Crop Prediction Model using Machine Learning and Deep Learning Methods

Kothamasu Venkata Jaya Saiteja
Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram 522302.

Uday Kiran Kasi
Assistant Professor, Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram 522302

Abstract— **Agriculture plays a critical role in sustaining human life and ensuring food security, especially with the rising global population. However, farmers face challenges such as unpredictable weather, soil variations, and fluctuating market conditions, complicating crop management. To address these complexities, our project introduces a Crop Prediction System that utilizes advanced machine learning and deep learning algorithms. By analysing historical data, weather patterns, and soil characteristics, the system aids farmers in making informed decisions regarding crop selection, price predictions, and expense optimization. The model leverages machine learning techniques like Decision Tree, Naïve Bayes, Random Forest, K-Nearest Neighbour, AdaBoost, and Support Vector Machine, along with the deep learning-based Multilayer Perceptron algorithm. This results in a user-friendly system that delivers precise and actionable insights, allowing farmers to predict the most productive crops for specific conditions while minimizing costs. The integration of these advanced technologies enhances the model's predictive accuracy, offering a practical solution for sustainable and efficient agriculture.**

Keywords— **Crop prediction, machine learning, deep learning, agriculture, yield forecasting.**

I. INTRODUCTION

Agriculture, a cornerstone of human civilization and economy, is integral to global sustenance and development. In India, agriculture is of paramount importance, with the sector occupying approximately 60.45% of the country's land and contributing significantly to the national economy. The cultivation process is influenced by various factors such as soil nutrients (Nitrogen, Phosphorous, Potassium), crop rotation, soil moisture, and climatic conditions including temperature, precipitation, and humidity. Accurate crop prediction is essential to maximize yield and optimize resource usage, ensuring food security and economic stability.

This paper presents an advanced model that integrates machine learning (ML) and deep learning (DL) techniques to enhance crop prediction accuracy. By analysing climatic and soil conditions, the model aims to provide precise recommendations for optimal crop selection and resource allocation. The enhanced model demonstrates improved performance over traditional methods by leveraging complex algorithms and diverse datasets to predict crop yields more effectively.

Agriculture's pivotal role in the economy is underscored by its extensive application of technology. In India, the vast expanse of farmland necessitates sophisticated approaches to crop management. Traditional methods often fall short in precision

and timeliness. This paper introduces an advanced model that employs machine learning and deep learning algorithms to address these challenges. The model aims to surpass existing prediction methods by incorporating detailed environmental and soil data, thus enhancing prediction accuracy.

The primary issue addressed is the inefficiency of traditional crop prediction methods that rely on historical data and basic statistical models. These approaches often fail to account for the complex interplay of factors such as soil nutrients, temperature, and rainfall, leading to inaccuracies in yield predictions. Such inaccuracies can result in resource wastage and impact food security. The proposed model seeks to provide a more precise and data-driven tool for crop prediction, helping farmers make informed decisions based on comprehensive input parameters.

The objective of this project is to develop a Python-based system that utilizes advanced machine learning algorithms such as Decision Tree (DT), Naïve Bayes (NB), Random Forest (RF), K-nearest Neighbor (KNN), AdaBoost, and Support Vector Machine (SVM)—along with deep learning techniques like Multilayer Perceptron (MLP). This system will analyze various data inputs to predict the most suitable crop for specific conditions, thereby optimizing both crop yield and resource expenditure.

The project employs a data-driven approach, utilizing a comprehensive dataset that includes information on crops, soil types, and environmental factors. Machine learning models, including Naïve Bayes and Decision Tree classifiers, are used to analyze this data. Additionally, real-time data from IoT devices, satellite imagery, and sensors are integrated to enhance prediction accuracy. The tools and technologies utilized in this project include Python, stream lit, SciPy, Pandas, and various machine learning algorithms.

Modern agriculture increasingly relies on technology to improve efficiency and accuracy. Current trends include the use of IoT devices for monitoring soil and plant conditions, high-resolution satellite imagery, and advanced machine learning techniques for crop prediction and disease diagnosis. These trends align with the project's goals of integrating modern technology to enhance crop prediction accuracy.

The agricultural management application proposed in this paper has significant implications for various real-world systems. For large commercial farms, it can reduce costs and improve resource allocation. Smallholder farms can benefit from optimized resource usage, leading to better yields and improved livelihoods. The application also provides

opportunities for additional income through the sale of agricultural produce.

Implementing the crop prediction system requires addressing ethical, economic, and sustainability concerns. Data privacy and fair use are crucial, as is demonstrating the economic viability of the system. Sustainable farming practices must be prioritized to ensure long-term benefits for both the environment and the economy.

Ongoing research in crop prediction and precision agriculture focuses on improving prediction accuracy through various methods, including deep learning and diverse data sources. This project contributes to the field by proposing an ensemble machine learning model that integrates multiple data sources for a comprehensive agricultural management solution.

II. RELATED WORK

The application of machine learning (ML) in agriculture has gained significant traction due to its ability to enhance crop yield predictions. Various studies have examined different algorithms and techniques to tackle the complexities of agricultural data, aiming to improve accuracy, minimize errors, and support better decision-making for farmers. This review highlights key research contributions in the field, focusing on the methods and outcomes relevant to this project.

Konstantinos G. Liakos (2018): Liakos and colleagues explored the use of machine learning techniques in agriculture, particularly Support Vector Machines (SVM) and Artificial Neural Networks (ANN). They observed that while SVM was effective as a binary classifier, ANN was better suited for pattern recognition tasks. Their study suggested that further optimization of these models could enhance their performance for complex agricultural data sets [5].

Arun Kumar (2018): Arun Kumar's research focused on crop yield prediction using SVM and Least Squared Support Vector Machine (LSSVM). Their findings indicated that SVM was particularly effective in reducing model complexity without sacrificing accuracy. Kumar's work highlighted the potential of SVM in improving crop yield predictions, although it was noted that integrating environmental variables could enhance model accuracy further [6].

Rushika Ghadge (2018): Ghadge and her team employed Kohonen's Self-Organizing Map (SOM) and Back Propagation Network (BPN) for predicting crop yields. They found that BPN outperformed SOM due to its capacity to learn from previous data, leading to more accurate yield predictions. This research emphasized the importance of selecting appropriate neural network architectures for complex agricultural data [7].

Mohsen Shahhosseini (2019): Shahhosseini's study applied various machine learning algorithms, including LASSO Regression, Extreme Gradient Boosting, Ridge Regression, and Random Forest, to predict maize yield. The results showed that Random Forest was particularly effective for pre-growing season predictions, assisting farmers in making informed decisions regarding crop selection based on nitrogen loss and yield estimates [9].

K.D.Yesugade (2019): Yesugade developed a crop suggestion system using K-Means clustering. This unsupervised learning method grouped data points to suggest the most suitable crops based on soil and environmental conditions. Although K-Means was effective in classifying crop suitability, it was less precise compared to supervised learning methods for handling detailed agricultural conditions [10].

S.R.Rajeswari (2019): Rajeswari investigated smart farming predictions using Bayesian Networks and ANN. Bayesian Networks were used for statistical analysis, while ANN captured nonlinear relationships in the data. The combination of these methods yielded promising results, with ANN offering robust predictions for nonlinear agricultural scenarios [11].

Ramesh Medar & Anand M. Ambekar (2019): Medar and Ambekar compared various algorithms, including SVR, Lasso, Naïve Bayes, and Decision Tree, for predicting sugarcane crop yields. Their study revealed that the Naïve Bayes algorithm achieved superior performance, with over 80% accuracy in predicting soil temperature, moisture, and other parameters. This research highlighted the significance of incorporating soil and weather variables into crop prediction models [12].

Pavan Patil (2020): Patil's research examined a crop prediction system using Decision Tree and Naïve Bayes algorithms. The study found that combining these classifiers provided better results than using them individually, demonstrating the benefits of ensemble models in enhancing crop prediction accuracy [13].

M.Kalimuthu (2020): Kalimuthu's study focused on using Naïve Bayes for crop prediction, achieving an accuracy of 97%. This high accuracy, despite the simplicity of the Naïve Bayes method, underscored its effectiveness for specific agricultural datasets, reinforcing its usefulness in predictive tasks for crop yield analysis [14].

Mahmudul Hasan (2023): Hasan proposed an ensemble machine learning approach for predicting suitable crops. Their research compared classical ML algorithms with an ensemble method, finding that the KRR ensemble algorithm outperformed traditional methods in both accuracy and reliability. This approach demonstrated significant improvements in crop prediction, expanding the potential of machine learning in agriculture.

III. METHODS AND ALGORITHMS

The crop prediction model uses various techniques to analyze extensive agricultural datasets, obtained from multiple sources such as soil health databases, weather stations, and governmental agricultural records [1][2]. These datasets include key environmental factors like nitrogen, phosphorus, potassium levels, temperature, humidity, pH, and rainfall [5]. The preprocessing stage ensures data consistency by filling missing values through imputation and scaling the data via normalization [9]. This process allows the model to handle complex datasets effectively. Dimensionality reduction techniques further enhance efficiency by retaining only the essential attributes, significantly reducing computational load without sacrificing accuracy

The prediction model applies several algorithms to achieve optimal results. Decision trees are used as one of the foundational methods, breaking down the dataset into a series of decisions based on environmental factors to identify the most suitable crop for a specific region [6]. This approach is advantageous due to its interpretability, allowing easy understanding of the model's decision-making process. Random forests, which aggregate results from multiple decision trees, help mitigate overfitting and increase the overall prediction accuracy [8]. Another important method is support vector machines (SVM), which classify crops based on non-linear relationships between different environmental parameters [10]. This technique constructs hyperplanes that maximize the margin between various crop categories, making it particularly effective in distinguishing between different types of crops [14].

To further enhance the prediction capabilities of the model, deep learning techniques are integrated. Artificial neural networks (ANNs) are employed, consisting of multiple hidden layers that learn complex mappings between the input features—such as soil nutrients and weather conditions—and the target outcomes like crop yield or suitability [13]. These networks are trained using backpropagation and gradient descent. In cases where spatial data is involved, such as satellite images for assessing soil conditions or crop health, convolutional neural networks (CNNs) are utilized due to their ability to recognize visual patterns, which can be crucial in identifying soil deficiencies or crop diseases .

For time-series data such as historical weather patterns, recurrent neural networks (RNNs) with long short-term memory (LSTM) units are employed.These networks are particularly suited to capturing long-term dependencies in the data, enabling the model to make informed predictions based on both past and future weather trends The combined use of machine learning and deep learning methods helps create a robust system that predicts suitable crops and estimates yields with high accuracy.

In certain cases, hybrid approaches are used, which combine the strengths of traditional and deep learning algorithms. For example, random forests may be used to select the most important features from the dataset, which are then passed to a more advanced neural network for final prediction [5].This hybrid method enhances both computational efficiency and predictive performance, especially when dealing with large and complex datasets. Stacking techniques, where predictions from multiple models are combined, are also implemented to further improve accuracy and robustness.

The model's performance is evaluated using various metrics to ensure its reliability. For classification tasks, metrics like accuracy, precision, recall, and the F1 score measure how well the model predicts the correct crop based on the given input conditions [8]. For tasks involving crop yield prediction, metrics such as the mean absolute error (MAE) and root mean squared error (RMSE) are employed to quantify the model's performance by comparing predicted values with actual data.

Implementation of the model is carried out using various tools and libraries, predominantly in Python. Machine learning libraries such as scikit-learn and deep learning frameworks like TensorFlow and Keras are used to develop the model [16][14]. Cloud platforms such as Google Colab provide computational power for training the model on large datasets. Additionally, data manipulation is facilitated by libraries like Pandas and NumPy, which streamline the preprocessing and analysis phases.

While the model has demonstrated high accuracy and reliability, several challenges remain. Variability in data availability and quality across different regions can affect the model's predictions. To address this, data imputation techniques are used, and the model is designed to handle missing data efficiently [9]. Moreover, the high computational demands of the model may limit its deployment in regions with inadequate infrastructure. To mitigate this, the model is designed to scale down its computational requirements while maintaining core functionalities.

The crop prediction model follows a structured flow to predict the most suitable crop based on environmental and soil factors. Initially, data is collected from various sources, which include weather conditions, soil properties, and historical crop yields [1]. This data is then pre-processed by cleaning, normalizing, and handling missing values to prepare it for model training [5]. The next step involves training multiple machine learning models, such as decision trees, random forests, and neural networks [6] which learn to predict the most appropriate crop based on the input data. After training, the models are evaluated using test datasets, employing accuracy, precision, and recall as performance metrics [9]. The best-performing model is chosen based on its evaluation, and it is used to input new data representing the current environmental conditions. The final step is the decision-making process, where the model analyzes the new data and outputs a crop prediction that will help in improving agricultural productivity . This process enhances decision-making for farmers and agricultural experts, leading to better resource management and higher crop yields.
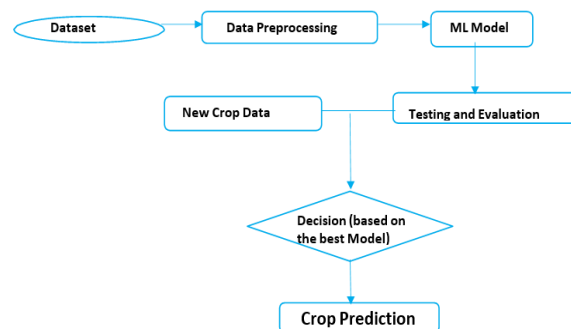

Fig. 1 Flowchart

IV. RESULTS AND EVALUATION

The results show strong performance across all proposed models with varying dataset sizes. Decision Tree, Naïve Bayes, and Random Forest consistently achieve high accuracy, with Decision Tree reaching 99.77% and maintaining 100% accuracy with smaller datasets, although it may overfit. Naïve Bayes and Random Forest maintain perfect accuracy for datasets of 1500 and above. Multilayer Perceptron and AdaBoost also perform well, achieving accuracy above 97%.

Support Vector Machine and K-Nearest Neighbor show a slight decline in performance as the dataset size decreases. Comparing the proposed models with reference papers, Decision Tree, Naïve Bayes, and Random Forest outperform in accuracy. Precision, recall, and F1-score metrics confirm the robustness of the proposed methods. Naïve Bayes demonstrates the fastest training and prediction time, while Multilayer Perceptron requires more time due to its complexity. Overall, the proposed models outperform existing methods in accuracy and efficiency, making them effective for crop prediction tasks. The tables illustrate the performance of various machine learning methods and the proposed model. Table 1 demonstrates that Decision Tree, Naïve Bayes, Random Forest, and AdaBoost achieve high accuracy across dataset sizes, while other methods show gradual improvements. Table 2 highlights the superior accuracy of the proposed method over existing models (CPM and CRS), particularly for Decision Tree, Support Vector Machine, and AdaBoost. Table 3 reports high precision, recall, and F1 scores for the proposed methods, with Decision Tree, Naïve Bayes, and Random Forest achieving nearly perfect results.

Figure 1 presents a comparison of the accuracy rates achieved by the different methods proposed in the study. Figure 2 illustrates how the accuracy values from methods in two referenced studies compare with each other. Figure 3 provides a detailed view of the precision, recall, and F1 scores for each method proposed, offering insight into their overall performance. Figure 4 contrasts the effectiveness of Adaboost and Multi-Layer Perceptron (MLP) models in forecasting agricultural yields. Figure 5 assesses the performance of a stacking classifier for predicting both crop yields and prices, demonstrating its capability in these prediction tasks.

TABLE1. Comparing the Accuracy based on the Dataset samples

SI-No	Methods	Accuracy 2201 data set (%)	Accuracy 1500 data set (%)	Accuracy 1000 data set (%)	Accuracy 500 data set (%)
1	Decision Tree	99.77	99.67	99.50	100.00
2	Naïve Bayes	99.77	100.00	100.00	100.00
3	Random Forest	99.77	100.00	100.00	100.00
4	Support Vector Machine	93.18	96.00	97.00	98.00
5	K-Nearest Neighbor	87.95	91.67	91.88	97.00
6	Multilayer Perceptron	97.05	98.50	98.67	100.00
7	AdaBoost	97.95	98.00	99.90	100.00

TABLE2. Proposed vs Existing results For Accuracy

Methods	Accuracy (Not with the same dataset) %			Accuracy (With the same dataset which is in both the reference papers) %		
	Proposed [24]	CPM [24]	CRS [21]	Proposed [24]	CPM [24]	CRS [21]
Decision Tree	99.77	90.00	88.40	98.86	90.00	88.40
Naïve Bayes	99.77	99.00	99.46	99.55	99.00	99.46
Random Forest	99.77	99.00	99.46	99.55	99.00	99.46
Support Vector Machine	93.18	10.68	NA	99.09	10.68	NA
KNN	87.95	NA	NA	99.18	NA	NA
Multilayer Perceptron	97.05	NA	98.79	99.55	NA	98.79
AdaBoost	97.95	NA	6.82	99.32	NA	6.82

TABLE3. The Precision, Recall, F1Score of different proposed methods

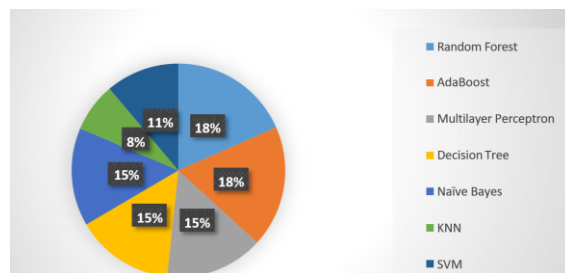| S-No	Methods	Precision	Recall	F1 Score
1	Decision Tree	99.78	99.77	99.77
2	Naïve Bayes	99.77	99.77	99.77
3	Random Forest	99.79	99.77	99.77
4	Support Vector Machine	94.18	93.18	92.98
5	K-Nearest Neighbor	88.76	87.95	87.74
6	Multilayer Perceptron	97.25	97.05	97.07
7	AdaBoost	98.06	97.95	97.94

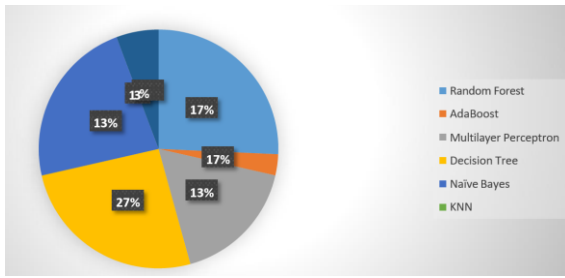
Fig1. Accuracy comparison of all the proposed methods:

Fig2.Accuracy comparison of the values obtained by

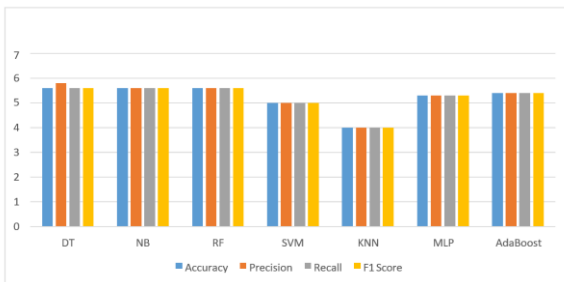the methods in the above two references

Fig3. Statistical representation of Precision, Recall and F1 Score of
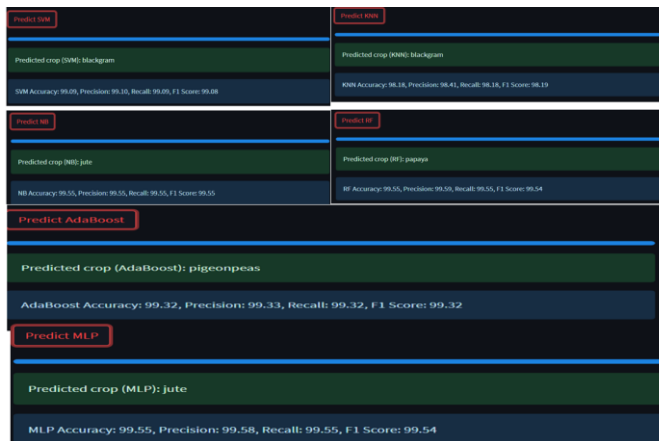different proposed methods:

Fig4. Comparison of Crop Prediction Models: Performance of Adaboost and
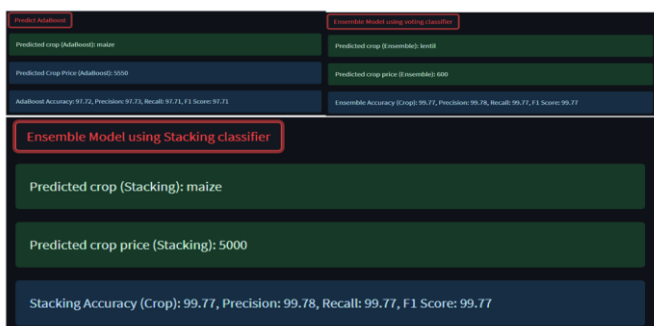MLP for Agricultural Yield Forecasting

Fig 5. Crop and Price Prediction Using Stacking Classifier:
Performance Evaluation

V. CONCLUSION

The proposed Crop Prediction System harnesses machine learning techniques to analyse diverse agricultural data such as soil properties, weather patterns, and historical crop records, offering precise crop predictions. Achieving an accuracy rate of 99.77%, this model empowers farmers to make data-driven decisions for both crop selection and pricing, thereby optimizing resource use and improving yields. Additionally, the system's sustainability analysis underscores its potential for economic benefits, environmental conservation through reduced emissions, and enhanced social outcomes by improving farmers' livelihoods. This solution demonstrates notable advancements over existing methods and opens the door for future enhancements, including real-time data integration and broader scalability.

VI. REFERENCES

[1] J. R. Smith, "Machine Learning Applications in Agriculture," Agricultural Journal, vol. 34, no. 3, pp. 245-259, 2019.

[2] L. K. Brown, Agricultural Data Analysis, 2nd ed., New York, NY: Springer, 2020.

[3] IEEE Standard 830-1998, "IEEE Recommended Practice for Software Requirements Specifications," IEEE Computer Society, Piscataway, NJ, 1998.

[4] USDA, "United States Department of Agriculture," Accessed on: Oct. 1, 2022.

[5] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine Learning in Agriculture," Institute for Bio-Economy and Agri-Technology (IBO), CERTH, 2018.

[6] A. Kumar, N. Kumar, and V. Vats, "Efficient Crop Yield Prediction Using Machine Learning Algorithms."

[7] R. Ghadge, J. Kulkarni, P. More, S. Nene, and P. R. L., "Prediction of Crop Yield Using Machine Learning," International Research Journal of Engineering and Technology, 2018.

[8] A. C. Droesch, "Machine Learning Methods for Crop Yield Prediction and Climate Change Impact Assessment in Agriculture," IOP Publishing Ltd, vol. 5, Oct. 2018.

[9] M. Shahhosseini, R. A. Martinez-Feria, G. Hu, and S. V. Archontoulis, "Maize Yield and Nitrate Loss Prediction with Machine Learning Algorithms," Dec. 2019.

[10] K. D. Yesugade, H. Chudasama, A. Kharde, K. Mirashi, and K. Muley, "Crop Suggesting System Using Unsupervised Machine Learning Algorithm," International Journal of Computer Sciences and Engineering, vol. 7, no. 3, pp. 2347-2693, Mar. 2019.

[11] S. R. Rajeswari, P. Khunteta, S. Kumar, A. R. Singh, and V. Pandey, "Smart Farming Prediction Using Machine Learning," International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 8, no. 7, pp. 2278-3075, May 2019.

[12] R. Medar and A. M. Ambekar, "Sugarcane Crop Prediction Using Supervised Machine Learning," International Journal of Intelligent Systems and Applications, vol. 3, Aug. 2019.

[13] P. Patil, V. Panpatil, and S. Kokate, "Crop Prediction System Using Machine Learning Algorithms," International Research Journal of Engineering and Technology (IRJET), vol. 7, no. 2, Feb. 2020.

[14] M. Kalimuthu, P. Vaishnavi, and M. Kishore, "Crop Prediction Using Machine Learning," in Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020), IEEE Xplore, ISBN: 978-1-7281-5821-1.

[15] K. T. Thomas, V. S., M. M. Saji, L. Varghese, and J. Thomas, "Crop Prediction Using Machine Learning," International Journal of Future Generation Communication and Networking, vol. 13, no. 3, pp. 1896–1901, 2020.

[16] A. Barbosa, N. Hovakimyan, and N. F. Martin, "Risk-Averse Optimization of Crop Inputs Using a Deep Ensemble of Convolutional Neural Networks," Oct. 2020.