IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

Crop Yield Prediction using CNN

Ritik Bohra, Shrunkhal More, Shwetal Kamat, Rishabh Pathak, Prof. Manisha Shitole
Department of Computer Science and Engineering
MIT ADT University, Pune, India

Abstract: *Deep learning is a branch of Machine Learning which is completely based on artificial neural networks, as neural networks are going to mimic the human brain so deep learning is also a kind of mimic of the human brain. Farming is the main occupation of India. Crop yield has a direct impact on nation and international economies annually and the yield predicted plays a significant part in the food management and agriculture sector. The task is to build a prediction model for crop production . A prerequisite of intelligent systems has brought artificial neural networks to become a new technology which provides assorted solutions for complex problems in agriculture research. Performance of the agriculture sector mainly hinges on natural forces such as spatio-temporal distribution of rainfall, temperature, climate etc, with the result any deviation of monsoon from the normal pattern brings about numerous fluctuations in area and production. Crop yield has a direct impact on nation and international economies annually and the yield predicted plays a significant part in the food management and agriculture sector. The task is to build a prediction model for crop production. The basic principle of ANN architecture, Data Modeling for Prediction involves four stages namely historical data analysis (Descriptive), Data preprocessing, modeling of Data and Performance Estimation. First classify data based on different attributes. Regression analysis using CNN, it observes the relation between an independent (predictor) and dependent (target) variables. Based on relation training the model will predict crop yield production. application of ANN in predicting crop yield by using various crop performance features as input parameter.*

Keywords: crop performance

I. INTRODUCTION

Farming is the main occupation of India. Crop yield has a direct impact on national and international economies annually and the yield predicted plays a significant part in the food management and agriculture sector. India's economy mainly depends on agriculture yield growth and their related agro industry. 61% percent of the total geographical area of India comes under agriculture. This model is taking input as a dataset of India. This particular dataset includes the data from all the states, districts of India and the crops in the respective regions. So, we are going to study different states and districts of India.

A prerequisite of intelligent systems has brought artificial neural networks to become a new technology which provides assorted solutions for complex problems in agriculture research. Performance of the agriculture sector mainly hinges on natural forces such as spatio-temporal distribution of rainfall, temperature, climate etc, with the result any deviation of monsoon from the normal pattern brings about numerous fluctuations in area and production. Crop yield has a direct impact on nation and international economies annually and the yield predicted plays a significant part in the food management and agriculture sector. The task is to build a prediction model for crop production using data prediction.

1.1 Problem Statement

The problem statement of our project is to design and develop a yield prediction model using CNN (Convolutional Neural Network). The aim is to accurately predict crop production, which has a direct impact on national and international economies. The yield prediction plays a significant role in food management and the agriculture sector. The project focuses on building a classification model and a yield prediction model to improve accuracy based on the MAPE(Mean Absolute Percentage Error) metric. The scope of the project includes predicting crop production with accuracy using a dataset of India, considering different attributes such as state, district, season, and crop. The limitations of the project include the requirement for realistic and known categorical inputs and the use of specific software tools for system implementation.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

II. METHODOLOGY

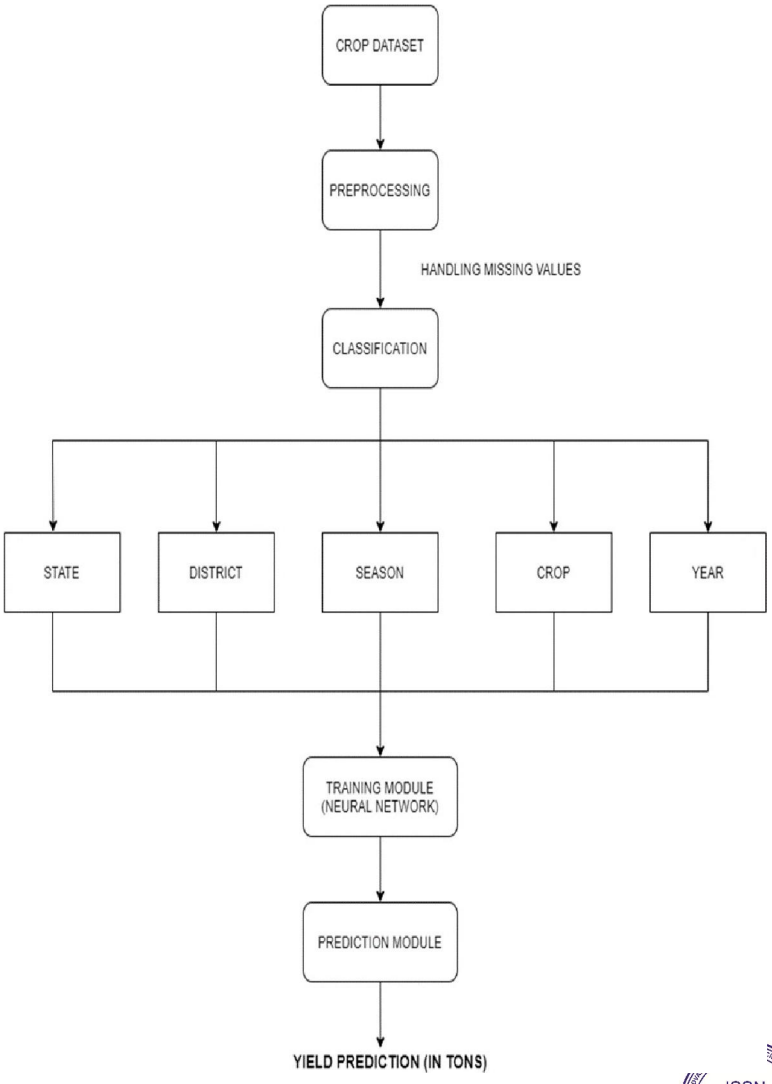
The methodology used in our project involves several steps, including data collection, preprocessing techniques, and the architecture of the Convolutional Neural Network (CNN) model. The model was trained and optimized for better accuracy using specific techniques.

Data Collection: The dataset used in our project was collected from the website of the agriculture government portal. It includes state-wise, district-wise, season-wise, year-wise, and crop-wise data on the area covered and production of crop yield in India

The attributes of area are measured in hectares, and crop production is measured in tonsper hectare.

Preprocessing Techniques: The preprocessing phase of our project involved several techniques to ensure the quality and reliability of the data. We eliminated redundant data features, removed null values, and filtered the dataset based on specific range values.Outliers were also removed to improve the accuracy of the model. Categorical data suchas district, season, and crop were encoded into numerical values using label encoding.

Architecture of the CNN Model: The CNN model architecture was designed to handle the prediction of crop yield. It involved input layers for each input parameter, such as district, year, season, and crop. Additionally, embedding layers were created for thecategorical parameters (district, season, and crop) using word embedding techniques.These layers were then concatenated and attached to the CNN model.

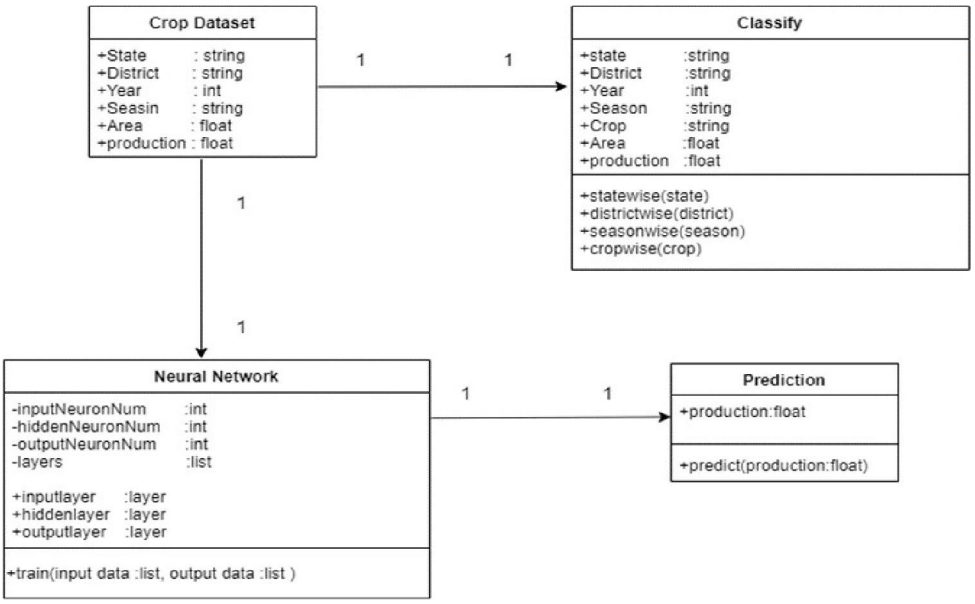


IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

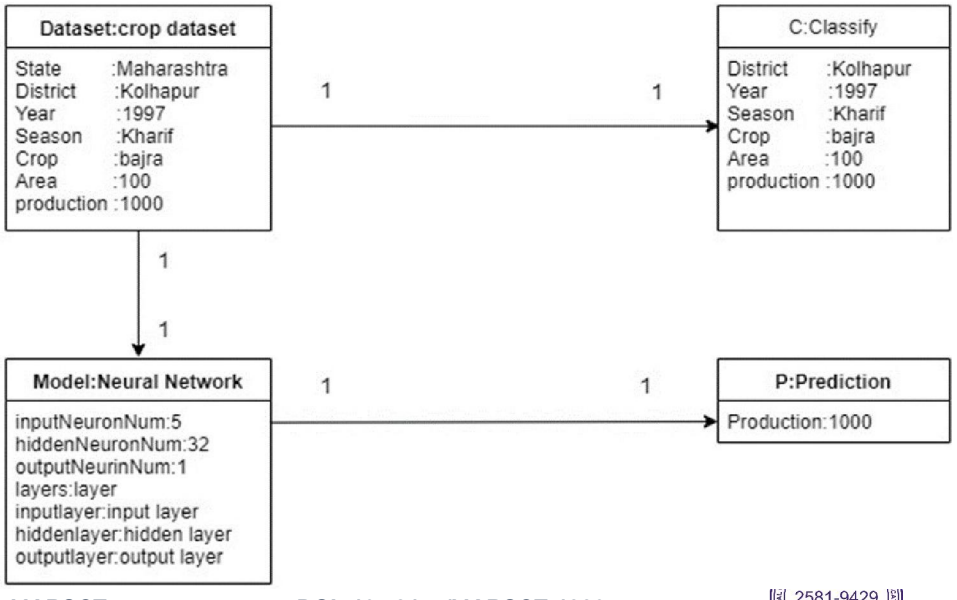
Training and Optimization: The training of the CNN model was performed using the preprocessed dataset. Numerical data, such as year, area, and production, werenormalized to ensure consistent scaling. The data from each column was stored inseparate arrays. The model was then compiled and optimized using suitable optimization techniques. The training process involved iterating over the dataset until better accuracywas achieved, measured using metrics such as MAPE (Mean Absolute Percentage Error)and MSE (Mean Squared Error). To optimize the model, we used techniques such as adjusting the learning rate, tuning hyperparameters, and employing regularizationtechniques like dropout or L1/L2 regularization.

These techniques helped in reducing overfitting and improving the generalization capability ofthe model. Overall, the methodology involved collecting the dataset, preprocessing the data to ensure quality, designing the CNN model architecture, training the model using the preprocessed dataset, and optimizing the model for better accuracy

Structural Design -



Object Diagrams -



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

III. OUR CONTRIBUTION

Our contribution for this project is the design and development of a crop yield prediction model using CNN (Convolutional Neural Network). We have conducted a literature review on deep learning and related work in crop yield prediction. We have analyzed the problem statementand defined the aim and objectives of the project. We have also determined the scope and limitations of the project. In terms of system design, we have implemented preprocessing techniques to eliminate redundant data features, remove null values and outliers, and filter the dataset. We have also developed a classification module to categorize the data based on district, year, season, and crop, and save the categorized data in separate CSV files. Furthermore, wehave implemented the training module of the neural network model, which involves normalizing the numerical data, creating input and embedding layers for the input parameters, and training the model using the training dataset. We have optimized and compiled the model, and evaluated its performance using metrics such as MAPE (Mean Absolute Percentage Error) and MSE (Mean Squared Error). Additionally, we have implemented the front-end design for input parameters and display of the predicted output using Flask and HTML. We have also provided an installation guide and user manual for the system. Overall, our contribution includes the design and implementation of various modules, data preprocessing, classification, training of the neural network model, and the development of the user interface.

Literature Survey -1. Year -2022

Authors -Seyed Mahdi Mirhoseini Nejad, Dariush Abbasi-Moghadam, Alireza Sharifi,
Nizom Farmonov, Khilola Amankulova and Mucsi László´z

Title -Multispectral Crop Yield Prediction Using 3D-Convolutional Neural Networksand Attention Convolutional LSTM Approaches

Description -The document discusses two proposed models for crop yield prediction using deep learning techniques. The models aim to improve the accuracy of predictions by extracting relevant features from MODIS satellite data and using advanced neural network architectures. The first model combines 2D-CNN, skip connections, and LSTM-Attentions, while the second model incorporates 3D-CNN, skip connections, and ConvLSTM. The models are trained and evaluated using data from over 1800counties in the USA where soybean is mainly cultivated.

The results show that the second proposed model outperforms the other techniques in terms of mean absolute error (MAE). The MAE for the second model is 4.3, while the other models achieve MAE values ranging from 6.003 to 7.002. The models are also evaluated using other metrics such as root mean square error (RMSE), mean absolute percent error (MAPE), and mean squared logarithmic error (MSLE), and the second model consistently performs better than the other models.

The document highlights the advantages of the proposed models, including the use of attention mechanisms to focus on relevant features, the inclusion of 3D-CNN for extracting spectral, spatial, and temporal features, and the combination of LSTM and convolution to reduce network complexity. The models have been compared with existing methods, including DeepYield, ConvLSTM, 3DCNN, and CNN-LSTM, and have shown superior performance.

However, the document acknowledges limitations in the input data, as full-size images were not used due to hardware and software limitations. Instead, histograms were generated as input data, which may affect the results compared to other studies. The document suggests future work could include the incorporation of climatic data and higher-resolution satellite data, as well as the use of spiking neural networks.

In conclusion, the proposed models show improved accuracy in crop yield prediction compared to existing methods. The models effectively extract relevant features from satellite data and utilize advanced neural network architectures to achieve accurate predictions. These models have the potential to benefit various sectors, including agriculturalists and agencies involved in crop forecasting.

2. Year -2022

Authors -K. Vignesh, A. Askarunisa and A. M.Abirami

Title -Optimized Deep Learning Methods for Crop Yield Prediction

Description -The document discusses the use of optimized deep learning methods for crop yield prediction. The authors propose a model that combines data mining and deep learning to predict crop yields based on

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

environmental, land, water, and crop characteristics. They use a Discrete Deep belief network with Visual Geometry Group(VGG) Net classification method and the Tweak Chick Swarm Optimization approachto estimate agricultural production. The model is able to accurately predict crop output with 97 percent accuracy, surpassing existing models. The authors also compare theirproposed model with other methods and demonstrate its superior performance in terms of precision, recall, accuracy, and mean square error. The document provides a detailed description of the methodology, including data preprocessing, feature extraction, featureselection, and classification. The authors suggest that their approach can be applied toother datasets, such as soil and rainfall data, to further test its scalability. Overall, the proposed model shows promise in accurately predicting crop yields and assisting farmers in making informed management and financial decisions.

3. Year -2022
Authors -Priyanka Sharma, Pankaj Dadheech, Nagender Aneja and Sandhya Aneja
Title -Predicting Agriculture Yields Based on Machine Learning Using Regression andDeep Learning
Description -The document discusses a research study on predicting crop yields in India using machine learning and deep learning techniques. The study aims to assistfarmers in increasing their productivity and making informed decisions about crop selection and farming practices. The researchers collected data from official government websites and used variables such as rainfall, crop type, meteorological conditions, area, production, and yield to develop models for crop yield prediction.
The study compares the performance of different machine learning algorithms, including decision tree, random forest, and XGBoost regression, as well as deep learning techniques such as convolutional neural network (CNN) and long-short term memory network (LSTM). The evaluation metrics used include accuracy, root mean square error, mean absolute error, and standard deviation.
The results show that the random forest algorithm and CNN perform the best in terms of accuracy and loss. The random forest algorithm achieves a maximum accuracy of 98.96%, while the CNN has a minimum loss of 0.00060. These models outperform other machine learning and deep learning methods in predicting crop yields.
The study highlights the importance of accurate crop yield prediction for farmers, especially in a country like India where agriculture plays a significant role in the economy. By using machine learning and deep learning techniques, farmers can make informed decisions about crop selection and farming practices, ultimately increasing their productivity and reducing losses.
The researchers suggest further research to improve the accuracy of the models, such as incorporating remote sensing data and satellite imagery for more precise predictions. They also emphasize the need for more historical data on climate and environment to enhance the models' performance.
In conclusion, the study demonstrates the effectiveness of machine learning and deep learning techniques in predicting crop yields. The random forest algorithm and CNN show the best performance in terms of accuracy and loss. The findings of this study can help farmers in India and other agricultural regions in making informed decisions to improve crop productivity and food security.

4. Year -2022
Authors -Uferah Shafi, Rafia Mumtaz, Zahid Anwar, Muhammad Muzyyab Ajmal, Muhamad Ajmal Khan , Zahid Mahmood, Maqsood Qamar, and Hafiz MuhamadJhanzab
Title -Tackling Food Insecurity Using Remote Sensing and Machine Learning-BasedCrop Yield Prediction
Description -This document presents a framework for predicting wheat grain yieldusing remote sensing data and machine learning algorithms. The study focuses on the importance of accurate yield prediction for food security and resource management in the agricultural industry. The authors use three regression techniques, including Random Forest, Xtreme Gradient Boosting (XGB) regression, and Least Absolute Shrinkage & Selection Operator (LASSO) regression, to predict wheat grain yield based on data collected from three experimental fields with different sowing dates. Theperformance of the models is assessed using evaluation metrics such as coefficient of determination (R2) and mean absolute error (MAE). The results show that LASSO achieved the highest performance, with an R2 of 0.93 and an MAE of 21.72. The studyalso explores the impact of sowing dates on crop

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

yield and finds that the wheat field with the earliest sowing date had the highest average yield. The research provides insights that can help farmers and agronomists make informed decisions about crop management activities. Future work includes integrating more predictors and exploring deep learning techniques for crop yield prediction.

5. Year -2021
Authors -Farhat Abbas, Hassan Afzaal, Aitazaz A. Farooque and Skylar Tang
Title -Crop Yield Prediction through Proximal Sensing and Machine LearningAlgorithms
Description -This document discusses the use of proximal sensing techniques and machine learning algorithms for predicting crop yield. The study focuses on potatotuber yield in six fields in Atlantic Canada over two growing seasons. The researchers collected data on soil and crop properties, including soil electrical conductivity, soil moisture content, soil slope, normalized-difference vegetative index (NDVI), and soil chemistry. Four machine learning algorithms - linear regression (LR), elastic net (EN), k-nearest neighbor (k-NN), and support vector regression (SVR) - were used to predict potato tuber yield based on this data. The results showed that SVR outperformed theother algorithms in terms of accuracy, with the lowest root mean square error (RMSE) values for all datasets. The performance of k-NN was poor in three out of four datasets. The study also highlighted the importance of large datasets for generating accurate predictions using any model. The findings of this study can contribute to the development of site-specific management zones for potato crops, which are crucial for food security initiatives globally. Overall, the research demonstrates the potential of machine learning algorithms in improving crop yield prediction using proximal sensingdata.

Year -2021
Authors -Thomas van Klompenburg, Ayalew Kassahuna, Cagatay Catal
Title -Crop yield prediction using machine learning: A systematic literature
Description -This document is a systematic literature review on the use of machine learning for crop yield prediction. The review aimed to extract and synthesize thealgorithms and features used in crop yield prediction studies. The authors retrieved 567 relevant studies from six electronic databases and selected 50 studies for further analysis based on inclusion and exclusion criteria. The most commonly used features in the selected studies were temperature, rainfall, and soil type, and the most appliedalgorithm was Artificial Neural Networks.
Additionally, the authors performed an additional search to identify deep learning-based studies and found 30 papers that applied deep learning algorithms. The most widely used deep learning algorithm in these studies was Convolutional Neural Networks (CNN), followed by Long-Short Term Memory (LSTM) and Deep Neural Networks (DNN).
The document highlights that crop yield prediction is a challenging task in precision agriculture, as it depends on various factors such as climate, weather, soil, fertilizer use, and seed variety. Machine learning, as a branch of artificial intelligence, offers a practical approach to improve yield prediction by analyzing patterns and correlations in datasets.
The systematic literature review followed a defined protocol and aimed to provide an overview of the existing research on machine learning in crop yield prediction. The review found that previous survey studies did not systematically review the literature in this area. The document also presents the research questions defined for the review and the search strategy used to retrieve relevant studies from the electronic databases.
The results of the review are presented in tabular form, which shows the selected publications, their publication year, title, and algorithms used. The distribution of the selected publications per year is shown in Figure 4, indicating an increasing number of papers published in recent years.
The document also provides details on the methodology of the review, including the exclusion criteria used to filter out irrelevant studies. The exclusion criteria include factors such as publications not related to agriculture and yield prediction, non-English publications, duplications, and review papers.
In conclusion, this systematic literature review provides a comprehensive overview of the use of machine learning in crop yield prediction. It identifies the most commonly used features and algorithms and highlights the need for

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

further research in this field. The review also emphasizes the importance of machine learning in improving yield prediction and supporting decision-making in precision agriculture.

Learning's

Use of Machine Learning Algorithms: The document discusses the use of various machine learning algorithms such as linear regression, elastic net, k-nearest neighbor, support vector regression, decision tree, random forest, XGBoost regression, and more. These algorithms are applied to predict crop yields based on different input parameterssuch as rainfall, temperature, crop type, and soil characteristics.

Deep Learning Techniques: The document also explores the use of deep learning techniques, including Convolutional Neural Networks (CNN), Long-Short Term Memory (LSTM), and Deep Neural Networks (DNN). These techniques are employed to improve the accuracy of crop yield prediction by extracting relevant features from satellite data, considering spatial and temporal information, and reducing networkcomplexity.

Importance of Data Preprocessing: Data preprocessing techniques are crucial in crop yield prediction to ensure data quality and eliminate redundant features, null values, and outliers. Preprocessing steps include normalization of numerical data, featureextraction, and selection to enhance the performance of prediction models.

Evaluation Metrics: The document emphasizes the use of evaluation metrics such as Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), coefficient ofdetermination (R2), mean absolute error (MAE), and root mean square error (RMSE) to assess the performance of crop yield prediction models. These metrics help in comparing different algorithms and techniques and determining the accuracy ofpredictions.

Integration of Environmental Factors: Several studies highlight the importance of incorporating environmental factors such as rainfall, temperature, soil type, and vegetation indices in crop yield prediction models. These factors significantly impactcrop growth and production, and their inclusion improves the accuracy of predictions.

Potential Applications: The research demonstrates that accurate crop yield prediction models can assist farmers and agronomists in making informed decisions about crop management activities, resource allocation, and financial planning. These models cancontribute to improving crop productivity, food security, and economic growth in the agricultural sector.

In summary, the document provides insights into the use of machine learning and deep learning techniques for crop yield prediction. It highlights the significance of data preprocessing, the selection of appropriate algorithms, and the integration of environmental factors in improving prediction accuracy. The learnings from these studies can be applied to develop more accurate and efficient crop yield prediction models, benefiting farmers, agricultural agencies, and global food security initiatives.

IV. CONCLUSION AND FUTURE WORK

In conclusion, the project focused on designing and developing a crop yield prediction model using Convolutional Neural Networks (CNN). The systematic methodology involved data collection, preprocessing techniques, and the implementation of the CNN model. A classification model was built to categorize the data based on attributes such as state, district, season, and crop. Additionally, a yield prediction model was developed using the CNN architecture and trained with the preprocessed dataset. The performance analysis showed promising results, with improved accuracy in predicting crop yield. Evaluation metrics such as Mean Absolute Percentage Error (MAPE) and Mean Squared Error (MSE) were used to assess the model's performance.

There are several areas of improvement and expansion for future work. Firstly, exploring the use of more advanced deep learning techniques and architectures can further enhance the accuracy of the crop yield prediction model. Incorporating more diverse and comprehensive datasets from different regions and countries can improve the generalization capability of the model. Integrating real-time weather data and other environmental factors can provide more accurate predictions by considering the impact of natural forces on crop production. The user interface can be enhanced to provide a more user-friendly experience, allowing users to input various parameters and visualize the predicted crop yield in a more intuitive manner. Additionally, future work could involve the incorporation of climatic data and higher-resolution satellite data, as well as the use of spiking neural networks. Overall, with further research

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

and development, the crop yield prediction model using CNN has the potential to contribute to better agricultural planning, food management, and economic growth in the agriculture sector.

REFERENCES

[1]. Abbas, F., Afzaal, H., Farooque, A. A., & Tang, S. (2022). Crop Yield Prediction through Proximal Sensing and Machine Learning Algorithms.

[2]. Jahić, A., & Buzađija, N. (2023). Multispectral Crop Yield Prediction Using 3D-Convolutional Neural Networks and Attention Convolutional LSTM Approaches.

[3]. Sharma, P., Dadheech, P., Aneja, N., & Aneja, S. (2022). Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning.

[4]. van Klompenburg, T., Kassahun, A., & Catal, C. (2021). Crop yield prediction using machine learning: A systematic literature review.

[5]. Vignesh, K., Askarunisa, A., & Abirami, A. M. (2023). Optimized Deep Learning Methods for Crop Yield Prediction.