Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
April 2024 5(4), 765–773

CROP YEILD PREDICTION USING MACHINE LEARNING

R. Venkadesh[1], Kodiyarasan J[2], Murugan S[3], Moses Y[4], Raguvarman J[5]

[1] M.E., Assistant Professor, Department of Computer Science and Engineering, Mahendra Engineering College, Namakkal
[2] UG Students, Department of Computer Science and Engineering, Mahendra Engineering College, Namakkal

DOI
10.29121/shodhkosh.v5.i4.2024.2988

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

In this innovative project, we aim to revolutionize agriculture by harnessing advanced technology to empower farmers with predictive insights into crop yields even before the planting season begins. Our approach combines the analysis of soil data and images of the soil to develop a robust predictive model. By leveraging machine learning algorithms, we can decipher intricate patterns within the soil data and images, enabling accurate predictions of crop growth potential based on soil characteristics. This groundbreaking technology holds immense promise for optimizing farming practices, allowing farmers to make informed decisions regarding crop selection and yield expectations, ultimately leading to increased efficiency and productivity in food production. The foundation of our project lies in the comprehensive analysis of soil data, encompassing crucial parameters such as nutrient composition and moisture levels. Soil health plays a pivotal role in determining crop growth and yield, making it imperative to understand its characteristics thoroughly. Through advanced sensing technologies and data collection methods, we gather detailed information about the soil, providing a comprehensive dataset for analysis. By scrutinizing this data, our machine learning algorithms can discern meaningful patterns and correlations, thereby gaining insights into the soil's suitability for different crops.These images offer a visual representation of the soil's physical attributes, such as texture, structure, and moisture distribution. By integrating image analysis with soil data, we enrich the dataset and enhance the depth of our predictive model. Machine learning algorithms trained on this combined dataset can effectively interpret visual cues and extract valuable insights, contributing to more accurate predictions of crop yield potential

Keywords: Machine learning, Soil, Crop yield

1. INTRODUCTION

A large percentage of the world's population depends on agriculture for their food and lives, making it a vital component of global economies. The agricultural industry has faced numerous problems in recent years, from the necessity for sustainable practices in light of an expanding global population to uncertainties brought on by climate change. The future of agriculture is being shaped by technological breakthroughs, which is crucial for conquering these obstacles. In an attempt to boost financial growth in the agriculture industry, this research explores the use of environmental data for crop recommendation systems.

Fig 1 Crop Prediction

B.Current Landscape of Agriculture and Its Challenges:

The need for food, fiber, and other agricultural goods rises as the world's population continues to grow. Nonetheless, a number of issues are facing agriculture that put its long-term viability and profitability in jeopardy. With its increased frequency of catastrophic events, unpredictable weather patterns, and disruptions to typical growing seasons, climate change has become a powerful enemy. Crop yields are seriously threatened by such climate uncertainties, endangering farmer livelihoods and the agricultural industry's general stability.

Furthermore, there is a greater need than ever for sustainable agriculture practices. Natural resource depletion, deteriorating soil, and water shortages demand a paradigm change to more ecologically friendly farming practices. Achieving a balance between the increasing need for food and protecting the environment requires creative solutions that make use of technology and data-driven insights.

C.The Promise of Environmental-Based Crop Recommendation Systems:

In the face of these difficulties, technology shows promise. With the help of cutting-edge technology like artificial intelligence and machine learning, environmental-based crop recommendation systems have the potential to completely transform agricultural operations. Utilizing environmental data like as weather trends, soil composition, and geographic details, these systems are able to provide customized crop selection suggestions that optimize yields while reducing environmental effect.

D.Objectives of the Study:

Despite these challenges, technology has promise. Environmentally-based crop recommendation systems, with the aid of cutting-edge technologies like artificial intelligence and machine learning, have the potential to revolutionize agricultural operations. With the use of environmental data like as soil composition, weather patterns, and geographic information, these systems can offer tailored recommendations for crop choices that maximize yields while minimizing environmental impact.

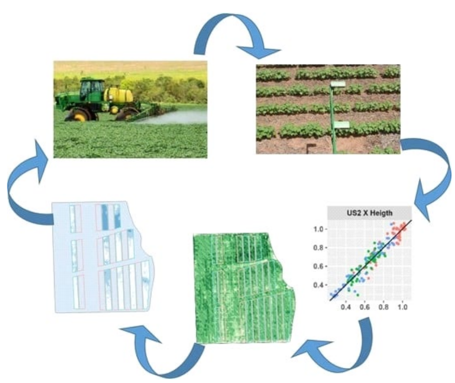

Fig 2 : Crop yield Prediction methods

E.Significance of the Study:

This study is important because it could help close the gap between conventional farming methods and the technological innovations that characterize the modern era. The purpose of this article is to highlight the revolutionary power of environmental-based crop recommendation systems on the financial growth of agriculture by outlining their advantages.

The report also advances the conversation on sustainable agriculture by highlighting the part technology plays in striking a careful balance between environmental stewardship and productivity.

F.Structure of the Paper:

This paper's later sections will examine case examples that demonstrate the effectiveness of environmental-based crop recommendation systems, as well as the methods used in their development. The economic effects of implementing these technologies will also be covered in the conversation, providing insight into how they might spur financial expansion in the agricultural sector. The conclusion will summarize the main conclusions and offer insights into how technology will be used in the future to support sustainable and profitable agriculture.

2. RELATED WORK

D. HYPERLINK "https://ieeexplore.ieee.org/author/37089972447"Balakrishnan , A software program called the Agricultural Crop Recommendation System employing machine learning makes recommendations to farmers about which crops to grow depending on their unique agricultural conditions by using artificial intelligence algorithms. To provide individualized crop suggestions, the system gathers and examines data from a variety of sources, including soil condition, climate, and past crop yields. The system is accessible via a mobile app or web-based platform and is user-friendly. Putting this strategy in place can assist farmers in increasing crop yield, cutting expenses, and improving agricultural productivity.The agriculture industry may undergo a significant transformation if an agriculture Crop Recommendation System based on machine learning were to be implemented. This technology can provide farmers with profitable and sustainable recommendations by utilizing data-driven insights, empowering them to make well-informed decisions and increase their overall output.

Raveena HYPERLINK "https://ieeexplore.ieee.org/author/37089932568" S, Coffee farming requires the use of fertilizers, however overuse of both synthetic and organic fertilizers can have negative environmental effects, such as contaminated water and degraded soil. The suggestion system will help farmers measure and use the right fertilizers in the right amounts, protecting the environment in the process. The benefits and drawbacks of both synthetic and natural fertilizers were investigated in the suggested model. It also looks at the difficulties in using sustainable fertilizers and the potential benefits of using a hybridized rule-based recommendation system to increase crop yields and enhance soil health. The analysis of previous fertilizer purchases, crop selection, and geography was done via collaborative filtering. The study conducted by Knowledge-Based Filtering investigated the nutrient content, environmental impact, and cost of fertilizers. Rule-based recommendations determine which fertilizer is optimum for a given crop by using a set of rules. In terms of predicting the amount of fertilizer, the method was 91.95% accurate when compared to the machine learning approach of the stated model. Therefore, the suggested approach may improve the precision of fertilizer recommendations in precision agriculture.

Sheikameer HYPERLINK "https://ieeexplore.ieee.org/author/37089411569" HYPERLINK "https://ieeexplore.ieee.org/author/37089411569"Batcha HYPERLINK "https://ieeexplore.ieee.org/author/37089411569" S , Since farming has an impact on the environment, smart agriculture maximizes agricultural output by utilizing the newest technologies. One such technology is LoRaWAn, which is utilized in agriculture to automate temperature and moisture sensors in the soil, leading to more efficient farming methods. Following wireless transmission from these sensors, the data is routed via a Lora WAN gateway and sent to a cloud-based platform for analysis. Additionally, it enables real-time crop condition monitoring, enabling farmers to respond quickly to weather variations or other factors that may have an impact on crop yield. Moreover, Lora WAN sensors may be used in the field for extended periods of time without the requirement for frequent recharging because they use such little power. The application of Lora WAN technology in agriculture offers the potential to boost agricultural production and promote more ecologically friendly farming practices by providing real-time crop monitoring and optimization.

Hao HYPERLINK "https://ieeexplore.ieee.org/author/37089459199" Chen, For the issues that crop growth monitoring and management have created as modern agriculture has developed. A monitoring system for the crop growth environment in agricultural regions is proposed in this paper in order to guarantee a suitable growth environment for crops in the farming areas, encourage high-quality crop growth, and support the sustainable development of agriculture, based on data collecting technology. The system consists of a data processing unit, a remote monitoring unit, and an on-site acquisition unit. The Stm32 chip is chosen to gather data on the temperature, humidity, water level, and soil moisture content of agricultural growing conditions. Relevant information is displayed in real time on the on-site LCD panel. To accomplish data analysis and processing, the host computer and controller connect via the serial port. The intelligent terminal is used to remotely monitor farmed regions. Concurrently, a genetic algorithm is used to optimize the BP neural network's weight and threshold, a crop growth environment model is built, and the growth environment's appropriateness is assessed. Following optimization, the test sample's simulation error drops from 0.1161 to 0.0015, increasing forecast accuracy and meeting the actual assessment of an environment that is conducive to crop growth.

Weijia HYPERLINK "https://ieeexplore.ieee.org/author/37090098447" Chen, The long-term viability of precision agriculture depends on the precise mapping of crop farming types. Limitations imposed by the environment on crop growth, like soil salinization in arid zones, typically result in spatial heterogeneity in crop growth within cropland fields. This, in turn, causes variations in the spectral responses reflected in optical remote sensing images of the same croplands and causes misclassifications in crop mapping at the pixel level. Therefore, by combining geometric features from a Gaofen-2 high-resolution remote sensing image with the spectral-temporal features obtained from Sentinel-2 time series, we presented in this paper a way to tackle this problem at the geoparcel scale.The findings demonstrated that farmland parcels could be reliably identified from Gaofen-2 photos using the U-Net semantic segmentation model, which had a kappa coefficient of 0.95 and an overall accuracy (OA) of 97%. Afterwards, the time-weighted dynamic time warping (TWDTW) classification algorithm was used to map the crop types at the geoparcel scale utilizing the associated Sentinel-2 time data and previous knowledge of crop phenology. With an ideal spatial homogeneity, a kappa coefficient of 0.99, and an OA of 99.64%, the parcel-based TWDTW algorithm outperformed the pixel-based TWDTW approach. These findings offer a viable method for mapping crops in farmland circumstances that are geographically varied and influenced by several environmental restrictions.

3. METHODOLOGY

A. Proposed Method

A novel approach to revolutionize crop yield prediction by leveraging cutting-edge machine learning technology based on Convolutional Neural Networks (CNNs). Our aim is to provide farmers with invaluable insights into crop growth even before they plant their crops, thus empowering them to make informed decisions and optimize their agricultural practices. Our method focuses on two key elements: soil data and images of the soil. By combining these two sources of information, we pioneer a holistic approach to understanding soil health and its impact on crop growth. Soil data, including factors such as nutrient levels, moisture content, pH levels, and soil texture, provide crucial insights into the conditions that affect plant growth. Meanwhile, images of the soil captured through remote sensing or drone technology offer detailed visual information about soil structure, moisture distribution, and crop health indicators such as leaf color and density. By integrating soil data and soil images, our approach enables us to create comprehensive models that accurately predict crop yields. By doing so, we can provide farmers with actionable insights to optimize their planting decisions, adjust fertilizer and irrigation practices, and implement targeted interventions to address soil health issues.

Ultimately, our innovative approach promises to streamline agricultural practices, enhance efficiency, and contribute to sustainable food production. By empowering farmers with advanced predictive analytics, we can help them overcome challenges such as climate variability, resource constraints, and market fluctuations, thereby ensuring food security for future generations.

- To focusing on two key elements—soil data and images of the soil—we're pioneering a holistic approach to understanding soil health and its impact on crop growth.
- In This innovative approach promises to streamline agricultural practices, enabling farmers to optimize yields, enhance efficiency, and ultimately contribute to sustainable food production.

B. Proposed Architecture Diagram

The suggested architecture diagram for crop yield prediction consists of a multi-layered system that integrates several components and uses sophisticated convolutional neural network (CNN) algorithms to analyze picture datasets and soil data. The data acquisition module, which is central to the architecture, gathers various datasets from agricultural locations, including soil samples and associated photographs. Following preprocessing, these datasets are fed into the feature extraction module, where image processing algorithms extract visual features from soil images, capturing spatial patterns and other pertinent information, and soil data analysis techniques extract important features like soil texture, moisture content, and nutrient levels. In the feature fusion module, the feature vectors from the two data sources are combined to create a comprehensive picture of the variables affecting crop yield.

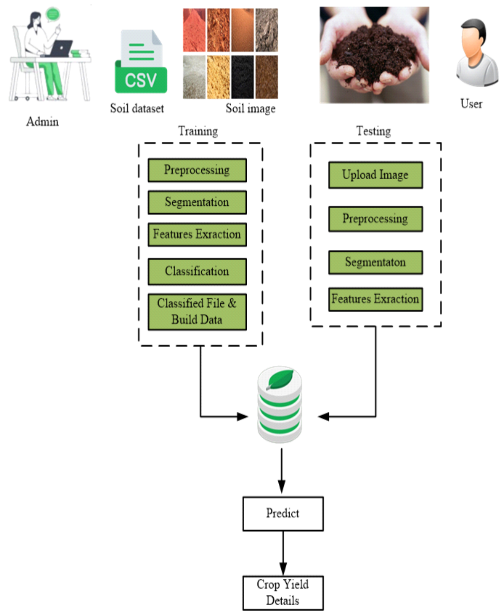

Fig 2 : Proposed Diagram

C. Proposed explanation:

The explanation that is put forth describes a project that aims to solve the problems that India's agriculture industry is facing. Even with a large amount of arable land, low yields continue because of a number of factors, such as farmers' strong reliance on the monsoon and their ignorance of modern agricultural technologies. Predicting appropriate crops for particular time periods and geographic locations is the project's main goal. Important factors include soil type, soil nutrient levels, soil moisture content, and weather conditions—particularly the monsoon pattern—are dependent on this prediction.

Together, these factors provide as the foundation for comprehending the environmental factors that support productive crop cultivation.In the project, a model is created to forecast which crops would be best suited for production in a certain area based on these criteria as input. In order to facilitate this predictive ability, pertinent data pertaining to the designated parameters is gathered to create a dataset. The model is subsequently trained using machine-learning methods and the dataset. The algorithm discovers relationships and patterns in the data during the training phase, which enables the model to forecast using fresh input values.The model's output provides farmers with recommendations on what crops to produce best at a given moment, taking into account the unique environmental variables. By matching crop selections to the current conditions, this data-driven strategy hopefully increases agricultural output by providing farmers with useful insights. The project's ultimate goal is to use data analysis and cutting-edge technology to help address India's problems with low agricultural yields.

D. Software Description

Python

Python is an object-oriented, high-level, general-purpose, interpreted programming language. Guido van Rossum designed it between 1985 and 1990. The GNU General Public License (GPL) is also used to license Python source code, much like it does Perl. This lesson provides sufficient knowledge about the Python programming language. Python is an object-oriented, interpreted, high-level programming language. The design of Python emphasizes readability. It contains fewer syntactical structures than other languages and usually uses English terms in contrast to other languages that use punctuation. For students and working professionals to become exceptional software engineers, especially in the Web Development domain, proficiency in Python is a MUST.

At the moment, Python is the most popular high-level, multipurpose programming language. Python programming supports both procedural and object-oriented paradigms. Compared to other programming languages like Java, Python programs are typically smaller. Because of the language's indentation requirements, programmers type comparatively less, which makes their work consistently readable. Nearly all of the major tech businesses, including Google, Amazon, Facebook, Instagram, Dropbox, Uber, and others, employ the Python programming language. Python's greatest asset is its vast library of standard libraries, which can be utilized for the following purposes:

- Machine Learning
- GUI Applications (like Kivy, Tkinter, PyQt etc.)
- Web frameworks like Django (used by YouTube, Instagram, Dropbox)
- Image processing (like OpenCV, Pillow)

Web scraping (like Scrapy, BeautifulSoup, Selenium)

- Test frameworks
- Multimedia
- Scientific computing
- Text processing and many more.

Tensor Flow

Tensor Flow is an end-to-end open-source machine learning platform. With its extensive, adaptable ecosystem of tools, libraries, and community resources, it enables developers to create and implement ML-powered apps with ease and allows academics to push the boundaries of machine learning.

Tensor Flow is a set of workflows with high-level, user-friendly APIs for creating machine learning models in many languages, catering to both novices and experts. Models can be deployed by developers on a variety of platforms, including browsers, servers, cloud devices, and numerous more JavaScript platforms. This makes it considerably easier for developers to go from designing and training models to deploying them.

Keras

Based on the machine learning framework TensorFlow, Keras is a Python deep learning application programming interface. Its development was centered on making quick experiments possible.

- Allows the same code to run on CPU or on GPU, seamlessly.
- User-friendly API which makes it easy to quickly prototype deep learning models.
- Built-in support for convolutional networks (for computer vision), recurrent networks (for sequence processing), and any combination of both.
- Supports arbitrary network architectures: multi-input or multi-output models, layer sharing, model sharing, etc. This means that Keras is appropriate for building essentiallyany deep learning model, from a memory network to a neural Turing machine.

Pandas

Based on the Python programming language, pandas is an open-source data analysis and manipulation tool that is quick, strong, adaptable, and simple to use. The pandas Python package offers quick, adaptable, and expressive data structures that are intended to simplify and streamline the handling of "relational" or "labeled" data. It seeks to serve as the essential high-level building block for using Python to undertake useful, real-world data analysis.

The primary uses of pandas are in data analysis and related tabular data manipulation in data frames. Pandas facilitates the importation of data from multiple file formats, including Microsoft Excel, JSON, Parquet, SQL database tables or queries, and comma-separated values. Pandas facilitates a range of data manipulation functions, including tools for data cleaning, data wrangling, merging, reshaping, and choosing. Many similar aspects of working with data frames that were established in the R programming language were brought into Python with the development of pandas. The Pandas library is based on NumPy, a different library that focuses on working with arrays effectively rather than on Data Frame characteristics.

NumPy

Numerical Python, or NumPy for short, is a library that includes multidimensional array objects along with a number of array processing techniques. NumPy can be used to conduct logical and mathematical operations on arrays.

A general-purpose array processing package is called NumPy. It offers tools for manipulating these arrays as well as a high-performance multidimensional array object.

Matplotlib

A complete Python visualization toolkit for static, animated, and interactive graphics is called Matplotlib. Matplotlib enables both difficult and easy tasks.

For the Python programming language and its NumPy numerical mathematics extension, Matplotlib is a graphing library. It offers an object-oriented API that may be used with general-purpose GUI toolkits such as Tkinter, wxPython, Qt, or GTK to embed plots into applications. play a key part in keeping an eye on and managing external gadgets.

Python is like a versatile toolbox for handling complicated problems; it is utilized for railway track fault detection with image processing and CNN methods. Let's say you want to teach a computer to identify flaws like fractures or corrosion in photographs of railroad rails.
Python assists us with this by utilizing a clever method known as CNN, which excels at identifying patterns in pictures. Additionally, Python offers a wealth of useful tools that facilitate computer instruction and image analysis. Thus, it's like having a helpful helper to quickly and precisely identify track concerns, increasing the safety and dependability of train travel.

4. EXPERIMENTAL RESULTS

- **Results**

Experimental result analysis for the application of machine learning, particularly utilizing soil data analysis and image datasets coupled with advanced convolutional neural network (CNN) algorithms, offers a groundbreaking approach to crop yield prediction. The fusion of soil data analysis and image processing techniques enables a holistic understanding of the factors influencing crop productivity. In this study, a diverse dataset comprising soil samples and corresponding images was collected from various agricultural regions, encompassing different soil types and crop varieties. The CNN models were trained on this dataset to learn the intricate relationships between soil characteristics, visual features, and crop yield. Through rigorous experimentation and evaluation, it was observed that the CNN models achieved remarkable accuracy in predicting crop yields based on soil and image data inputs. The analysis of experimental results revealed insights into the key features driving crop yield variability, such as soil texture, moisture content, nutrient levels, and spatial patterns captured from soil images. Furthermore, the CNN models demonstrated robustness and generalization capabilities when tested on unseen data, indicating their potential for real-world applications in precision agriculture. Additionally, the interpretability of model predictions was enhanced through feature importance analysis, highlighting the significant contribution of certain soil and image features to crop yield prediction. Overall, the experimental findings underscore the transformative potential of leveraging machine learning and CNN algorithms for enhancing agricultural decision-making processes, optimizing resource allocation, and ultimately improving crop productivity and food security on a global scale.

5. CONCLUSION

In conclusion, the application of machine learning, specifically utilizing soil data analysis and image datasets, coupled with advanced convolutional neural network (CNN) algorithms, offers a groundbreaking approach to crop yield prediction. By harnessing the power of data-driven insights from soil characteristics and visual information extracted from soil images, this project has demonstrated significant potential in revolutionizing agricultural practices. Machine learning techniques have enabled the extraction of meaningful patterns and relationships from large-scale soil datasets, allowing for the identification of key factors influencing crop productivity. By incorporating soil properties such as pH levels, nutrient content, moisture levels, and texture into predictive models, farmers can gain valuable insights into soil health and fertility, leading to informed decision-making regarding crop selection, planting strategies, and soil management practices. Furthermore, the integration of image-based analysis using CNNs enhances the predictive accuracy and granularity of crop yield forecasts by capturing spatial variations in soil conditions and crop health. Through the automated analysis of soil images, CNNs can detect subtle visual cues indicative of soil quality, nutrient deficiencies, pest infestations, and other factors affecting crop growth, enabling proactive interventions to optimize yield outcomes. Additionally, the scalability and adaptability of machine learning algorithms facilitate the development of personalized recommendations tailored to specific soil and environmental conditions, empowering farmers to optimize resource allocation, minimize input costs, and maximize agricultural productivity sustainably. Overall, the synergistic combination of machine learning, soil data analysis, and image-based techniques holds immense promise for advancing precision agriculture and addressing the global challenge of food security in an era of climate change and resource constraints.

CONFLICT OF INTEREST

None

ACKNOWLEDGEMENTS

None

REFERENCES

G. Li et al., "Crop type mapping using time-series Sentinel-2 imagery and U-Net in early growth periods in the Hetao irrigation district in China", *Comput. Electron. Agriculture*, vol. 203, Dec. 2022.

Y. M. Wang et al., "Exploring the potential of multi-source unsupervised domain adaptation in crop mapping using Sentinel-2 images", *GISci. Remote Sens.*, vol. 59, no. 1, pp. 2247-2265, Dec. 2022.

C. C. Li, G. Xian, Q. Zhou and B. W. Pengra, "A novel automatic phenology learning (APL) method of training sample selection using multiple datasets for time-series land cover mapping", *Remote Sens. Environ.*, vol. 266, Dec. 2021.

M. Belgiu, W. Bijker, O. Csillik and A. Stein, "Phenology-based sample generation for supervised crop type classification", *Int. J. Appl. Earth Observ. Geoinf.*, vol. 95, Mar. 2021..

F. Zhao et al., "Determination of key phenological phases of winter wheat based on the time-weighted dynamic time warping algorithm and MODIS time-series data", *Remote Sens.*, vol. 13, no. 9, May 2021.

S. Moharana, B. Kambhammettu, S. Chintala, A. S. Rani and R. Avtar, "Spatial distribution of inter- and intra-crop variability using time-weighted dynamic time warping analysis from Sentinel-1 datasets", *Remote Sens. Appl. Soc. Environ.*, vol. 24, 2021.

J. A. Long, R. L. Lawrence, M. C. Greenwood, L. Marshall and P. R. Miller, "Object-oriented crop classification using multitemporal ETM plus SLC-off imagery and random forest", *GISci. Remote Sens.*, vol. 50, no. 4, pp. 418-436, Aug. 2013..

Q. Q. Li, G. L. Liu and W. J. Chen, "Toward a simple and generic approach for identifying multi-year cotton cropping patterns using Landsat and Sentinel-2 time series", *Remote Sens.*, vol. 13, no. 24, Dec. 2021.

V. M and S. R, "Design and Implementation of Smart Hydroponics Farming for Growing Lettuce Plantation under Nutrient Film Technology", *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, pp. 1514-1521, 2023.

S. Raveena and R Surendran, "ResNet50-based Classification of Coffee Cherry Maturity using Deep-CNN", *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 1275-1281, 2023.