A Data-Driven Approach To Plant Suggestions Using Soil Health And Food Price Data

Dr.R.Nallakumar[1], S.Linkedh[2], M.Prasanth[3], K.Srikanth[4], S.Archana[5]
[1]Dept of Artificial Intelligence and Data Science
[2, 3, 4, 5]Dept of Artificial Intelligence and Data Science
[1, 2, 3, 4, 5] Karpagam Institute of Technology, Coimbatore.

Abstract- *This research introduces a data-driven framework for suggesting suitable plants by analyzing soil health and food prices. It aims to enhance agricultural decision-making by integrating these factors. Through data collection, preprocessing, and feature engineering, a predictive model is developed, correlating soil attributes and market costs with plant performance. This model aids in identifying optimal crops for specific locations. By omitting the user interface aspect, the focus remains on technical aspects. The project holds potential for optimizing crop choices, leading to improved agricultural yield and economic outcomes.*

Keywords- Plant Suggestion, Data Analysis, Feature Engineering, Soil Attributes, Yield, Decision Tree.

I. INTRODUCTION

The paper presents an innovative approach to data-driven plant recommendations using soil health and food price data, addressing pressing challenges in modern agriculture. This research provides a complete approach to improve crop selection, promote sustainability, and boost economic viability in an era of data-driven decision-making.Plant selection is the main difficulty facing agriculture because poor selections can result in lower yields and environmental damage. This work presents a novel system architecture that integrates feature engineering, model building and training, validation, and optimization with data gathering and processing. The technology provides farmers and gardeners with practical insights into which crops are best suited for their unique soil conditions and market dynamics by utilizing a wide range of data sources.

Plant suggestions can be made with accuracy and data-driven reasoning because to the integration of sophisticated machine learning algorithms including decision trees, random forests, support vector machines, and neural networks. The models undergo validation procedures to guarantee precision, and optimization methods are implemented to consistently enhance the recommendations.This research tackles problems like resource inefficiency, market volatility, and environmental sustainability in addition to revolutionizing crop choices. It promotes increased agricultural productivity, economic stability, and ecological preservation by making planting decisions more informed. The study is a major step in the direction of bringing the agriculture industry into the era of data-driven sustainability and prosperity.

A. OBJECTIVES

The study article lays forth a wide range of goals intended to revolutionize farming methods by creatively utilizing data-driven approaches. Optimizing crop selection is the main objective of this project. The goal of the research is to create a system that can provide accurate and knowledgeable recommendations to farmers, gardeners, and policymakers so they can select crops that suit their unique soil conditions, climate constraints, and financial concerns. This will be accomplished by utilizing data on food prices and soil health. The ultimate goal is to increase agricultural output and profitability while reducing the ongoing problem of poor crop choices. Concurrent with the goal of maximizing crop choice, the research aims to promote agricultural sustainability. The long-term health of the soil and the preservation of the environment depend heavily on sustainable practices. This mission strives to achieve a balance between ecological preservation and productivity by boosting crop diversification, encouraging resource-efficient consumption, and minimizing soil deterioration. The goal of the project is to provide agricultural stakeholders with knowledge and resources that will help them maximize productivity while also protecting and preserving the environment.

The development of an intuitive system is the third and most important goal. It is critical that a wide range of consumers, from experienced farmers to inexperienced gardeners, benefit from data-driven plant recommendations. The goal of the project is to create a user-friendly interface that makes plant recommendations practical and intelligible. By focusing on the needs of the user, the system's advantages are made accessible to a larger group of people, which helps to democratize sustainable practices and data-driven agriculture.

B. EXISTING SYSTEM

The existing systems for plant suggestions in agriculture are typically manual, reliant on traditional knowledge, and often lack the advantages of data-driven approaches. These systems are characterized by several key features. Traditional agriculture relies heavily on the expertise and experience of farmers and agricultural specialists. Plant selection is often based on accumulated knowledge passed down through generations, which can lead to suboptimal decisions due to changing environmental conditions and market dynamics. Many farmers resort to trial-and-error methods, where they experiment with different crops to see what works best in their specific soil conditions and climate. This approach can be time-consuming and costly. Farmers often have limited access to comprehensive soil health and food price data. Access to relevant information and technology is not uniform, making it challenging for all agricultural stakeholders to make informed decisions. Without data-driven guidance, there is a risk of resource inefficiency, including excessive water usage, fertilizer application, and other inputs. Inefficient resource usage can lead to financial burdens and environmental harm. Traditional systems may not prioritize long-term soil health and environmental sustainability, potentially contributing to soil degradation and other ecological issues. The existing systems do not account for market dynamics and food prices, leaving farmers exposed to market volatility and financial uncertainty. In summary, the existing systems for plant suggestions in agriculture primarily rely on traditional knowledge and experience, often resulting in suboptimal crop selection, resource inefficiency, and limited environmental sustainability. These systems lack the advantages of data-driven approaches that can provide more informed, efficient, and sustainable plant recommendations, as proposed in the research project.

C. PROPOSED SYSTEM

The suggested system combines a decision tree algorithm with several cutting-edge machine learning techniques to bring a novel approach to plant recommendations in agriculture. The decision tree algorithm is fundamental to the process, providing a hierarchical set of rules for classifying plant species according to factors like soil health, climate, and market dynamics. This gives users accurate and simple-to-understand recommendations. Simultaneously, a variety of algorithms, such as neural networks, random forests, and support vector machines, are utilized to improve the accuracy and resilience of the system. For example, the random forest combines several decision trees to provide overall insights, while neural networks and support vector machines deal with classification and complex relationships in the data.

II. FEASIBILITY ANALYSIS

Nothing we think has to be practical. It is a good idea to consider the viability of any problem we attempt. The examination of influence that results from the creation of a system in an organisation is called feasibility. The effect could be favourable or unfavourable. The system is deemed workable when the positives nominate the negatives. In this case, there are two approaches to conduct the feasibility study: technically and economically.

TECHNICAL FEASIBILITY :We can confidently state that it is technically possible because it won't be too difficult to obtain the resources needed for both system development and maintenance. We are making use of the resources that are already on hand because the organisation has all the resources required for both the creation and maintenance of the programme.

ECONOMIC FEASIBILITY :Development of this application is highly economically feasible. The organization needed not spend much money for the development of the system already available. The only thing is to be done is making an environment for the development with an effective supervision. If we are doing so, we can attain the maximum usability of the corresponding resources. Even after the development, the organization will not be in condition to invest more in the organization. Therefore, the system is economically feasible.

III. MODULES IN THE PROJECT

A. OVERALL ARCHITECTURE OF SYSTEM

The system architecture for plant suggestions is a multifaceted framework that encompasses a series of interconnected processes, each with a specific role in providing data-driven and precise recommendations for crop selection based on soil health and food price data. This architecture consists of four key phases: data collection and processing, feature engineering, model development and training, and validation and optimization. Here's a comprehensive explanation of each phase:

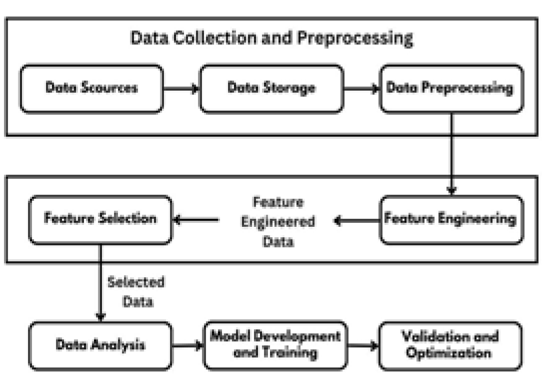
Feature engineering is a critical step in the architecture, where relevant attributes or features are selected, created, or transformed from the raw data to serve as input for the machine learning models. This process aims to extract meaningful and informative features that capture essential information for plant suitability and market dynamics. For example, features could include soil suitability indices, seasonality effects, historical market trends, and climate indices.

Effective feature engineering enhances the models' ability to make precise recommendations, as it focuses on extracting the most relevant information while reducing noise. By creating well-structured features, the system can uncover complex relationships and patterns in the data, thus improving the accuracy of its plant suggestions.

Each algorithm serves a specific purpose. Decision trees create a hierarchical set of rules that can help interpret the significance of individual factors. Random forests combine multiple decision trees to enhance accuracy and robustness. Support vector machines excel in classification and ranking, while neural networks can capture complex non-linear relationships. The diversity of algorithms ensures a holistic analysis of the data, allowing the system to provide informed and precise plant recommendations.

The final phase of the architecture is validation and optimization. Model validation is essential to ensure that the plant suggestions are accurate and reliable. Techniques like cross-validation, data splitting, and performance metrics assessment are employed to evaluate model performance and identify potential issues such as overfitting or underfitting. Validation ensures that the recommendations align with the ground truth and can be trusted by users.

Optimization is an iterative process aimed at enhancing the models and recommendations continuously. It includes hyperparameter tuning, feature selection, and the incorporation of user feedback. The system evolves and adapts to changing environmental and market conditions to remain relevant and effective over time.



System Architecture

B. DATA COLLECTION AND PREPROCESSING

The Data Collection and Preprocessing module is a fundamental component of the plant suggestion system's architecture, responsible for gathering and preparing the diverse data required for the recommendation process. This module encompasses a series of crucial tasks.

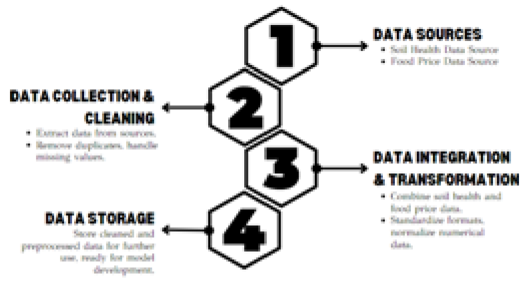
Data Collection: It involves sourcing data from various repositories, including soil health parameters (such as pH levels, nutrient concentrations, and organic matter content), climate data (such as temperature, precipitation, and historical weather patterns), crop performance records, and real-time or historical food price data. These datasets provide the essential information required to make informed plant recommendations.

Data Preprocessing: Once the data is collected, it undergoes rigorous preprocessing to ensure its quality and uniformity. This phase includes data cleaning to eliminate errors and inconsistencies, quality checks to verify accuracy, and data integration to ensure compatibility. Additionally, preprocessing techniques are applied, including data normalization to standardize scales, handling of missing values, and data transformation to make it suitable for analysis.

The Data Collection and Preprocessing module plays a pivotal role in providing high-quality data that serves as the foundation for the subsequent phases of feature engineering, model development, and validation. Ensuring data integrity and accuracy is essential for generating precise and reliable plant suggestions, as it guarantees that the recommendations are based on trustworthy information, ultimately benefiting farmers, gardeners, and agricultural stakeholders.

In the Data Collection and Preprocessing module, the quest for data quality and relevance remains paramount. Data collection continues to encompass diverse sources, including soil health parameters, climate data, historical crop records, and food price information. Ensuring that these datasets are not only extensive but also up-to-date is crucial, as it guarantees that the recommendations provided to agricultural stakeholders are based on the most current and comprehensive information available.

Furthermore, data preprocessing techniques, such as data cleaning, quality checks, and integration, are continuously refined. Advanced methods for data normalization, imputation of missing values, and transformation are explored to maintain data integrity and prepare it for analysis. The ongoing pursuit of data excellence ensures that the plant suggestion system maintains its capacity to provide accurate, reliable, and relevant recommendations, ultimately benefiting farmers, gardeners, and policymakers in their agricultural endeavors.



Data Flow Diagram for Data Collection and Data Preprocessing

C. FEATURE ENGINEERING AND FEATURE SELECTION

The Feature Engineering and Selection Module is a critical component of the plant suggestion system's architecture, responsible for preparing the raw data for model development and training. This phase plays a pivotal role in improving the quality of the input data and ensuring that the machine learning models receive relevant and meaningful features for accurate plant recommendations.

Feature Engineering: Feature engineering is the process of selecting, creating, or transforming attributes from the raw data to construct valuable features. In the context of plant suggestions, this module extracts and crafts features that encapsulate essential information related to soil health, climate conditions, and market dynamics. For example, features may encompass soil suitability indices, historical price trends, climate indices, and seasonality effects. By designing these features, the system gains the ability to represent complex relationships within the data effectively.

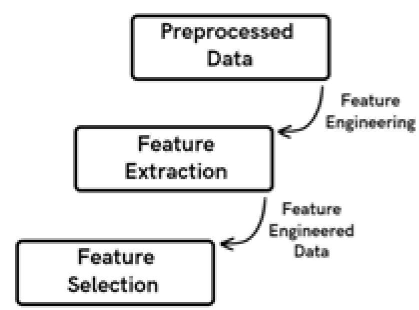
Feature Selection: Feature selection is an integral part of this module. It involves choosing the most relevant attributes while discarding irrelevant or redundant ones. This process ensures that only the most informative features are passed on to the machine learning models, reducing noise and enhancing model performance. Feature selection techniques such as recursive feature elimination or mutual information scores help prioritize the attributes that have the most impact on plant recommendations.

Data Transformation: Data transformation may also be employed in this module to convert data into suitable formats for modeling. This includes encoding categorical variables, scaling numerical attributes, and addressing missing data points.

In the Feature Engineering and Selection Module, continuous efforts are focused on enhancing the quality and informativeness of features. Advanced feature engineering techniques are explored to extract additional valuable insights from the data, ensuring that the system captures complex relationships and dependencies that influence plant recommendations. Furthermore, feature selection methods are fine-tuned to prioritize attributes that have the most significant impact on plant suitability, reducing noise and improving the model's efficiency.

The Feature Engineering and Selection Module empowers the plant suggestion system by enhancing the quality of the input data. This, in turn, enables the machine learning algorithms to make precise and informed recommendations. It plays a crucial role in extracting valuable insights from the dataset and ensuring that the selected features align with the project's goals and objectives. The efficiency of this module is a critical determinant in the overall performance of the recommendation system, ultimately benefiting farmers, gardeners, and agricultural stakeholders by providing more accurate and actionable plant suggestions.

New features are created by combining or transforming existing ones. For instance, a composite feature representing the combined impact of soil pH, nutrient levels, and climate conditions can be generated to evaluate overall plant suitability.



Data Flow Diagram for Feature Engineering and Feature Selection

D. MODEL DEVELOPMENT AND MODEL TRAINING

The Model Development and Training Module is a central and intricate part of the plant suggestion system's architecture, where the machine learning algorithms, specifically the Support Vector Machine (SVM) and Decision

Tree, are employed to create models that can make informed and precise recommendations for crop selection based on soil health and food price data.

Support Vector Machine (SVM): SVM is a powerful machine learning algorithm employed in this module. It excels in classification tasks, making it invaluable for categorizing and ranking plant suitability. SVM works by finding the optimal hyperplane that maximizes the margin between different classes of plant recommendations. This margin ensures a clear and robust separation, providing accurate results. For this project, SVM can effectively classify which crops are most suitable for specific soil and environmental conditions, enhancing the precision of the recommendations.

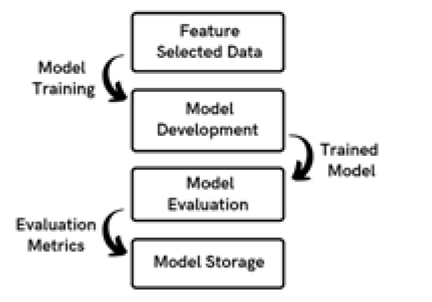
Decision Tree: Decision trees are another key component of the module. Decision trees create a hierarchical set of rules based on the input data, making them adept at interpreting the significance of individual factors in plant selection. They are known for their transparency and ease of understanding, which is essential for providing insights into why certain plant recommendations are made. Decision trees can also handle both categorical and numerical data, making them versatile in capturing the diverse attributes influencing crop selection.

The Model Development and Training Module leverages these algorithms to build robust and interpretable models. During the training phase, the algorithms learn the intricate relationships between soil health, climate conditions, and market dynamics and their influence on crop selection. The SVM focuses on classifying plant suitability, while the Decision Tree provides transparency in the decision-making process.

The Support Vector Machine (SVM) and Decision Tree models, as part of the Model Development and Training Module, are crucial in making the plant suggestion system accurate and interpretable. The SVM, with its effective classification capabilities, ensures that the recommendations align with specific soil and climate conditions. Its ability to find the optimal separation boundary results in well-defined and reliable categorizations of plant suitability.

On the other hand, Decision Trees provide transparency and comprehensibility. They construct a hierarchy of rules, making it clear why certain plant recommendations are made based on specific features. This transparency is vital in providing users with insights into the decision-making process and enabling them to trust and understand the recommendations.

By combining the strengths of these two algorithms, the module ensures that the system is both accurate and interpretable. It captures the complexity of factors influencing crop selection while making the recommendations easily understandable to a wide range of users, from experienced farmers to novice gardeners. The models trained in this module are the backbone of the system, making it a valuable tool for modernizing agriculture, optimizing crop selection, and fostering sustainability.



Data Flow Diagram for Model Development

E. *VALIDATION AND OPTIMIZATION*

The Validation and Optimization Module serves as a critical phase in the plant suggestion system's architecture, ensuring the accuracy, reliability, and continual improvement of the recommendations generated by the machine learning models. This module consists of two interrelated components: validation and optimization.

Validation: Validation is the process of assessing the performance of the machine learning models to ensure that the plant recommendations they provide are accurate and reliable. Several techniques are employed for this purpose:

Cross-Validation: Cross-validation divides the dataset into multiple subsets, allowing for iterative model training and testing. This technique assesses the model's performance under various scenarios, reducing the risk of overfitting (fitting the model too closely to the training data) and ensuring that the recommendations generalize well to unseen data.

Performance Metrics: Metrics such as accuracy, precision, recall, and F1-score are calculated to evaluate how well the models are performing. These metrics help assess the models' ability to correctly classify and rank plant suitability based on soil health and market dynamics.

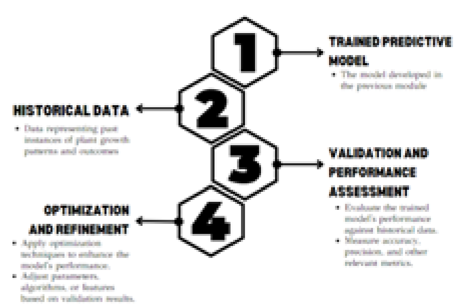
Model Comparison: The performance of different models, including Support Vector Machines (SVM), Decision Trees, and others, is compared to identify which one provides the most accurate and reliable recommendations. This enables the selection of the best-performing algorithm for the specific task.

Optimization: Optimization is the phase where the machine learning models and the plant suggestion system as a whole are refined to enhance their effectiveness. This iterative process ensures that the recommendations align with the project's goals and adapt to changing conditions:

Hyperparameter Tuning: Parameters that govern the behavior of the machine learning algorithms are fine-tuned to optimize model performance. This includes adjusting parameters for algorithms like SVM, Decision Trees, and any other models used in the system.

User Feedback: Feedback from users, such as farmers and gardeners, is integrated into the system. This feedback loop allows the system to learn and adapt based on the experiences and observations of its users.

Continuous Monitoring: The system is continuously monitored to ensure that it remains effective and up to date. It adapts to changes in environmental conditions, market dynamics, and technological advancements.



Data Flow Diagram for Validation and Optimization

IV. SOFTWARE TESTING

Software testing is a crucial phase in the development of a project focused on plant suggestion based on soil health and food prices. It ensures the reliability, functionality, and performance of the software, ultimately delivering a high-quality product to users.

Unit testing is essential to verify the correctness of individual components, such as data preprocessing, feature engineering, and model training. Integration testing ensures that these components work seamlessly together, and regression testing prevents the introduction of new errors when updates are made.

Functional testing validates the core functionality, ensuring that plant suggestions based on soil health and food prices are accurate and aligned with project requirements.

Usability testing focuses on the user interface and overall user experience, while security testing identifies vulnerabilities and protects user data. Compatibility testing ensures the software works across different platforms, browsers, and devices.

Documentation review ensures that project documents are accurate, and compliance testing addresses any legal or industry-specific standards. Exploratory testing uncovers hidden issues, and scalability, failover, and recovery testing validate system reliability. Cross-browser, cross-device, and accessibility testing guarantee a broad user reach.

V. CONCLUSION

In conclusion, the data-driven plant suggestion system represents a transformative approach to modernizing agriculture, offering a holistic solution for optimized crop selection. This project leverages innovative technologies, advanced machine learning algorithms, and comprehensive data sources to provide farmers, gardeners, and agricultural stakeholders with accurate, data-driven, and actionable plant recommendations based on soil health and market dynamics.
The integration of Support Vector Machine and Decision Trees facilitates accurate and interpretable recommendations, addressing the challenges of modern farming. Through rigorous validation and optimization processes, the system ensures that recommendations remain reliable and up-to-date, continuously adapting to changing conditions.

The transparent decision-making process, user feedback integration, and ongoing monitoring measures further enhance the system's trustworthiness and utility. By empowering agricultural stakeholders with insightful recommendations, the plant suggestion system contributes to improved crop productivity and sustainability, helping to tackle the multifaceted challenges faced by the agricultural industry. This project signifies a significant step towards informed and sustainable agriculture, ultimately benefiting not only the agricultural community but also the environment and the broader global population.

REFERENCES

[1] N. Srikanth; Bolla Tirupathi Rao; Gutla Sri Lakshmi Bhargavi; Mandadi Lakshmi Sai Likhitha ; "Deep Learning Model for Plant Disease Detection and Classification with Pesticide Suggestion", 4th International Conference on Electronics and Sustainable Communication Systems (ICESC) 2023.

[2] Shyam Chand G; Hari R ; "Plant Disease Identification and Suggestion of Remedial Measures using Machine Learning", 6th International Conference on Computing Methodologies and Communication (ICCMC) 2022

[3] R.M.N.P Karunarathna; T D D Senadeera; M B Sumesh Ranka; D V R Gunasinghe; U.U. Samantha Rajapaksha; S.M.B. Harshanath; "Plant Suggestion and Monitoring Robot", 4th International Conference on Advancements in Computing (ICAC) 2022

[4] Peng Qian; Allam Maalla; He WenHai; Li ShaoQiang ; "Agricultural Planting Big Data Q & A System Technology Research Based on Knowledge Graph", IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA) 2023

[5] Weiru Wang; Yueshuang Bao; Yulong Jin; Xinyuan Liu; Jie Hao; "Research on market operation of power supply virtual power plant based on system dynamics", IEEE Sustainable Power and Energy Conference (iSPEC) 2021

[6] S. Ganesh Sundaram; A. Ponmalar; Vijay Priya V; S. Deeba; Harish Anantha Krishnan R; J.H Vishwath; "Detection and Providing Suggestion for Removal of Weeds Using Machine Learning Techniques", International Conference on Computer, Power and Communications (ICCPC) 2022

[7] A. Thanushree; K.R. Shobha; Parimala Prabhakar; S Chandrashekhar; "Automated Soil Moisture and Nutrient Analyzer for Mulberry Plants Using IoT", IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC) 2021

[8] Hongwei Dai; Weijin Wang; "Summary of Collecting and Processing Methods for Common Cause Failure Data of PSA Components in Nuclear Power Plant", 4th International Conference on System Reliability and Safety Engineering (SRSE) 2022

[9] Ch. Rakesh D; Vishnu Vardhan; Babu Bhavani Vasantha; G. Sai Krishna; "Crop Recommendation and Prediction System", 9th International Conference on Advanced Computing and Communication Systems (ICACCS) 2023

[10] Latha Banda; Aarushi Rai; Ankit Kansal; Animesh Kumar Vashisth; "Suitable Crop Prediction based on affecting parameters using Naïve Bayes Classification Machine Learning Technique", International Conference on Disruptive Technologies (ICDT) 2023

[11] S. Thirumal; R. Latha; "Automated Rice Crop Yield Prediction using Sine Cosine Algorithm with Weighted Regularized Extreme Learning Machine", 7th International Conference on Intelligent Computing and Control Systems (ICICCS) 2023

[12] Hui Li; Liping Di; Chen Zhang; Li Lin; Liying Guo; Haoteng Zhao; "Prediction of Crop Planting Map Using One-dimensional Convolutional Neural Network and Decision Tree Algorithm", 11th International Conference on Agro-Geoinformatics (Agro-Geoinformatics) 2023

[13] Avdesh Kumar Sharma; Anand Singh Rajawat; "Crop Yield Prediction using Hybrid Deep Learning Algorithm for Smart Agriculture", Second International Conference on Artificial Intelligence and Smart Energy (ICAIS) 2022

[14] Vijay Choudhary; Archana Thakur; "Comparative Analysis of Machine Learning Techniques for Disease Prediction in Crops", IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT) 2022

[15] Dung Nguyen; Yan Zhao; Yifan Zhang; Anh Ngoc-Lan Huynh; Fred Roosta; Graeme Hammer; Scott Chapman; Andries Potgieter; "Crop Type Prediction Utilising a Long Short-Term Memory with a Self-Attention for Winter Crops in Australia", IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium

[16] Tanvi Deshmukh; Anand Rajawat; S.B. Goyal; Jugnesh Kumar; Amol Potgantwar; "Analysis of Machine Learning Technique for Crop Selection and Prediction of Crop Cultivation", International Conference on Inventive Computation Technologies (ICICT) 2023