Technical Report: AgriAI: Empowering Farmers with Predictive Crop Analysis and Disease Management

DARSHAN BR, Lakehead University, Canada
SANNDEEP NARINDI, Lakehead University, Canada

Machine learning (ML) and deep learning (DL) are pivotal tools in the domain of agriculture, playing crucial roles in crop yield prediction and disease detection, respectively. These technologies facilitate decisions that span from selecting the most suitable crops to managing them effectively throughout the growing season, as well as identifying and managing diseases that may affect crop health. A plethora of ML and DL algorithms have been harnessed in these fields, each offering unique insights and capabilities. This study, with its focus on identifying the most frequently employed models for crop yield prediction and disease detection, aims to contribute significantly to the existing body of knowledge. By applying these models to publicly available datasets, the study seeks to provide valuable insights and predictions that can be utilized by farmers and agricultural experts alike. Furthermore, the study endeavors to craft user-friendly interfaces that are specifically tailored to meet the needs of farmers, ensuring that the insights and predictions generated by the models are easily accessible and comprehensible. Through these efforts, the study aims to empower farmers with the tools and knowledge they need to make informed decisions and optimize their crop yields while effectively managing crop diseases.

Additional Key Words and Phrases: Machine Learning, Crop yield prediction, Systematic Literature Review

ACM Reference Format:
Darshan BR and Sanndeep Narindi. 2018. Technical Report: AgriAI: Empowering Farmers with Predictive Crop Analysis and Disease Management. 1, 1 (April 2018), 11 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Crop yield prediction is the process of calculating the amount of crops that will be harvested in a certain agricultural region, which is usually defined in terms of quantity per unit area (for example, bushels per acre or tons per hectare) [3]. Crop yield prediction is performed using a variety of methodologies, including traditional statistical approaches as well as modern technology such as ML and DL. ML is an area of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that allow computer systems to improve their performance on a specific job by learning from data without being explicitly programmed. The core idea is to enable machines to discover patterns, make predictions, or take actions based on input data rather than relying on predetermined rules. Explicit instructions are used in traditional programming to specify the reasoning and decision-making procedures involved in a particular activity [2].

In contrast, a ML system adjusts its behavior to maximize a predetermined goal by learning from examples and experiences. Three primary categories are commonly used to describe this learning process: Supervised Learning: The algorithm is trained using a labeled dataset in which the input data is coupled with the relevant output labels. The model learns to translate input data to the correct output, allowing it to anticipate new, previously unseen data. Unsupervised Learning: The algorithm is given unlabeled data and must identify patterns or structures within it without explicit supervision. Clustering and dimensionality reduction are two prominent tasks in unsupervised learning. Reinforcement Learning: The algorithm learns by interaction with the environment and feedback in the form of rewards or penalties. The goal is to learn a policy that maximizes the total reward over time. ML algorithms employ a variety of methodologies, including Linear Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Neural Networks (NN), and more. These algorithms use statistical models to generalize from training data to make predictions or conclusions about new, previously unknown data. Overall, the goal of ML is to enable computers to learn and develop via experience, allowing them to execute complex tasks and make intelligent decisions without having to be explicitly programmed for each circumstance.DL is a branch of ML that focuses on the usage of artificial NN with numerous layers, sometimes known as deep neural networks [1].

In our research, we have conducted a comparative analysis of ML and DL algorithms within the agricultural domain. Given the evolving nature of ML/DL, our objective is to assess the current landscape in agriculture. Specifically, we will compare the effectiveness of the most commonly used algorithms for predicting crop yield and detecting crop diseases. To select the algorithms for comparison, we conducted a survey on AI in agriculture, the findings of which serve as our baseline. Following the comparative analysis, our future plans involve developing a user-friendly website tailored for farmers based on the insights gained from the study.

Authors' addresses: Darshan BR, Lakehead University, 1 Thørväld Circle, Thunderbay, Canada, dbilasan@lakeheadu.ca; Sanndeep Narindi, Lakehead University, Thunderbay, Canada, snarindi@lakeheadu.ca.

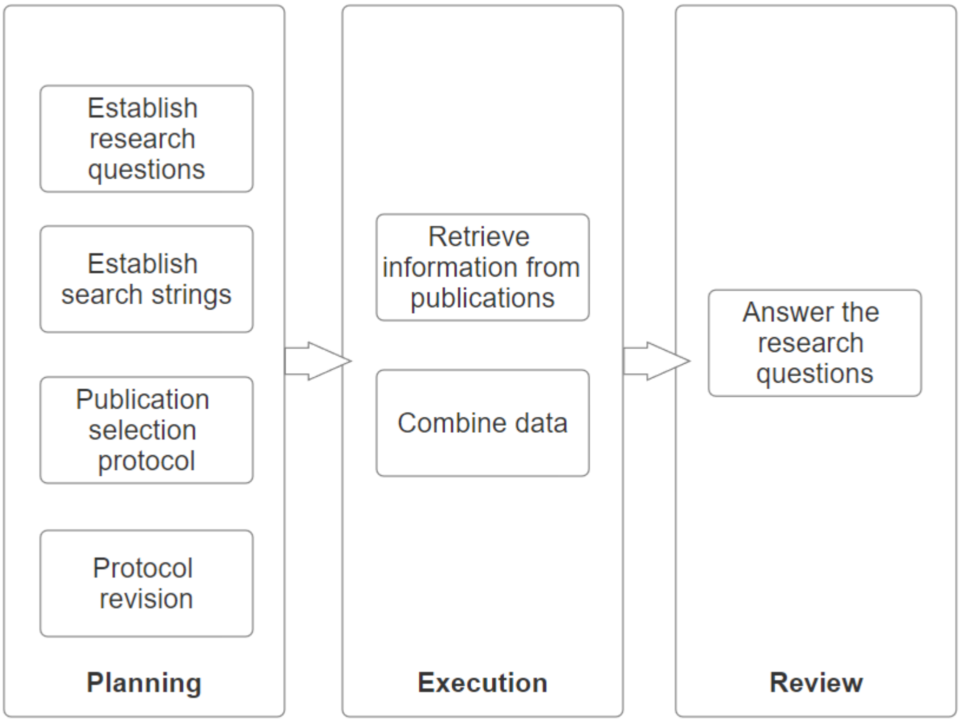

Fig. 1. Review protocol

2 RESEARCH OBJECTIVE

The primary aim of this research is to conduct a comparative analysis of ML and DL algorithms concerning crop yield prediction and crop disease prediction. To determine which algorithms to compare, we conducted a survey. The review protocol is outlined in Figure 1. After obtaining the replies of this survey, we have undertaken an in-depth comparison study. We want to identify the strengths and shortcomings of various ML and DL algorithms in agricultural applications by comparing their performance, accuracy, and other relevant characteristics. Furthermore, based on the findings of this comparative analysis, We have created a user-friendly website designed exclusively for farmers. This platform will use our study's findings to create intuitive tools and resources that promote smooth decision-making and increase agricultural output. We hope that by developing such a user-friendly interface, we will be able to bridge the gap between cutting-edge technology and practical on-the-ground application, ultimately empowering farmers with accessible and effective solutions.

The following describes what kind of benefits all stakeholders could achieve through this objective.

1. Farmers: Make informed decisions for higher productivity
   Crop prediction models offer farmers with critical information for making informed decisions throughout the farming cycle, including optimising tasks such as planting and harvesting. Farmers may use precise forecasts to adjust their methods, lowering risks and increasing profitability. These models also stimulate the use of precision agriculture, which enables resource optimisation and sustainable farming approaches. Technology-driven insights are transforming conventional farming, allowing farmers to efficiently address new agricultural challenges.
2. Cooperatives, Agencies, and Organizations in Agriculture: Leading Sustainable Farming Methods
   Crop prediction models are essential tools for agricultural organisations, cooperatives, and support groups. They use these models to provide customised recommendations to farmers, supporting sustainable farming practices. These models improve production by offering early advise on crop selection, planting, and harvesting. They also promote resource-efficient farming and reduce environmental impact through precision agriculture. Using technology, stakeholders can streamline decision-making, resulting in greater crop management and sector resilience.
3. Organizations: Crop and soil monitoring in real time for better management
   Crop and soil monitoring enabled by AI and ML are changing the way firms manage agriculture. Real-time monitoring offers early problem detection, allowing for proactive interventions to sustain yields and optimise harvesting schedules. This system improves yield prediction accuracy and overall agricultural health by responding quickly to soil conditions, weather patterns,

and disease outbreaks. Businesses profit from strategic harvesting planning and lower losses, establishing themselves as leaders in sustainable and precision agriculture.

4 Insurance Companies: Crop Insurance and Risk Assessment Offering
Crop prediction models are critical for agricultural insurance businesses because they help with risk assessment and management, allowing them to offer specialised crop insurance packages. These models analyse crop and area risks by analysing historical data and weather trends, enabling the construction of insurance plans to cover losses caused by poor weather, pests, or other disasters. Crop insurance not only protects farmers' finances, but it also promotes sector stability and sustainability. Insurance companies can use predictive algorithms to provide risk mitigation strategies, thereby increasing the resilience of the agricultural environment.

5 Policy Makers and Stake holders: Crop yield predictions empower policymakers and stakeholders in agriculture by optimizing resource allocation, enhancing food security planning, stabilizing market prices, managing risks, facilitating insurance and financial planning, supporting climate change adaptation, influencing international trade, and guiding research and innovation efforts for sustainable and resilient agricultural practices.

6 Consumers: Maintaining Stable Food Supplies and High-Quality Produce
Crop prediction models help customers by stabilising food supplies, affecting pricing, and enhancing food quality. These models enable farmers predict and respond to production fluctuations, resulting in more reliable and constant food availability. They also help to balance supply and demand, which could lead to more stable and inexpensive food prices. Furthermore, precision agriculture, powered by predictive algorithms, enables farmers to provide higher-quality products that satisfy consumer expectations for freshness and nutritional value. Crop prediction models benefit consumers by improving food stability, perhaps saving money, and providing access to higher-quality agricultural products.

The above beneficiaries are related to each other as follows: Farmers are crucial to agricultural output, as they cultivate crops and raise livestock. Cooperatives help farmers by providing resources, collective negotiating power, and market access. Agricultural development organizations provide expertise, training, and advocacy to help farmers improve their practices and achieve sustainability. Insurance firms provide risk management solutions that protect farmers from agricultural losses caused by poor weather, pests, or other unforeseen events. Governmental and regulatory policies impact the agricultural landscape, influencing farming techniques, subsidies, and environmental laws. Stakeholders in this ecosystem include farmers, cooperatives, organizations, insurance firms, and policymakers who work together to promote sustainable agricultural output, equitable market access, and food security. Ultimately, customers profit from this interconnected network by having access to a consistent and diverse food supply.

3 EXPLORING THE STATE OF THE ARTS: INSIGHTS FROM SURVEY RESULTS

As mentioned in the objectives to find out the most used ML and DL algorithm, we have addressed the following research questions.

RQ1. What data sets have been used in the literature for crop yield prediction?
RQ2. Which ML algorithms are used in the literature for crop yield prediction? [8]
RQ3. Which evaluation matrix has been used in the ML literature for crop yield prediction? [8]
RQ4. Who will be benefited from this crop yield prediction using AI?

To answer this research questions we have carried out a survey in six different databases to find out which recent algorithms are frequently used the initial results were broad we found a total of 1190 papers to narrow it down we defined an exclusion criteria which helped us bring it down to 70 papers. The search strings used at the beginning was 'precision agriculture with ML OR AI', moving further we used the search strings 'crop prediction with ML and AI'. Finally we have used 'crop yield prediction and ML OR AI'. The same string has been used to search in all mentioned six databases. During this search Corpus and Web of Science did not return any results. We have found a lot of papers published, So we have defined the exclusion criteria. In fig. 2, the center of the radial map is where we began with search phrases that returned 1190 publications. The first bar stretches from the center to the outer circle, representing the number of papers after the first criteria. The following bar begins near the outside edge of the first bar, reflecting the cumulative impact of the second criterion, Following that, bars repeat the pattern, demonstrating the cumulative effect of each exclusion criterion based on the quantity of papers. The outermost point represents the final number of papers after all exclusion rules have been applied.

3.1 Search Strings and Exclusion criteria

- Papers published after 2019: We have included papers published only after 2019. We obtained 1190 papers after applying the first exclusion criteria.
- Including only research articles: We excluded review articles, discussion papers, and short communications from our analysis. Following this criterion, the initial pool of 1190 articles was reduced to 910.

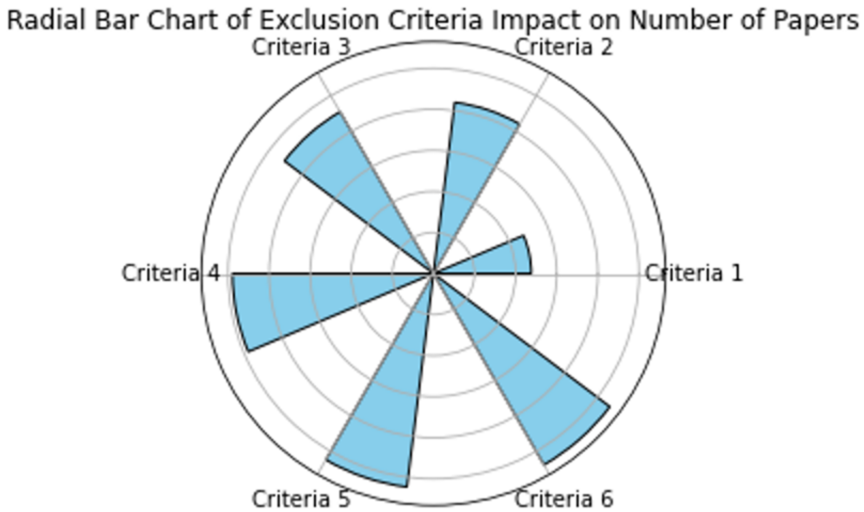

Fig. 2. Impact of exclusion criteria on number of papers

- Publication title not relevant to agricultural sector We have considered publications done in Computer and Electronics in Agriculture, Artificial Intelligence in Agriculture and Agriculture and Forest Meteorology
- Not written in English We have considered publication that are written in English.
- Duplicate publications Publications that are found across different databases are only considered once. After applying these criteria's the results were narrowed down to 173.
- Relevant to research questions: The information that was retrieved was focused on determining if the studies met the exclusion criteria requirements and answering the research questions. We have narrowed it down to 70 papers.

From the papers obtained the most used algorithms can been seen in Table 1 and most used evaluation matrix can be seen in Table 2. In these publications, RF , gradient boosting tree (GBT), LR, SVM and K-Nearest Neighbor Method (KNN) are the most often utilized models for crop yield prediction. Most of the research focused on evaluating several ML models to discover which one could produce the most accurate predictions, using a broad range of models for assessment. In addition, IoT devices are used to gather data in real time and generate predictions based on it. DL is being used to develop disease prediction models, We observed that Convolutional neural networks (CNN) and LSTM were most used algorithms in NN. Due to the limited amount of accessible data we noticed the avalibility of dataset was an issue

Algorithm	Number of Papers
RF	35
SVM	18
KNN	15
NN	14
LSTM	12
CNN, LR	11
OWN, DT	10
LASSO, NB	7
SVR	6
BOOSTING, XG	5
ANN	4
XGBOOST, RIDGE	3
Others	1

Table 1. Number of Papers for Different Algorithms

Evaluation Matrix	Number of Papers
RMSE	36
R2, Justification	35
LOYOCV validation strategy	34
MAE	12
accuracy	10
Partial dependence plots	7
MSE, RRMSE	2
Others	1

Table 2. Number of Papers for Different Evaluation Matrices

4 PREDICTIVE CROP ANALYSIS

In the development of our crop yield prediction system, we employed Python scripts utilizing various machine learning algorithms, including Decision Trees, Random Forest, Logistic Regression, and Support Vector Machines (SVM). These algorithms classify crops based on environmental factors such as temperature, humidity, pH, water availability, and season.

4.1 Dataset Description

The dataset used for predictive crop analysis comprises several key columns, each providing crucial insights into the environmental conditions and crop types:

- **Temperature:** This feature denotes the average temperature in degrees Celsius, offering insights into the climatic conditions of the region where the crops are cultivated. Temperature plays a critical role in crop growth and development, influencing processes such as germination, flowering, and fruit set.
- **Humidity:** Expressed as a percentage, the relative humidity level indicates the moisture content in the air. Humidity affects various physiological processes in plants, including transpiration, photosynthesis, and nutrient uptake. Understanding humidity levels helps assess the risk of diseases, pests, and drought stress on crops.
- **pH:** The pH value of the soil determines its acidity or alkalinity, significantly impacting nutrient availability and uptake by plants. Different crops have specific pH requirements for optimal growth, and deviations from the ideal pH range can lead to nutrient deficiencies or toxicities, affecting crop yield and quality.
- **Water Availability:** This feature represents the moisture content or water availability in the soil, indicating the amount of water accessible to plants for absorption through their roots. Adequate water availability is crucial for sustaining plant growth, development, and overall crop productivity. Insufficient water can lead to drought stress, reduced photosynthesis, and yield losses.
- **Season:** Seasonal variations play a vital role in agricultural production, influencing crop growth patterns, phenology, and environmental conditions. The season feature provides information about the time of the year when the data was recorded, offering insights into the annual climate cycle and its impact on crop cultivation practices.
- **Label (Crop Type):** The crop label column specifies the type of crop cultivated in the given conditions. The dataset includes a variety of crops such as blackgram, chickpea, cotton, jute, kidneybeans, lentil, maize, mothbeans, mungbean, muskmelon, pigeonpeas, rice, and watermelon. Each crop has distinct requirements in terms of temperature, humidity, soil pH, and water availability for optimal growth and yield.

These features collectively form the basis for predicting crop types and optimizing agricultural practices for improved yield and sustainability.

4.2 Data Handling and Exploration

The script begins by importing necessary libraries such as pandas, numpy, matplotlib, seaborn, and scikit-learn. It loads the dataset named "Crop_recommendation.csv" using pandas, which contains information about environmental factors and corresponding crop labels. Following the data loading, the script conducts initial data exploration tasks.

4.2.1 Data Loading and Checking. The dataset is loaded using the pandas library. The script then checks for missing values in the dataset, which are found to be absent.

4.2.2 Visualizations. Using seaborn, the script visualizes the distribution of crops across different seasons and water availability levels.

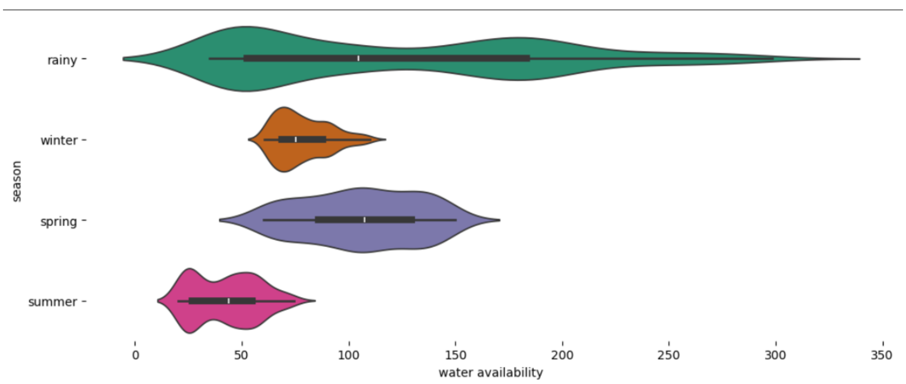

Fig. 3. Distribution of Crops Across Seasons and Water Availability Levels

Figure 3 illustrates the distribution of crops across various seasons and water availability levels using a violin plot and a bar plot, respectively.

Figure 4 shows the correlation matrix of environmental factors, providing insights into the relationships between temperature, humidity, pH, water availability, and season.

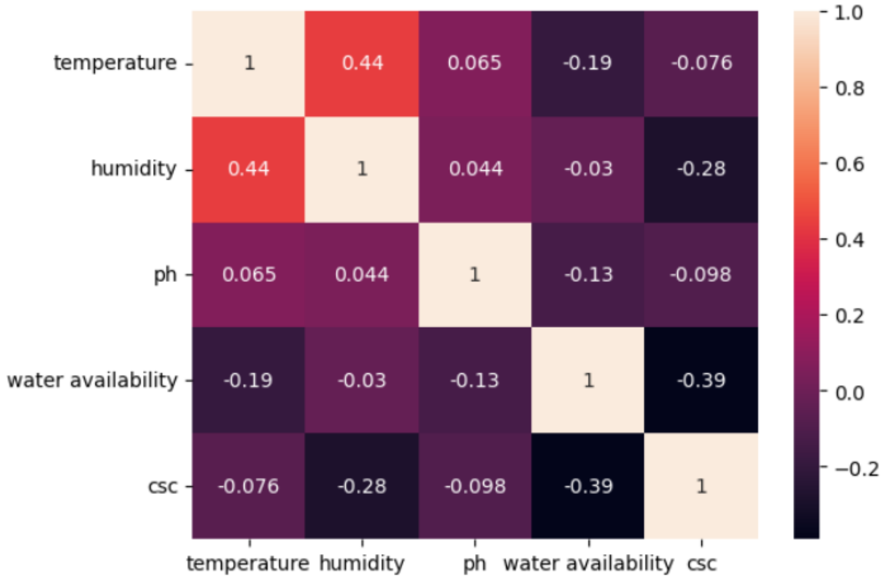

Fig. 4. Correlation Matrix of Environmental Factors

These visualizations aid in understanding the relationships between environmental factors and crop types, guiding subsequent analysis and model development.

4.3 Data Pre-Processing

We conducted several steps to preprocess the dataset before model training:

4.3.1 Mapping Categorical Season Column. To prepare the data for analysis, we mapped the categorical 'season' column to numerical values using a predefined mapping. The mapping assigned numerical values to each season category ('spring', 'summer', 'autumn', 'winter'). This transformation allows for the inclusion of seasonality as a numerical feature in our predictive models.

4.3.2 Feature Engineering and Data Transformation. After mapping the 'season' column, we dropped the original categorical column from the dataset while retaining the newly created numerical representation, labeled as 'converted_season_column'. This ensures that our dataset contains only numerical features, facilitating further analysis and model training.

4.3.3 Dataset Splitting. We divided the preprocessed dataset into training and testing sets, with 80

4.3.4 Addressing Class Imbalance. To address class imbalance in the training data, we employed the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE generates synthetic samples for the minority classes, thereby balancing the class distribution. This technique helps prevent bias towards the majority classes during model training and improves the overall performance of our predictive models.

This series of preprocessing steps ensures that our dataset is appropriately formatted and balanced, laying the foundation for accurate and reliable predictive modeling.

4.4 Results

We trained multiple machine learning models, including Decision Tree, Random Forest, Logistic Regression, Support Vector Machines (SVM) and K-Nearest Neighbor (KNN), to predict crop types based on environmental factors. For each model, we evaluated performance metrics such as accuracy, precision, recall, and F1-score using the testing data. Additionally, we generated confusion matrices and visualizations to further analyze the models' performance.

4.4.1 Model Performance Evaluation. The performance metrics for each model are summarized as follows:
Decision Tree: Accuracy: 0.93, Precision: 0.94, Recall: 0.95, F1-score: 0.94
Random Forest: Accuracy: 0.98, Precision: 0.98, Recall: 0.98, F1-score: 0.98
Logistic Regression: Accuracy: 0.85, Precision: 0.86, Recall: 0.86, F1-score: 0.85
Support Vector Machines (SVM): Accuracy: 0.96, Precision: 0.97, Recall: 0.96, F1-score: 0.96
K-Nearest Neighbor (KNN): Accuracy: 0.97, Precision: 0.97, Recall: 0.96, F1-score: 0.97

4.4.2 Model Selection. From the evaluation results, it is evident that the Random Forest (RF) model outperformed all other models, achieving the highest accuracy of 0.90. This superior performance is reflected in precision, recall, and F1-score as well. Consequently, we have selected the Random Forest model for deployment on our website.

4.4.3 Visualization. The accuracy of the Random Forest model is illustrated in **Figure 5**, and the detailed metrics evaluation is depicted in **Figure 6**.

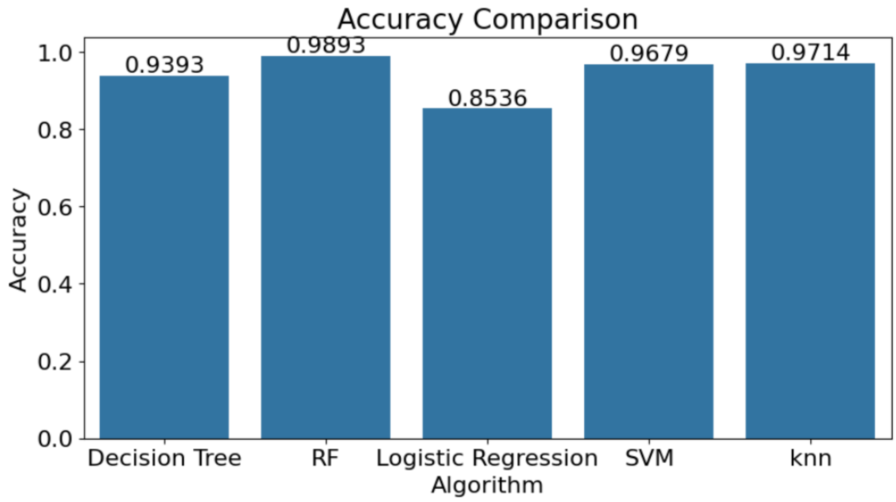

Fig. 5. Accuracy of Random Forest Model

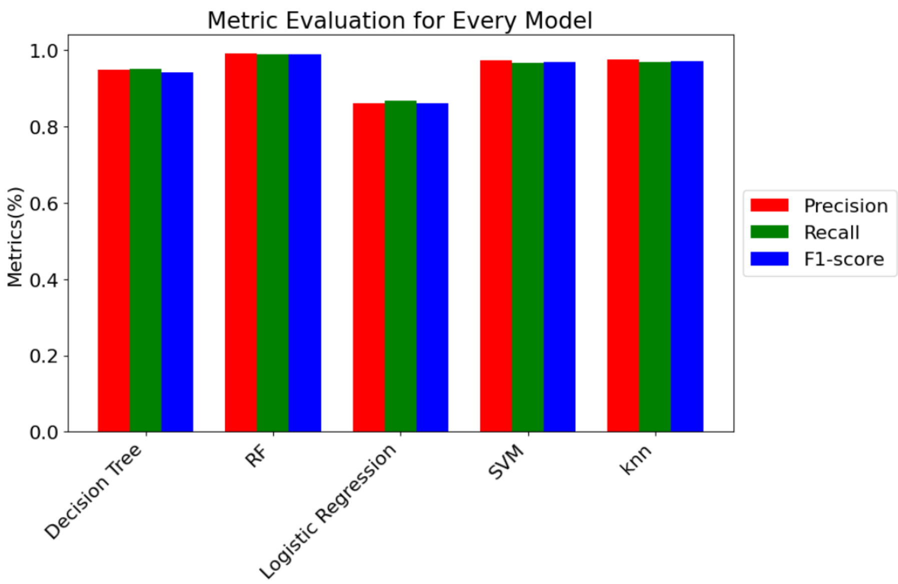

Fig. 6. Metrics Evaluation of Random Forest Model

This comprehensive analysis allows us to make informed decisions regarding model selection and deployment, ensuring optimal performance for our predictive crop analysis system.

5 DISEASE MANAGEMENT

In our disease management system, we employed a deep learning approach using Convolutional Neural Networks (CNNs). Unlike traditional machine learning algorithms, CNNs are specifically designed to handle image data effectively. Our system classifies crop diseases based on visual features extracted from images of the plants.

5.1 Dataset Description

The dataset used in our disease management system is a combination of five different datasets, encompassing a total of 17 classes representing plant species. The dataset comprises a total of 13,324 images, covering diseases and health conditions found in corn,

potato, rice, wheat, and sugarcane plant species. **Corn:** The corn class consists of four unique categories, with a total image count of 3,852. These images are derived from the widely used PlantVillage dataset [7].
Potato: The potato class is divided into three sections, containing 2,152 photos, all obtained from PlantVillage [7].
Rice: Rice classes are divided into four categories, with a total of 4,078 photos sourced from various datasets such as Dhan-Shomadhan and Kaggle [4]. Additional photographs were added for diversity.
Wheat: Wheat species are divided into three groups, each comprising 2,942 photos sourced from Kaggle's "Wheat Disease Detection" dataset [5].
Sugarcane: Sugarcane species are divided into three groups, with 300 photos collected from Kaggle's "Sugarcane Disease Dataset" [6]. This diverse dataset enables our disease management system to effectively recognize and classify various plant diseases across multiple plant species, contributing to improved agricultural practices and crop health monitoring.

5.2 Importing Libraries and Dataset Preparation

The code begins by importing essential libraries to facilitate deep learning operations. TensorFlow is imported for its powerful capabilities in building and training neural networks, while the `layers` module is used for defining the architecture of the neural network. Additionally, the `matplotlib.pyplot` module is imported for visualization purposes.

Images are loaded from a directory and converted into a TensorFlow dataset using the `tf.keras.preprocessing.image_dataset_from_directory` function. This function automatically reads images from the specified directory structure and prepares them for training.

The dataset is then partitioned into training, validation, and test sets based on predefined split ratios. Specifically, 80% of the data is allocated for training, while 20% is divided equally between validation (10%) and test (10%) sets. This ensures that the model is trained, validated, and tested on separate subsets of data, facilitating robust evaluation.

Prior to feeding images into the neural network, preprocessing techniques are applied to enhance training diversity. This includes:

- Rescaling and resizing images to a consistent size to ensure uniformity across the dataset.
- Employing data augmentation methods such as rotation and flipping to increase dataset variability. This helps prevent overfitting and improves the generalization ability of the model.

These preprocessing steps contribute to the overall effectiveness of the deep learning model in recognizing and classifying plant diseases accurately.

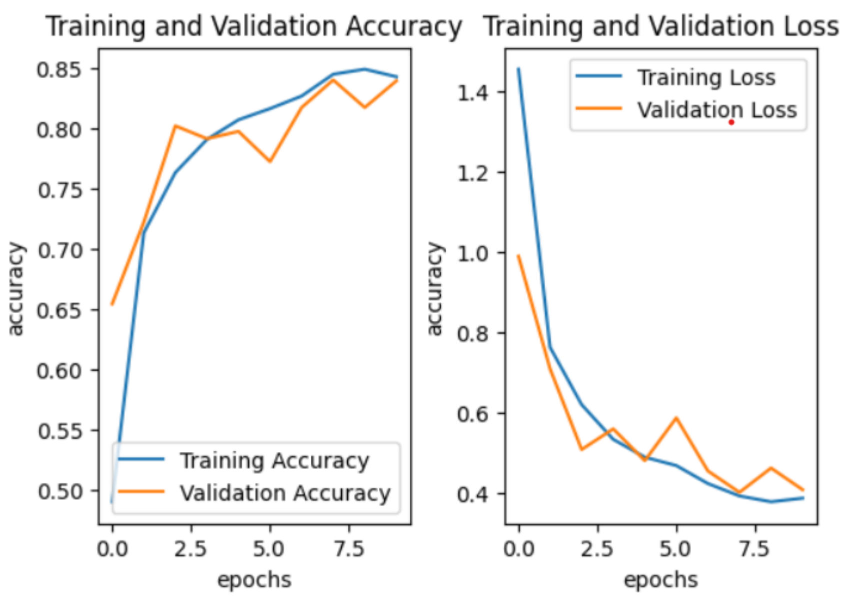

Fig. 7. Model Evaluation Metrics

5.3 Model Definition and Evaluation

The CNN model architecture is defined using the Sequential API. It consists of convolutional layers for feature extraction, max-pooling layers for downsampling, and fully connected layers for classification. The image size was set to 256, batch size to 32, with 3 channels, and trained for 10 epochs.

The model is compiled with specific configurations such as the Adam optimizer, Sparse Categorical Crossentropy loss function, and accuracy as the evaluation metric, preparing it for training. The model is then trained on the training dataset for a predetermined number of epochs using the `fit` method. Performance is monitored and evaluated using the validation dataset at the end of each epoch.

After training, the model is evaluated on the test dataset using the `evaluate` method, which provides metrics such as accuracy and loss. These metrics are depicted in **Figure** 7.

Matplotlib is utilized to visualize the training history, including accuracy and loss curves. This allows for tracking the model's performance and identifying potential overfitting or underfitting issues.

Finally, a prediction function is defined to make predictions on a subset of photos from the test dataset. After preprocessing an image, the function applies the trained model to predict the class label and confidence score of the image. These predictions are illustrated in **Figure** 8.

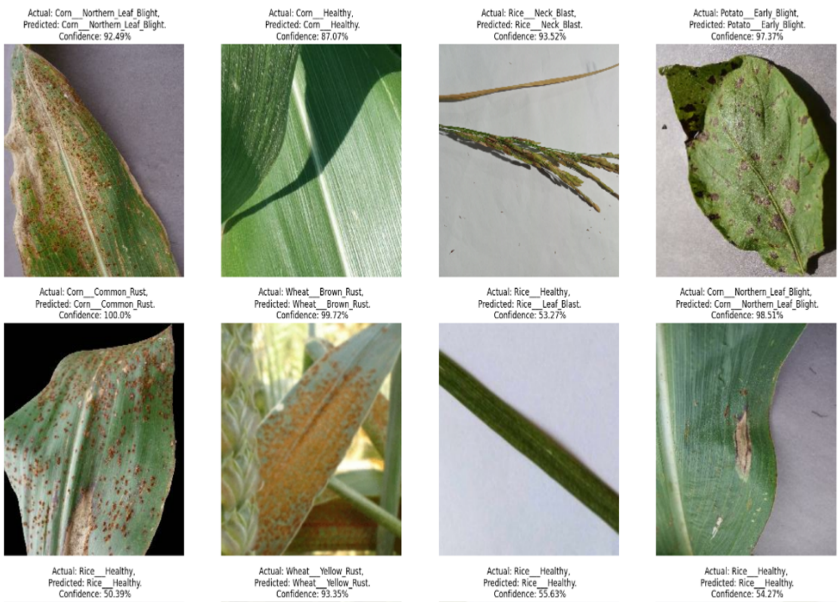

Fig. 8. Model Predictions

6 WEBSITE DEPLOYMENT

We have deployed our crop yield prediction model to a website, allowing users to input environmental features such as temperature, humidity, pH, water availability, and season to predict the suitable crop for cultivation. The website was developed using Anvil, a platform for building full-stack web apps with nothing but Python.

6.1 Functionality

The website provides a user-friendly interface where users can input the environmental conditions of their farm. Upon entering the required features, the model predicts the most suitable crop based on the provided data. This functionality assists farmers in making informed decisions about crop selection, optimizing agricultural practices, and maximizing yield.

6.2 Technology Used

The website leverages the following technologies:

- **Anvil:** Anvil enables the development of web applications entirely in Python, eliminating the need for front-end and back-end programming languages. It provides a drag-and-drop interface for building user interfaces and integrates seamlessly with Python code for server-side logic.
- **Python:** Python serves as the primary programming language for both front-end and back-end development in Anvil. It allows for rapid development and easy integration with machine learning models for predictive analytics.

6.3 Benefits

Deploying our crop yield prediction model to a website offers several benefits:

- **Accessibility:** Farmers can access the prediction model from any device with internet connectivity, making it convenient to use in the field or from home.
- **Ease of Use:** The website provides a simple and intuitive interface for entering environmental data, making it accessible to users with varying levels of technical expertise.
- **Real-time Predictions:** Users receive instant predictions based on the input data, allowing for quick decision-making in crop selection and planning.
- **Scalability:** The website can be scaled to accommodate additional features, datasets, and functionalities to meet the evolving needs of users.

6.4 Website Interface

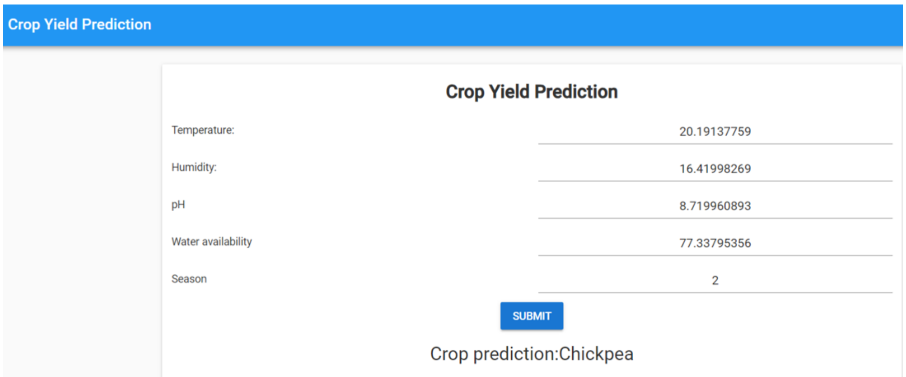

Fig. 9. Screenshot of the website interface

Overall, deploying our crop yield prediction model to a website empowers farmers with valuable insights and decision-making tools to optimize agricultural productivity and sustainability.

7 CONCLUSION

In this study, we conducted a comprehensive analysis of machine learning (ML) and deep learning (DL) algorithms for crop yield prediction and disease management in agriculture. Leveraging the power of ML and DL, we aimed to develop effective tools and resources to support farmers in making informed decisions and enhancing agricultural productivity.

We began by exploring the landscape of ML and DL algorithms commonly used in agriculture. Through a comparative analysis, we identified the most effective algorithms for crop yield prediction and disease detection. Our findings revealed that Random Forest (RF) emerged as the top-performing model for crop yield prediction, while Convolutional Neural Networks (CNNs) were most effective for disease classification.

Based on these insights, we developed predictive models for crop yield analysis and disease management. Our crop yield prediction model utilized environmental features such as temperature, humidity, pH, water availability, and season to recommend suitable crops for cultivation. Meanwhile, our disease management system classified plant diseases based on image data obtained from various plant species.

To facilitate easy access to these predictive tools, we deployed our crop yield prediction model to a user-friendly website. Built using Anvil, the website allows farmers to input environmental parameters and receive instant predictions on suitable crops. This platform offers real-time insights and supports data-driven decision-making in agriculture.

Moving forward, our future work will focus on further enhancing the capabilities of our predictive models and website interface. We aim to incorporate additional features, such as soil analysis and pest detection, to provide more comprehensive support to farmers. Additionally, we plan to integrate feedback mechanisms to continuously improve the accuracy and usability of our tools.

In conclusion, our research demonstrates the potential of ML and DL algorithms in revolutionizing agriculture by empowering farmers with valuable insights and decision-making tools. By bridging the gap between cutting-edge technology and practical on-the-ground application, we strive to contribute to sustainable and resilient agricultural practices for the benefit of farmers and stakeholders worldwide.

8 ACKNOWLEDGMENT

We would like to thank Dr. Xing Tan for providing valuable insights and assistance during the course of this research. This work was supported by Lakehead University. The authors declare that they have no conflict of interest to disclose.

REFERENCES

[1] Shaveta Dargan, Munish Kumar, Maruthi Rohit Ayyagari, and Gulshan Kumar. 2020. A survey of deep learning and its applications: a new paradigm to machine learning. *Archives of Computational Methods in Engineering* 27 (2020), 1071–1092.

[2] Ika Darnhofer, Walter Schneeberger, and Bernhard Freyer. 2005. Converting or not converting to organic farming in Austria: Farmer types and their rationale. *Agriculture and human values* 22 (2005), 39–52.

[3] Petteri Nevavuori, Nathaniel Narra, and Tarmo Lipping. 2019. Crop yield prediction with deep convolutional neural networks. *Computers and electronics in agriculture* 163 (2019), 104859.

[4] Md. Mahfuzur Rahman, Md. Mahfuzur Rahman, and Md. Mahfuzur Rahman. 2022. Dhan-Shomadhan: A Dataset of Rice Leaf Disease Classification for Bangladeshi Local Rice. https://data.mendeley.com/datasets/znsxdctwtt/1. Accessed: 2022-01-01.

[5] Behzad Safari Jalal. 2022. Behzad Safari Jalal. https://www.kaggle.com/datasets/sinadunk23/behzad-safari-jalal. Accessed: 2022-01-01.

[6] Prabhakaran Soundar. 2022. Sugarcane Disease Dataset. https://www.kaggle.com/datasets/prabhakaransoundar/sugarcane-disease-dataset. Accessed: 2022-01-01.

[7] Tushar5harma. 2022. Plant Village Dataset Updated. https://www.kaggle.com/datasets/tushar5harma/plant-village-dataset-updated. Accessed: 2022-01-01.

[8] Thomas Van Klompenburg, Ayalew Kassahun, and Cagatay Catal. 2020. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture* 177 (2020), 105709.