supply for the growing population of Telangana.

5.2.2 Foster Climate Resilience

In the face of changing climate patterns, the dataset equips stakeholders with the knowledge needed to develop resilient agricultural systems. By understanding how crops respond to varying environmental conditions, farmers can adapt their practices to mitigate the impacts of climate change and ensure continued productivity.

5.2.3 Inform Evidence-Based Policies

Policymakers and government agencies can leverage the insights gained from this dataset to formulate policies that are rooted in empirical evidence. Water management strategies, subsidy programs, and infrastructure investments can be tailored to address the specific needs and challenges of the agricultural sector in Telangana.

5.2.4 Drive Research and Innovation

The dataset serves as a catalyst for innovation in agriculture. Researchers can use this resource to push the boundaries of agricultural science, developing new varieties, cultivation techniques, and technologies that have the potential to revolutionize farming practices in the region.

5.3 A Call to Action

As this dataset finds its way into the hands of farmers, researchers, policymakers, and educators, it is incumbent upon all stakeholders to seize the opportunities it presents. Collaboration and knowledge-sharing will be essential in unlocking the full potential of this resource and driving meaningful advancements in agricultural development in Telangana.

In conclusion, the dataset introduced in this paper is not merely a collection of data points; it is a catalyst for progress, a tool for empowerment, and a foundation for a more sustainable and resilient agricultural sector in the Telangana region. Its significance lies not only in its contents but in the

potential, it holds to shape the future of agriculture in this dynamic and agriculturally vibrant region.

6. Data Quality Assurance and Preprocessing

The quality and reliability of the dataset are paramount for meaningful analysis and accurate insights. This section outlines the steps taken to ensure the integrity and reliability of the dataset.

6.1 Data Cleaning

The initial phase of data preparation involved a meticulous examination of the dataset for any inconsistencies, missing values, or erroneous entries. This process included:

- **Identification of Inconsistencies:** Any discrepancies or irregularities in the data were identified. This ranged from outliers that may skew analysis to discrepancies in categorical variables.
- **Handling Duplicate Records:** Duplicate entries, if present, were carefully examined and removed to prevent redundancy in the dataset.
- **Addressing Formatting Issues:** Inconsistencies in data formatting, such as date formats or numerical representations, were standardized for uniformity.

6.2 Outlier Detection and Handling

Outliers, data points significantly deviating from the norm, were identified using robust statistical techniques. These outliers were then carefully evaluated to determine the appropriate course of action:

Assessment of Outlier Impact: The impact of outliers on the analysis was thoroughly assessed. Depending on the context, outliers were either retained if they represented valid data points with potential significance or were appropriately addressed through techniques such as imputation or removal. Application of Outlier Handling Techniques: Techniques like Winsorization or transformation methods were applied to mitigate the potential influence of outliers on subsequent analyses.

6.3 Normalization and Standardization

To ensure uniformity and comparability across different features, normalization and standardization techniques were applied:

- Normalization: This process scaled numerical features to a consistent range, typically [0, 1], ensuring that no single feature disproportionately influenced the analysis due to its magnitude.
- Standardization: Standardization involved centering the data around the mean and scaling it to have a standard deviation of 1. This transformation ensured that features with different units or scales contributed equally to the analysis.

6.4 Missing Data Imputation:

In instances where data was missing, a rigorous imputation process was undertaken:

- Imputation Techniques: Various imputation techniques were applied based on the nature of the missing data. Mean imputation, where missing values were replaced with the mean of the feature, and predictive modeling approaches, where missing values were estimated based on relationships with other variables, were employed to accurately estimate missing values.
- Sensitivity Analysis: Sensitivity analysis was conducted to assess the impact of different imputation methods on subsequent analyses, ensuring that imputed values were consistent with the overall patterns in the data.

6.5 Data Validation and Verification

Before finalizing the dataset, a series of validation checks and verification procedures were conducted:

- External Source Cross-Referencing: The dataset was cross-referenced with external sources, including agricultural reports and meteorological records, to validate the accuracy of the data. Any discrepancies or inconsistencies were carefully reviewed and resolved.
- Expert Review: Domain experts in agriculture and meteorology were consulted to verify the relevance and accuracy of the dataset, providing valuable insights and expertise in validating the data.

6.6 Summary of Preprocessing Steps

In summary, the data preprocessing steps undertaken in this study were integral in ensuring the integrity and reliability of the dataset. These rigorous procedures laid the foundation for meaningful analysis and robust insights into the factors influencing crop yields in the Telangana region.

7. Conclusion

In culmination, this paper presents a meticulously compiled and extensive dataset tailored for in-depth analysis of crop yields in the Telangana region. Spanning a period of [1966-2020], the dataset encompasses 496 records, each meticulously sourced from reputable agricultural and meteorological authorities. The dataset encapsulates a diverse array of features, ranging from district-specific identifiers to critical environmental parameters.

The significance of this dataset cannot be overstated. It serves as a powerful tool for understanding the intricate relationships between environmental conditions, agronomic practices, and crop yields. By offering a granular view of the agricultural landscape, it empowers stakeholders with the knowledge needed to make informed decisions in a rapidly evolving agricultural sector.

The potential impact of this dataset on agricultural development in the Telangana region is profound. It holds the promise of revolutionizing the way farmers, researchers, policymakers, and advisors' approach agricultural practices. Through accurate crop yield predictions, optimized resource allocation, climate resilience strategies, evidence-based

policies, and research-driven innovation, this dataset has the potential to substantially increase agricultural productivity, ensure food security, and foster sustainable practices.

As this dataset disseminates into the hands of those driving agricultural progress, it is essential for stakeholders to collaborate, share knowledge, and apply the insights gained. Through collective effort, we can unlock the full potential of this resource and drive meaningful advancements in agricultural development in Telangana.

References

[1] http://data.icrisat.org/dld/src/additional.html
[2] https://www.ifpri.org
[3] https://ecostat.telangana.gov.in/national_agricultural_insurance.html
[4] https://agricoop.nic.in/en/agriculturecontingency/telangana
[5] https://data.telangana.gov.in/collection/agriculture
[6] https://agri.telangana.gov.in/
[7] Son, N.; Chen, C.; Chen, C.; Chang, L.; Chiang, S. Rice yield estimation through assimilating satellite data into a crop simumlation model. The Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2016, 41, 993–996.
[8] Bhanusree, D.; Srinivasachary, D.; Nirmala, B.; Supriya, K.; Naik, B.B.; Rathod, S.; Brajendra; Reddy, B.N.K.; Bellamkonda, J. Application of the CERES-Rice Model for Rice Yield Gap Analysis. Int. J. Environ. Clim. Change 2022, 12, 3471–3478.
[9] Kumar, K. S. (2009). Climate sensitivity of Indian agriculture (Working Paper No.43/2009). Madras School of Economics, 1–33.
[10] Padakandla, S. R. (2016). Climate sensitivity of crop yields in the former state of Andhra Pradesh, India. Ecological indicators, 70, 431–438.
[11] Singh, P.; Singh, K.; Baxla, A.; Rathore, L. Impact of climatic variability on wheat yield predication using DSSAT v 4.5 (CERES-wheat) model for the different agroclimatic zones in India. In Climate Change Modelling, Planning and Policy for Agriculture; Springer: Berlin/Heidelberg, Germany, 2015; pp. 45–55.