FDRP Journal's

INDJCST-0000942

15

FDRP

Journal's

Document Details

25% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

15 Not Cited or Quoted 25%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

22% Internet sources
23% Publications
3% Submitted works (Student Papers)

Match Groups

- **15** Not Cited or Quoted 25%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

22% Internet sources
23% Publications
3% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Publication	**S Vaishnavi., M Shobana., R Sabitha., S Karthik.. "Agricultural Crop Recommendati...**	**10%**
2	Internet	**ijmrset.com**	**8%**
3	Publication	**Biplob Dey, Jannatul Ferdous, Romel Ahmed. "Machine learning based recommen...**	**2%**
4	Internet	**ijircce.com**	**2%**
5	Internet	**ijrpr.com**	**<1%**
6	Internet	**digital.lib.usu.edu**	**<1%**
7	Publication	**"Merged Papers", 2020 7th International Conference on Smart Structures and Sys...**	**<1%**
8	Internet	**123dok.com**	**<1%**
9	Internet	**www.ijtsrd.com**	**<1%**

Agricultural Products: CVF Yield Prediction Using Ensemble Methods And Machine Learning Models

Dr D J Samatha Naidu, K. Aruna,
Department of MCA, Annamacharya PG College of Computer Studies ,Rajampet ,Andhra Pradesh ,India
Department of MCA, Annamacharya PG College of Computer Studies ,Rajampet ,Andhra Pradesh ,India
Corresponding Author: kisugallaaruna@ gmail.com

Abstract: The goal of this research is to improve agricultural decision-making by employing machine learning models and ensemble approaches to forecast the production of agricultural goods, particularly Crop, Vegetables, and Fruits (CVF). Several machine learning techniques, including Random Forest, Gradient Boosting, and ensemble approaches, are used to increase forecast accuracy by utilizing historical climate, soil, and yield data. By combining several models, prediction errors are reduced and reliability is increased, giving farmers and other stakeholders important information for maximizing resource allocation, raising productivity, and guaranteeing food security. In order to facilitate data-driven agricultural planning, this research attempts to close the gap between conventional farming methods and sophisticated predictive analytics.

Key Word: CVF Yield Prediction, Ensemble Methods, Machine Learning models

I. Introduction

[illegible] in this below.

- [illegible] March.
- [illegible]
- [illegible]
- [illegible]

II. Material And Methods

1 Material

1.1. Data Collection

Usually, a variety of sources provide the data utilized in crop advice. The following are some of the most significant datasets:

- **Soil Data**: Details on the pH level, texture, fertility, and type of soil. Government agencies, agricultural research institutes, and soil sensors can all provide this data.
  Eg: pH, organic matter, potassium, phosphorus, and nitrogen levels in the soil.
- **Climate data**: weather information for a specific time period, including temperature, humidity, rainfall, wind speed, and hours of sunshine.
  Eg: Local weather stations, online APIs (like Open Weather Map), and government weather bureaus.
- **Crop data:** historical information on crop yields, crop types, and farming regions' locations.
  Eg: yield per hectare for crops such as vegetables, rice, corn, or wheat.

1.2. Tools and Libraries

Programming Languages: python R

1) **Machine Learning Libraries**:

- **Scikit-learn**: A well-liked package for putting fundamental machine learning methods like clustering, regression, and classification into practice.
- **XGBoost / Light GBM**: Libraries for gradient boosting are used to solve classification and regression issues.
- **TensorFlow / Keras**: Libraries for methods based on deep learning.
- **Pandas**: to manipulate and preprocess data.
- **NumPy / SciPy**: For calculations involving numbers.
- **Matplotlib / Seaborn:** for the visualization of data.

1.3. Hardware Requirements

- **Cloud Services**: For storing big datasets and executing computationally demanding machine learning models, use AWS, Google Cloud, or Microsoft Azure.
- **Computational Resources:** CPUs and GPUs for deep learning techniques in particular, as well as for training machine learning models.
- **IoT Devices**: Sensors installed on farms to track soil, climate, and other environmental parameters in real time.

2. Methods

2.1. Data Preprocessing

Prior to using machine learning methods, data pretreatment is an essential step. Among the actions involved are:

- **Data Cleaning**: addressing missing data, eliminating anomalies, and fixing dataset mistakes.
- **Feature Engineering**: Making the raw data more helpful for machine learning models by adding extra features.
- **Normalization/Standardization**: Features are scaled to prevent any one feature from taking over the model because of its size.
- **Categorical Data Encoding**: Converting soil type and other category features.

2.2. Feature Selection

- **Correlation Matrix**: removing strongly connected aspects by analyzing feature correlations.
- **Principal Component Analysis (PCA)**: A method for keeping as much variance in the data as feasible while reducing its dimensionality.
- **Random Forest Feature Importance**: assessing each feature's significance in forecasting crop suitability using tree-based models.

2.3. Model Selection

- **Decision Trees:** used to categorize crops according to a variety of input characteristics, such as climate and soil type.
- **Random Forest**: A decision tree ensemble approach that works well for problems involving regression and classification, particularly when working with structured data such as soil and farm data.
- **Support Vector Machines (SVM)**: can be applied to classification problems, particularly when there are distinct boundaries between the various classes (crop types) in the data.
- **K-Nearest Neighbors (KNN)**: classification based on a data point's "nearness" to other like points using a non-parametric approach.
- **Gradient Boosting Machines (GBM)**: For regression and classification problems, methods like XGBoost and Light GBM are effective and manage outliers and missing variables effectively.

2.4. Model Training

Splitting the Data: Make training and testing sets out of the dataset (e.g., 80% training, 20% testing).

Cross-Validation: Use cross-validation (such as k-fold) to make sure the model is reliable and works well with unknown data.

Hyperparameter Tuning: For best results, adjust hyperparameters using strategies like Grid Search or Random Search.

2.5. Crop Recommendation System

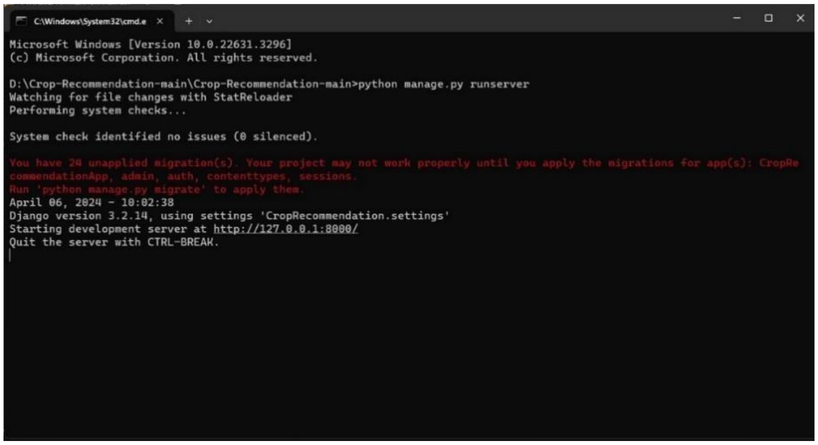
The model can be used to suggest crops for fresh input data after it has been trained. The input data could consist of:

- **Soil Conditions**: Moisture content, fertility, and soil type.
- **Climate Information**: sunshine hours, humidity, temperature, and precipitation.
- **Geographical Information:** Location, elevation, and past crop information.

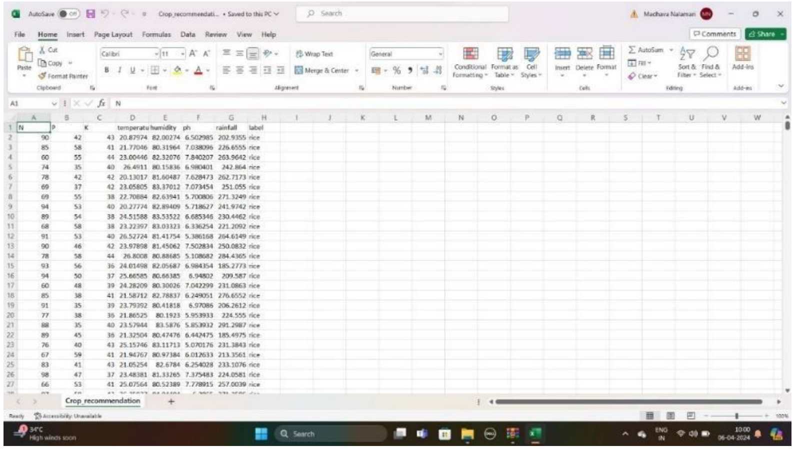
The method uses the trained model to forecast which crop or crops would be most suited for a certain location while taking environmental and climatic factors into account. Depending on regional differences in soil health, climate, and other local considerations, the suggestion can be tailored.

III. Result

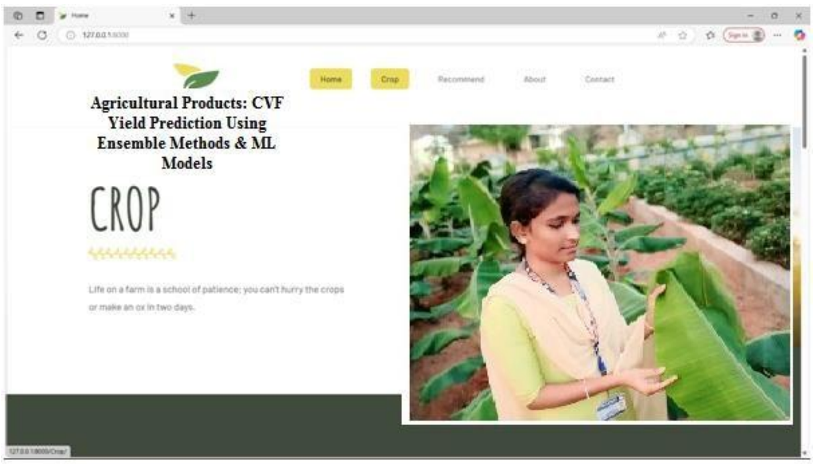
The study's findings show that machine learning models and ensemble approaches greatly increase the precision of yield prediction for crops, vegetables, and fruits (CVF). In terms of accuracy and dependability, ensemble methods like Random Forest and Gradient Boosting performed better than individual models among the models that were examined. More accurate forecasts were produced by combining data from several sources, such as historical yield records, soil characteristics, and climate. Ensemble models, according to comparative analysis, improved decision-making for farmers and stakeholders by lowering prediction errors and producing more reliable production predictions. These results demonstrate how machine learning may be used to maximize agricultural output and guarantee food security by providing data-driven insights.



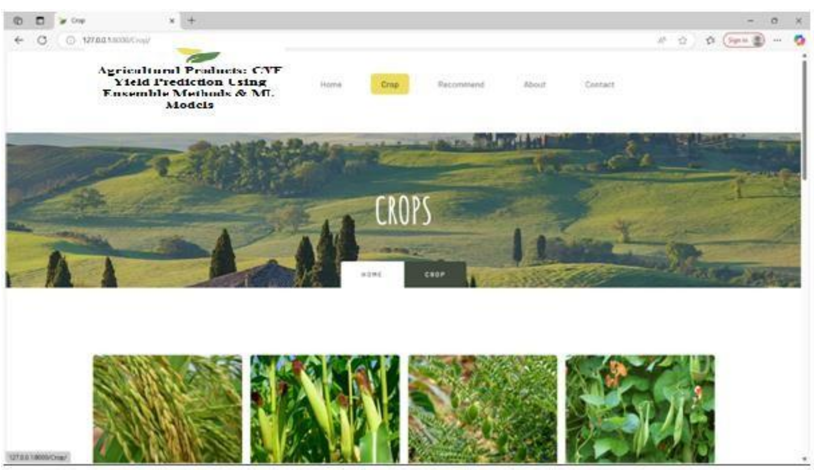
Screen1: CMD Running Process



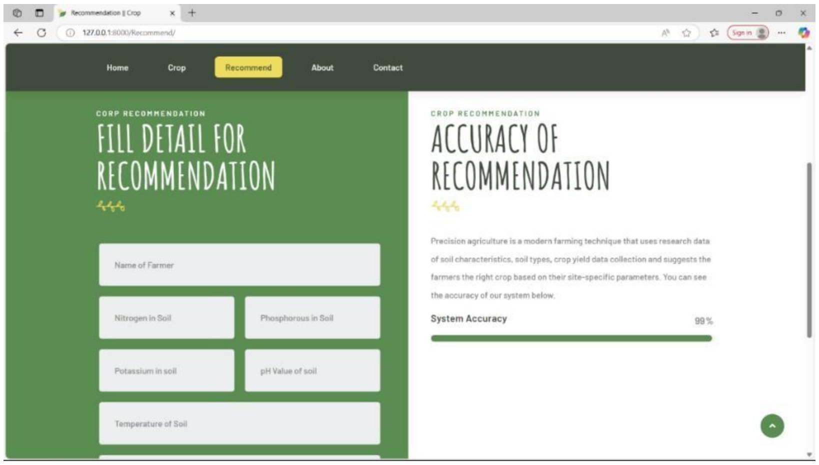
Screen 2: Showing Data Set



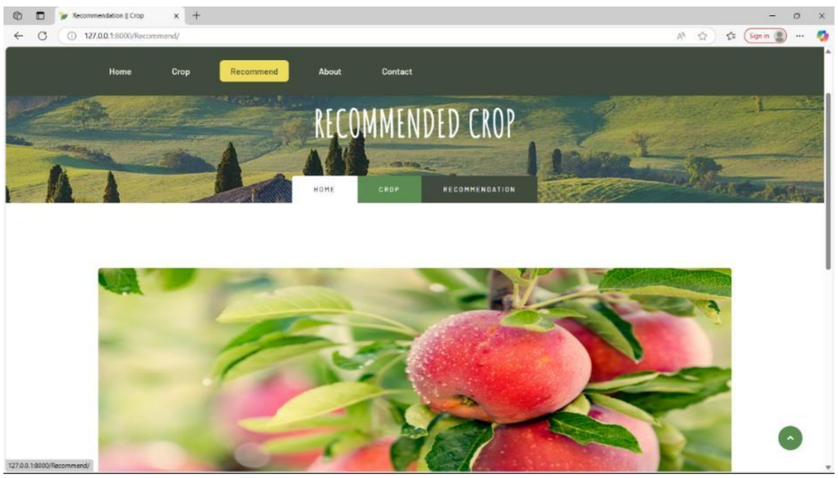
Screen 3: Show The Home Page



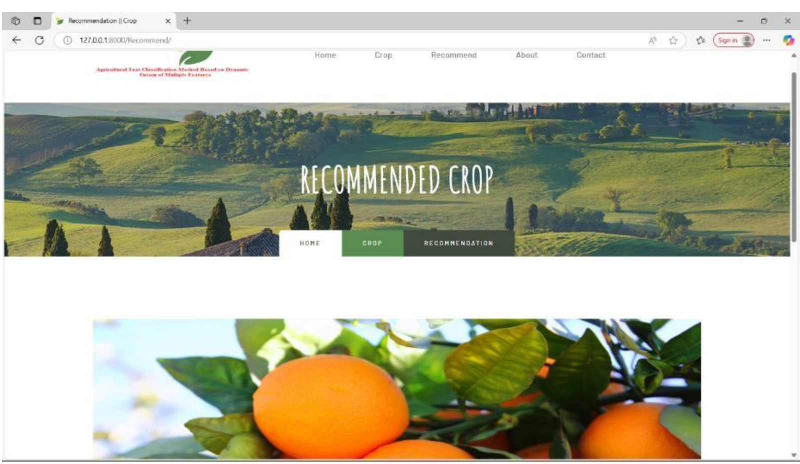
Screen 4: Show The Multiple Crops



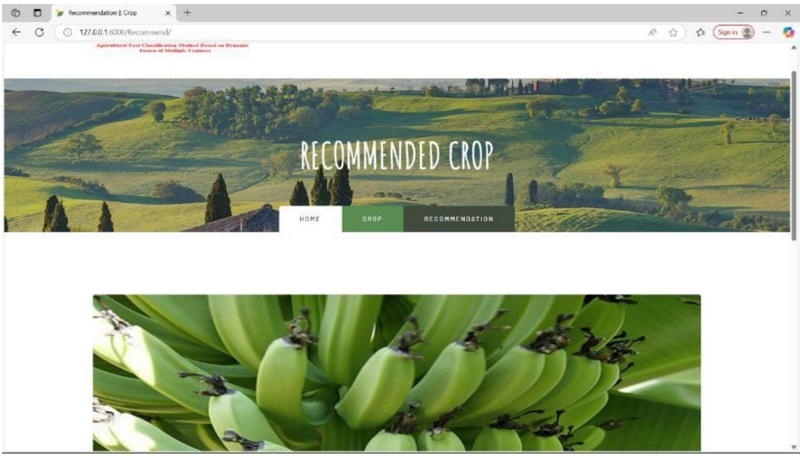
Screen 5: Fill Detail For Recommendation



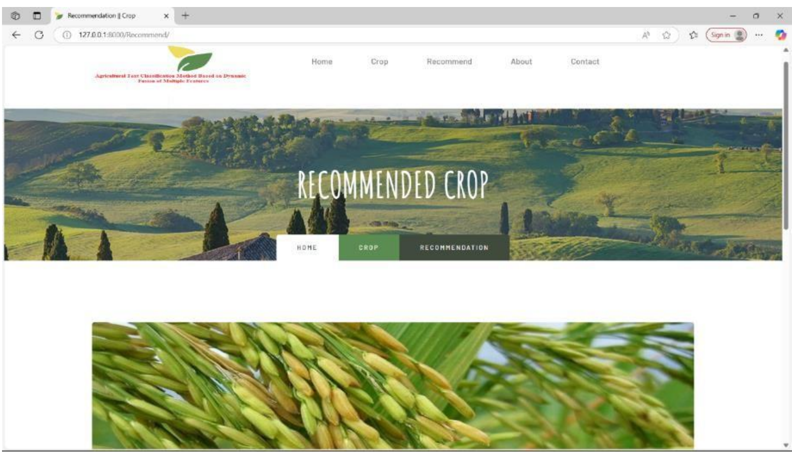
Screen 6: View Apple Crop Recommended



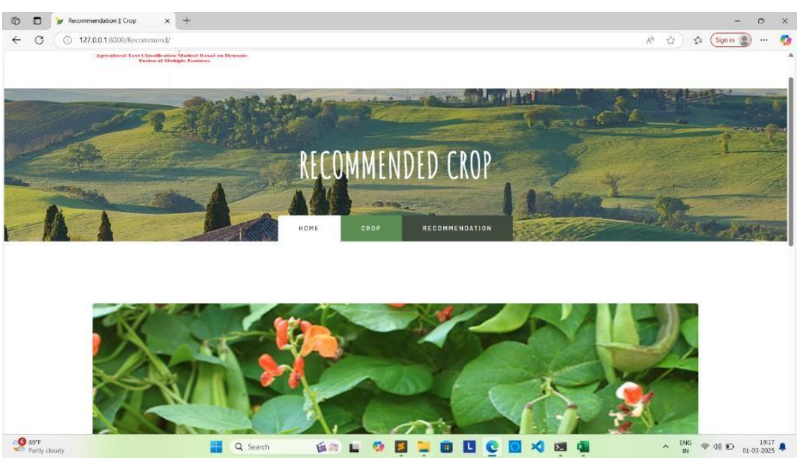
Screen 7: View Of Orange Crop Recommend



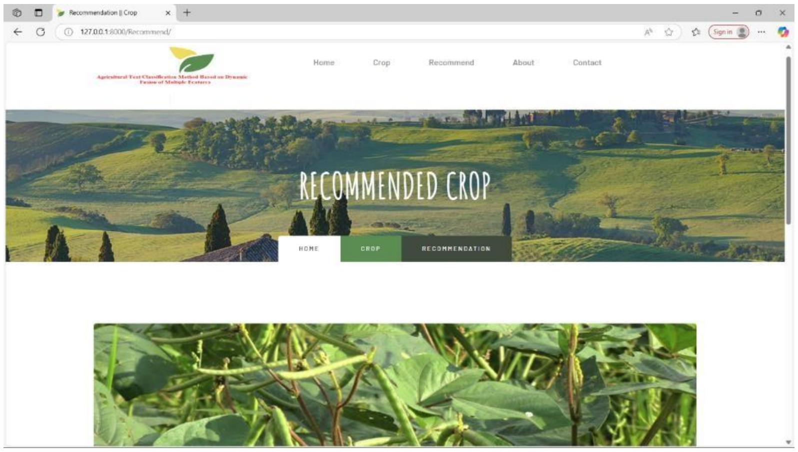
Screen 8: View Of Banana Crop Recommended



Screen 9: View Of Rice Crop Recommended



Screen 10: View Of Kidney Beans



Screen 11: View Of Moth beans

The process of using your "screen2" dataset to produce suggestions and show the outcomes is described below. When you enter the required information in this example, the project will filter the data and display items like.

IV. Discussion

A key component of contemporary farming is the ability to anticipate the yield of agricultural products, such as crops, vegetables, and fruits (CVF), which facilitates improved planning, resource allocation, and decision-making. By utilizing extensive datasets, weather patterns, soil conditions, and past yield records, ensemble methods and machine learning (ML) models have become effective instruments for raising the accuracy of yield forecasts. By merging several models to lower bias and variance, ensemble approaches like bagging, boosting, and stacking improve predictive performance. Accurate yield estimations are further enhanced by deep learning-based techniques (e.g., CNNs and LSTMs) and machine learning models such as Random Forest, Support Vector Machines (SVM), Gradient Boosting Machines (GBM), and Random Forest. Compared to conventional statistical methods, these models improve predictions by analyzing intricate interactions among numerous agricultural aspects. The predictive power of ML models is increased by combining them with remote sensing technologies, satellite imaging, and IoT-based sensor data. Nonetheless, there are still issues with data accessibility, computational complexity, and model interpretability. By tackling these issues using feature engineering, model optimization, and explainable AI methodologies, CVF yield prediction can become even more accurate and efficient, promoting food security and sustainable farming methods.

V. Conclusion

In this paper, significance of management of crops was studied vastly. Farmers need assistance with recent technology to grow their crops. Proper prediction of crops can be informed to agriculturists in time basis. Many Machine Learning techniques have been used to analyze the agriculture parameters. Some of the techniques in different aspects of agriculture are studied by a literature study. Blooming Neural networks, Soft computing techniques plays significant part in providing recommendations.Considering the parameter like production and season, more personalized and relevant recommendations can be given to farmers which makes them to yield good volume of production.

References

1) Amazon Web Services, [Online]. Available: https:// aws.amazon.com/cn/.D.A. Sarma, X. Don.

2) D.A. Sarma, X. Dong, and A. Halevy, Bootstrapping pay-as-you-go data [1] P. Chopade, J. Zhan, and M. Bikdash. Node attributes and edge structure for large-scale big data network analytics and community detection. In International Symposium on Technologies for Homeland Security (HST), pages 1–8, 2015.

3) X. Que, F. Checconi, F. Petrini, and J. A. Gunnels. Scalable community detection with the louvain algorithm. In Parallel and Distributed Processing Symposium (IPDPS), pages 28–37, 2015.

4) P. Cui, Z. Wang, and Z. Su. What videos are similar with you? :Learning a common attributed representation for video recommendation. In ACM International Conference on Multimedia (MM), pages 597–606, 2014.

5) **Dr D J Samatha Naidu**, completed MCA from S V University, Tirupati , MPhil computer science from Madurai Kamaraj University Madurai, MTech in Computer Science and Engineering in JNTUA, Anantapur, PhD in Computer Science from Vikrama Simha Puri University, Nellore, currently working as Professor and Principal Annamacharya PG College of computer studies, Rajampet since 20 years, 2 years industrial experience as network support engineer, 12 years research experience, Completed consultancy and major projects like AICTE and other IT industry.150 international Research journal papers published,100 national and international conferences are attended and presented papers.10 National And International Design Grant Patents, Utility Patents, Copy Rights, Patents Are Published. 12 Text Books are published.8 Theory and Lab Manuals are designed for MCA and MBA students, 22 national and international professional bodies Life member, associate member, fellow member for Edunix research university USA, ISTE, IE, IACSIT, IAENG, IMRF, IRDP, NITTE, GLOBAL PROFESSOR FOR ALUMINI ASSOCIATION, HRPC, UAE, EAI, KALA'S LIFE MEMBERSHIP, COUNCIL OF TEACHERS EDUCATION MEMBER, GLOBOL TEACHERS ASSOCIATE MEMBER,INSTITUTE OF GREEN ENGINEERS. Research papers are Reviewed as Editorial Member and Reviewer Member, 25 National and international awards are received from USA, MALAYSIA, Andhra Pradesh, Telangana, Tamilnadu state organizations, I received prestigious university best teacher award received from JNTUA Anantapur for 2022. Am very much honour getting second time award as university principal award for 2024 from JNTUA.

6) **K Aruna**, earned her Bachelor's degree in Computer Science (B.Sc.) from Rayalaseema University in 2023, where she developed a strong foundation in programming, data structures, and software development. Currently, she is pursuing a Master of Computer Applications (MCA) at Annamacharya PG College of Computer Studies, further enhancing her technical expertise and research skills. This publication marks her first contribution to the academic community, reflecting her keen interest in the intersection of technology and healthcare. Her primary research focuses on the Agricultural products: CVF yield prediction using ensemble methods & machine learning models.