Journal For Innovative Development in Pharmaceutical and Technical Science (JIDPTS)
Volume:6, Issue:2, Feb:2023

(J I D P T S)
ISSN(O):2581-6934

ANALYSIS OF CROP YIELD PREDICTION USING RECURRENT NEURAL NETWORKS -A DEEP LEARNING APPROACH

P Prathibha[1], Ramu V[2], Hymavathy S[3]

[1]Asst Professor in Computer Science
[2&3]Asst Professor in CSE (AI & ML) Department
[1]Pingle Govt College for Women (A), Waddepally, Hanamkonda
[1&3]Balaji Institute of Technology & Science (Autonomous) Narsampet, Warangal District

Abstract: Considering it depends on so many variables, including crop genotype, environmental factors, management practices, and their interactions, predicting crop output is quite difficult. Based on environmental data and management practices, this research proposes a deep learning framework for agricultural production prediction utilizing convolution neural networks (CNNs) and recurrent neural networks (RNNs). Using historical data, the proposed CNN-RNN model was used to predict corn and soybean yield across the years 2020, 2021, and 2022. Other popular techniques included random forest (RF), deep fully connected neural networks (DFNN), and LASSO. The new model significantly outperformed all previous examined approaches, with a root-mean-square-error (RMSE) of 9% and 8% of their respective average yields. Three distinguishing characteristics of the CNN-RNN make it a potentially valuable technique for additional crop-yield prediction study.

Keywords: convolution neural networks (CNNs), Long Short Term Memory Models

Introduction

Crop yield forecasting is essential for the global production of food. Managers can use accurate forecasting to make timely import and export decisions which will enhance the country's food security. In order to create superior kinds for all types of conditions, seed companies must foresee new hybrids. Farmers and growers can simply make financial decisions that will benefit them. Interactions between a person's genotype and environment are extremely complicated traits [1]. The terms "genotype" and "phenotype" refer to an individual's observable features, such as their physical, physiological, biochemical, and behavioral traits. One type of phenotypic trait, for instance, is plant length [2].For predicting of crop yield, many studies have used machine learning techniques such as regression trees, random forest modeling, multivariate regression, multivariate association rule mining, and artificial neural networks. Crop yield is treated by machine learning models as an implicit function of input variables including weather and soil conditions, which may be a very complex and nonlinear function. A Genotype is a representation of high dimensional marker data, which frequently includes millions of markers for each plant individual [3].It is essential to estimate whether genetic markers will affect the environment, which has an impact on how field management activities will interact with various environmental elements. Each person inherits two alleles for each character, one from the male and one from the female, referred to as alleles, and they're a different term of genes.

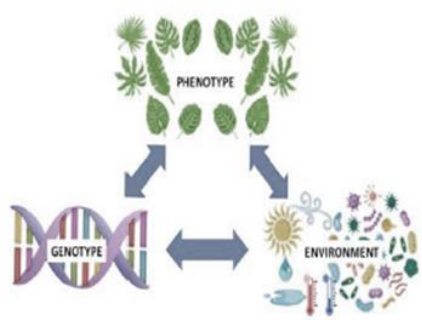


Figure 1: interactions between phenotype, genotype, and environment

It makes use of the moment simulation and solution techniques to predict agricultural production based on data on the environment and management approaches. The class of representation techniques for learning that deep learning methods belong includes multiple types of representation, each with nonlinear modules that transform what is represented at the present stage to a more or less abstract level [4].

Related Work

A review of research on estimate using machine learning has been initiated in order to collect and improve yield prediction data. A total of 567 comparable research studies have been identified after six databases were analyzed Fifty of these papers were selected chosen for additional research. Artificial intelligence takes advantages of soil moisture, and air temperature. The CNN, LSTM, and DNN deep neural networks are well-liked. For the training and testing phases of machine learning, historical data is used to assess performance. The foundation for descriptive and inferential machine learning algorithms is provided by the literature research and the study objectives [5].

Yield Prediction Using Predicted Weather Data

While it is unknown in the beginning, weather is one of the key variables in predicting crop output. As a result, predicting the weather must be a component of predicting agricultural productivity. We acquired the 2022 yield prediction results for the state of Iowa utilizing projected weather data in order to assess the effects of weather prediction on the performance of the CNN-RNN model. As a forecast for the same time period of the year 2021, we used the June to September weather records from 2022. We compared the CNN-RNN model's RMSE and anticipated state average yield utilizing perfect weather data to the model's RMSE and predicted state average yield using predicted weather data [6].Starting in June and extending through September, we updated the forecasted weather data every week with the actual ground truth 2022 weather data, and prediction results were gathered for each week. This allowed us to see the impact of weather prediction on the yield forecast with greater clarity. The predicted error decreased as we more closely matched the expected weather data with the actual weather data each week, demonstrating how weather prediction affects yield prediction [7].Large datasets and knowledge processes are involved in agriculture. The yield of the crops is predicted using a variety of ways. Massive amounts of data are processed using sophisticated technology and neural pathways [8]. In this research project, yield was determined using a replacement method. The characteristics include pH level, nitrogen and carbon content, temperature, weather patterns, soil types, and phosphate content.Maximising value is agriculture's main objective. It is crucial to understand the rainfall patterns of a chosen site since large-scale climatologically phenomena have an excessively detrimental impact on agriculture.

Methodology

Recurrent Neural Network (RNN)

The fundamental neural network is given a twist by a recurrent neural network model. A basic neural network is able to be utilized in cases if the input has the series type and had no predefined size. Recurrent neural structures retain recollections of the past and are impacted by decisions made in the past. They may take one or more inputs and produce one or more vector outputs. However, for this prediction study, genetic data wasn't made available to people in general. As a result, the model must use available data to indirectly reflect the effect of genotype [8]. A directed graph serves as a reflection of the temporal dependencies of nodes in an RNN, a type of artificial neural network.LSTM cells, which are exactly designed recurrent neurons to capture input dependencies throughout time, have been added to these RNNs to enhance performance. LSTM networks perform more compared to other time series models in an array of sequence modeling applications since they don't need the nonlinear functions to be anticipated to be provided. The RNN model possesses k LSTM cells and displayed the county's crop yield for year t using data from year's t k to t. The cell's inputs include data on average yield, management information, and the output of the FC layer, which gathered significant characteristics from the weather and soil data and processed them using the W-CNN and S-CNN models [9]. The S-CNN and FC models were created expressly for transmitting the soil data gathered at the soil surface directly to the LSTM cells.

 ISSN:-2581-6934 |Available at :www.jidps.com | pp (6-11)

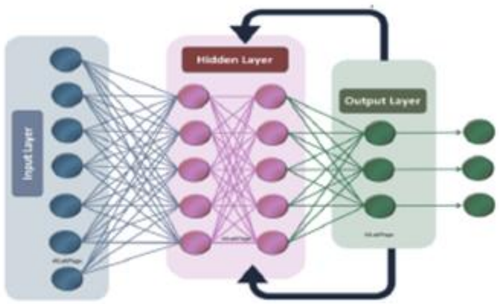


Figure 2: Recurrent Neural Network

Long Short Term Memory (LSTM)

In order to address the long-term correlation problem that arises in recurrent neural networks as a result of the vanishing gradient issues, Long Short Term Memory Models (LSTMMs) are critically built. To create a feed-forward neural network that is more traditional LSTMs include connections to feedback. The useful information about previous data is kept to help in processing the new data points. Algorithms manage the sequence of data entirely without considering every data point separately [10]. Text, speech, and general-time series LSTMs are appropriate for such data sequences. The LSTM can gather data about the current cell state with agility and care under careful control of solid frameworks known as gates. These are a way to allow data pass through it opportunistically. They consist of a point-by-point multiplication operation and a sigmoid or tanh neural net layer.

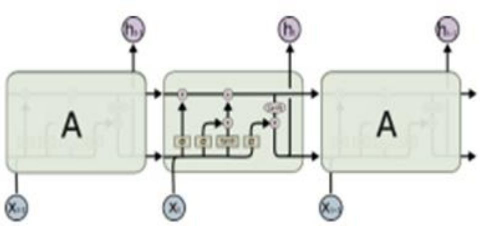


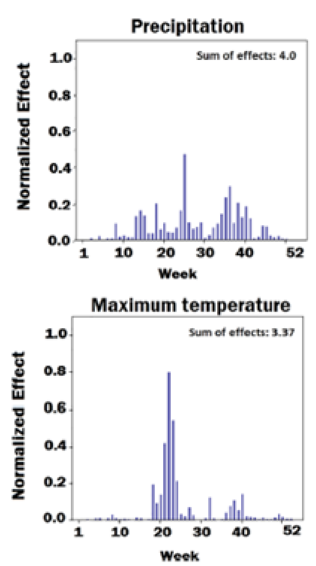
Figure 3: An LSTM's repeating module contains four interconnected layers.

Random Forest

Every decision tree, thus, has considerable variance; but, when we aggregate them all, the variance decreases because each tree is fully trained on appropriate sample information, and the result does not exclusively depend on tree structure but also on numerous call trees. The final result is generated by the bulk vote classifier in the event of a classification error [11]. The final result in the case of a regression defect is the mean of all the results.

Experimental Results

For the purpose of predicting soybean production and maize yield, respectively, we conducted two feature choices. We employed the guided back propagation method to select input variables that maximize the activation of our target neurons by backpropogating the positive gradients. First, we fed the CNN-RNN model all validation samples, and then we calculated the average activation of every neuron in the RNN cell's output at time step tithe other neurons were set to 0, and the gradient of active neurons was set to 1.Then, based on the size of the gradient, we back propagated the gradients of the activated neurons to the input space to identify the key input variables [12].The estimated impact of weather variables, soil conditions assessed at different levels and at the soil surface, and management methods. To make the consequences comparable, the effects have been normalized within each group, particularly the weather conditions, the soil features, and the management procedures. The temporal resolution of our significant analysis allows identification of crucial moments for a greater understanding of the complex agronomic system's operation. Solar radiation was the most sensitive component and snows the least sensitive factor in this study's examination of six weather variables that affect maize results.



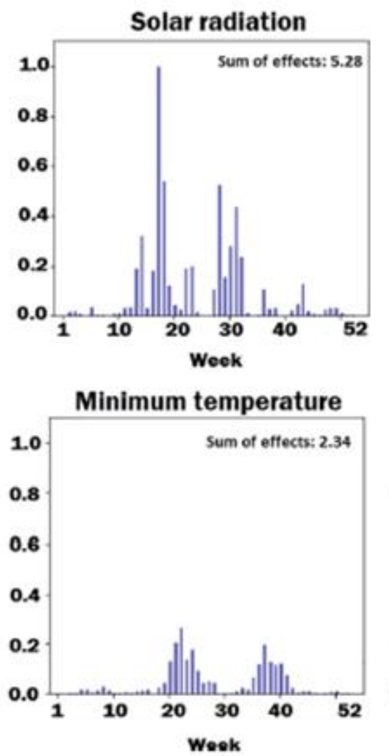
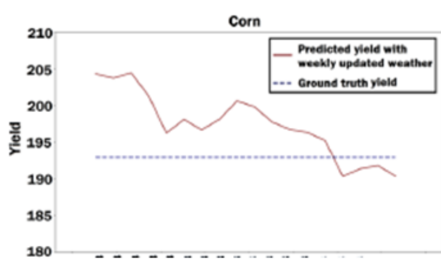


Figure 4: Estimated effects of six weather-related factors on maize, monitored for 52 weeks of every year commencing in January, are shown in a bar graph.

The RNN component of the model was created to accurately represent the rising trend in crop production over time as a result of ongoing advancements in plant breeding and management techniques. Numerous factors, including as weather, soil, and management, had a pretty large impact on the model's success [13]. Future projects requiring yield prediction could make use of the suggested approach because it successfully forecasted yields in untested conditions.



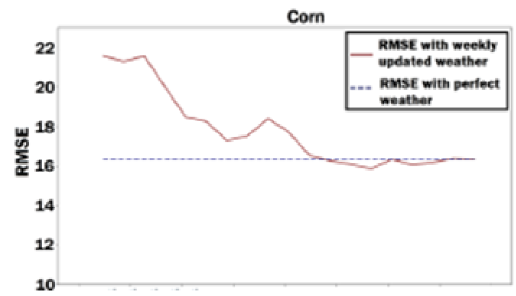


Figure 5: Every week, shown weather data was synchronized with the connected actual data. For each week, 2022 weather data and predicted results was obtained.

Conclusion

The paper includes an in-depth evaluation of all the various methods for predicting agricultural yields. The datasets and results will be discussed together with several algorithms. There is mention of many crop varieties used for annual prediction. Different classification approaches are used to predict different crops, and the accuracy of the results depends on the parameters utilized. The climatic environment has a significant impact on yield. Crop yield prediction largely depends on several elements, thus it is important to precisely take these into account when utilizing any approaches. Following so will produce effective results and allow you to make wise decisions regarding crop import and export in addition to crop pricing. This study's uniqueness was its application of the feature selection look at to evaluate the particular impacts of weather factors, soil conditions, and management variables as well as the time period over which these variables become critical.

References

[1]. Abadi,M.,Barham,P.,Chen, J.,Chen, Z.,Davis,A.,Dean, J., et al. (2016). "TensorFlow:A system for large scale machine learning," in 12th USENIX Symposium onOperating Systems Design and Implementation (Savannah,GA,USA), vol. 16. 265–283.

[2]. Andrade, F. H., Vega, C., Uhart, S., Cirilo, A., Cantarero, M., and Valentinuz, O.(1999). Kernel number determination in maize. Crop

Sci. 39, 453–459. doi:10.2135/cropsci1999.0011183x0039000200026x

[3]. Archontoulis, S. V., Huber, I., Miguez, F. E., Thorburn, P. J., Rogovska, N., andLaird, D. A. (2016). A model for mechanistic and system assessments ofbiochar effects on soils and crops and trade-offs. GCB Bioenergy 8, 1028–1045.doi: 10.1111/gcbb.12314

[4]. Awad, M. M. (2019). Toward precision in crop yield estimation using remotesensing and optimization techniques. Agriculture 9, 54. doi: 10.3390/agriculture9030054

[5]. Baum, M., Archontoulis, S., and Licht, M. (2018). Planting date, hybrid maturity,and weather effects on maize yield and crop stage. Agron. J. 111, 1–11. doi:10.2134/agronj2018.04.0297

[6]. Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencieswith gradient descent is difficult. IEEE Trans. Neural Networks 5, 157–166

[7]. Borovykh, A., Bohte, S., and Oosterlee, C. W. (2017). Conditional time seriesforecasting with convolutional neural networks. arXiv preprintarXiv:1703.04691.

[8]. Breiman, L. (2001). Random forests. Mach. Learn. 45, 5–32. doi: 10.1023/A:1010933404324

[9]. Drummond, S. T., Sudduth, K. A., Joshi, A., Birrell, S. J., and Kitchen, N. R. (2003).Statistical and neural methods for site–specific yield prediction. Trans. ASAE46, 5. doi: 10.13031/2013.12541

[10]. Egli, D., and Bruening, W. (1992). Planting date and soybean yield: evaluation ofenvironmental effects with a crop simulation model: Soygro. Agric. For.Meteorol. 62, 19–29.

[11]. S.Ramana, M.Pavan Kumar, N.Bhaskar, S. China Ramu, & G.R. Ramadevi. (2018). Security tool for IOT and IMAGE compression techniques. Online International Interdisciplinary Research Journal, {BiMonthly}, 08(02), 214–223. ISSN Number: 2249-9598

[12]. Peddyreddy. Swathi. (2022). Implications For Research In Artificial Intelligence. Journal of Electronics,Computer Networking and Applied Mathematics(JECNAM) ISSN : 2799-1156, 2(02), 25–28.

[13]. A. Yeboah-Ofori, and S. Islam, "Cyber Security Threat Modelling for Supply Chain Organizational Environments." MDPI. Future Internet. 11, (3), 63, March 2019. doi: 10.3390/611030063.

[14]. B. Woods, and A. Bochman, "Supply Chain in the Software Era" Scowcroft Center for Strategic and Security. Atlantic Council: Washington DC, USA, May 2018.

[15]. ENISA "Exploring the Opportunities and Limitations of Current Threat Intelligence Platforms" Version 1. December 2017. [online]

[16]. A. Yeboah-Ofori, J. D. Abduli, F. Katsriku, "Cybercrime and Risks for Cyber Physical Systems" International Journal of Cyber Security and Digital Forensics. Vol.8 No1, pp 43-57. 2019.

[17]. S. Ramana, N. Bhaskar, S. China Ramu, M. V. Ramana Murthy." A Two-Level Authentication Protocol for Secure M-Commerce Transactions using AMQP Protocol", Vol 2021: Issue 06, pp:844– 861.

[18]. Peddyreddy. Swathi. (2022). Industry Applications of Augmented Reality and Virtual Reality. Journal of Environmental Impact and Management Policy(JEIMP) ISSN:2799-113X, 2(02), 7–11. Retrieved from http://journal.hmjournals.com/index.php/JEIMP/article/view/453

[19]. Keerthi Vuppula, "An advanced machine learning algorithm for fraud financial transaction detection" Journal For Innovative Development in Pharmaceutical and Technical Science (JIDPTS) Volume:4, Issue:9, Sep:2021

[20]. S. G. Wu, F. S. Bao, E. Y. Xu, Y.-X. Wang, Y.-F. Chang, Q.-L. Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural

network," In Proceedings of 2007 IEEE International Symposium on Signal Processing and Information Technology, pp. 11-16, 2007.

[21]. N.-H. Pham, T.-L. Le, P. Grard, V.-N. Nguyen, "Computer aided plant identification system," In Proceedings of 2013 International Conference on Computing, Management and Telecommunications (ComManTel), pp. 134-139, 2013.

[22]. Ahmad and K. Pothuganti, "Design & implementation of real-time autonomous car by using image processing & IoT," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2020, pp. 107-113, doi: 10.1109/ICSSIT48917.2020.9214125.

[23]. Sirisati, Ranga Swamy, et al. "An Enhanced Multi-Layer Neural Network to Detect Early Cardiac Arrests." *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. IEEE, 2021.

[24]. Keerthi Vuppula, "Computer-Aided Diagnosis for Diseases using Machine Learning", International Journal of Scientific Research in Engineering and Management (IJSREM) Volume: 04 Issue: 12 | November - 2020 ISSN: 2582-3930.

[25]. PoornachanderVadicherla, MarupakaNagaraju, Dhanalakshmi Vadlakonda, "Enhanced Adaptive Ranking based Energy efficient Opportunistic Routing (EAREEOR) Scheme in Heterogeneous Wireless Sensor Network" Test Engineering & Management, Vol. 83: May/June 2020.

[26]. I. Ahmad and K. Pothuganti, "Smart Field Monitoring using ToxTrac: A Cyber-Physical System Approach in Agriculture," 2020 International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2020, pp. 723-727, doi: 10.1109/ICOSEC49089.2020.9215282.