Machine Learning-Based Soil Moisture Prediction System with Real-Time Weather Integration for Smart Agriculture

Uday Singh,
Department of Computing Technologies,
SRM Institue of Science And Technology,
Chennai, Kattankulathur, Tamil Nadu, India. Pin:-603203.
us6262@srmist.edu.in

Dr. R. Jeya,
Associate Professor,
Department of Computing Technologies,
SRM Institue of Science And Technology,
(of Affiliation)
Chennai, Kattankulathur, Tamil Nadu, India. Pin:-603203.
jeyar@srmist.edu.in

***Abstract*— Agriculture is a primary sector that supports the food of the world and the strength of the world's economies. Soil moisture content is important for crop growth and yield. The moisture content of the soil determines many characteristics of the soil which is especially relevant in agricultural fields. It helps the growers to better understand the soil and therefore grow better crops. The moisture content in the soil helps the farmers manage irrigation and water resources more efficiently. To evaluate the moisture levels in the soil, the researcher had to follow certain processes earlier, which was costly. Without access to this, the farmer had to use expensive sensors, which did not work with other systems. Due to such factors, the sensor-based measurement of.**

This project presents a Smart Agriculture Decision Support System that utilizes historical satellite data and machine learning techniques to estimate soil moisture. The raw SMAP data, available in .h5 format, is processed and converted into a structured time-series dataset. Feature engineering methods, including lag values and rolling averages, are applied to capture temporal patterns in the data. A Random Forest regression model is then trained to predict soil moisture values based on these features. To improve the practical relevance of the system, real-time weather parameters such as temperature, humidity, and rainfall are integrated using a weather API. These parameters are used to refine the model's predictions, enabling location-based adjustments and improving overall accuracy.

In addition to making predictions, the system contains a decision-making layer that provides helpful recommendations. According to the estimated soil moisture values and other environmental conditions, the system gives suggestions about which crop should be suitable, how much irrigation is required, and an approximate idea about the fertility of the soil. The straightforward, user-friendly interface allows non-technical users to access the results generated by the system. The majority of the currently available approaches focus on prediction phenomenon but not on the usability of the model. The decision machine with machine learning working in real-time together will offer practical easily expandable solutions.

Keywords*— *Soil Moisture Monitoring, Weather Data, Satellite Remote Sensing, Machine Learning, Random Forest Regression, Time- Series Analysis, Smart Agriculture, Explainable Artificial Intelligence, Visual Decision Support System, Sustainable Agriculture.

I. INTODUCTION

Agriculture continues to be one of the major sectors that contribute towards economic development and food security. With the growing population, the demand for food has been increasing. As a result, agricultural efficiency has become essential. Soil moisture is one out of the many factors which can affect the yield of crops. Moisture content of soil is very important as it helps it grow and develop properly. Additionally, it aids in planning irrigation and nutrient availability. To avert water deficiency and excessive irrigation that can affect the crop and soil, it becomes very necessary to maintain the moisture content of the soil of any area.

Conventional ways of estimating soil moisture are utilizing field sensors and manual sampling. Though effective, these approaches are generally expensive, slow, and often limited in scope. In vast agricultural areas, it is not feasible to place sensors at various locations to acquire the necessary data. Remote sensing technology has emerged as a viable solution. NASA, for example, features the soil moisture active passive (SMAP), which is a satellite mission bringing large-scale soil moisture data. The mission is global. Not directly usable. The raw data is complex and stored in formats like .h5. Thus it can't be directly used but need pre processing before using it for the intended practical.

Time series regression helps in looking at how two variables change with their relative position to each other, that is, the position of the independent and the dependent variable over a period of time. Thus, it can also be used to forecast the value of the dependent variable in the time series regression. If each timestep gets predicted, then aggregation will be too jagged. It assists you from the inside of the actual process which are working and resulting in the patterns in the data. Soil moisture prediction task which has the nonlinear nature is not suitable for a conventional approach. Soil moisture measurement and prediction can improve irrigation efficiency, relieve the drought situation and help manage waterways more effectively. Predicting soil moisture is essential for managing water resources. Our method converts data from environmental and policy settings into a clear forecast of future water quality.

Machine learning models can identify complex patterns and relationships in data that traditional models may overlook, resulting in more accurate predictions. The moisture present in the soil is one of the basic requirements of hydrological cycle and it affects plant growth, ecosystem health, and groundwater storage. Soil moisture in real life scenarios is a function not only on historical values but also on dynamic environmental properties like temperature, rainfall, humidity and so on. By integrating real-time weather data, it would be possible to increase the accuracy of the prediction system. The farmer also needs some recommendations based on the prediction and not just numbers to take a decision regarding irrigation, crop, soil, etc.

Change has only been constant for ages it seems like. Since the Greeks in the 5th and 6th century B.C defined change as the form becoming something else many researchers were developing varied concepts of change. Change has not been defined in an 'absolute' manner majorly because of the changing environment. When old factors stop being relevant, progress reverses and science reiterates. Also, something appropriate for a person may not be appropriate for the same person years later. Consequently, change invokes a circular process of effect. There exists a wide variety of change. Organizational change is one of the most common observed. We should research about it.

Change management can be viewed as the alteration in the structure, functions, technologies or culture of the organization to bring about changes in the system in order to keep up with changing times. For instance, after weighing the costs and benefits, a company may decide to outsource manufacturing operations to save costs, create competitive advantage, and earn better profits.

This work makes a key contribution by integrating a variety of components into a single system. The system processes satellite data and uses machine learning integrates real-time weather data and is also a decision support system. Other literature primarily focuses on either calculating NDVI derived from satellite datasets or anticipating NDVI based on historical NDVI datasets. Nevertheless, none of them are usable agricultural systems.

II. LITERATURE REIVEW

Advancing soil moisture estimation requires complex integration of field-based, remote sensing, and AI methodologies, supported by systematic approaches, thus the prediction of soil moisture dynamics at high spatial and temporal resolution may be improved with machine learning techniques that integrate all relevant data. The NASA Soil Moisture Active Passive (SMAP) mission, launched in 2015 [1], has provided high resolution global mapping of soil moisture and freeze/thaw state, and the overall objective of SMAP is to monitor global soil moisture mapping with unprecedented resolution, sensitivity, area coverage, and revisit times [1]. The SMAP mission will make global measurements of soil moisture present at the Earth's land surface and will distinguish frozen from thawed land surfaces, utilizing the SMAP L-band TB and radar data for soil moisture production [2]. Data preprocessing involves the steps to change or encode data so that it can be effortlessly analyzed by the machine, converting unprocessed data into analytical-ready data while resolving inconsistencies and errors and handling missing values [6].

Random Forest is an ensemble-learning algorithm that constructs multiple decision trees and aggregates their predictions to improve accuracy and reduce errors [4], it is used for regression and classification tasks and has gained popularity for its high accuracy and ability to explore non-linear relationships between predictor and response variables [4]. Random forests are commonly used for image classification, natural language processing, and time series forecasting tasks in environmental monitoring, they can be implemented using tools such as Scikit-learn [6]. The study further shows that ensemble ML models are better at capturing nonlinear interactions among soil, climate, and crop factors compared to traditional methods, and process-based and data-driven modeling have complementary strengths, they can be combined to yield more accurate predictions [7].

Deep learning algorithms have been shown to be effective in environmental modeling, and data-driven approaches have been used to build models with parameters learned from observations' data [7]. Reichstein et al. proposed a fusion of machine learning and mechanistic modelling approaches [7], and deep learning has achieved notable success in modelling ordered sequences and data with spatial and temporal context. Techniques such as SHAP assign each feature an importance value for a specific prediction, and various methods have been proposed to help users interpret the predictions of complex models, increasing transparency and accountability [8]. However, recent advances in AI have been driven by data abundance and computational scale, which rarely hold in low-resource environments, and deep learning is not usually compatible with resource-constrained devices [7].

The role of smallholder farmers and their families in increasing agricultural productivity growth sustainably will be crucial. There is an urgent need to use water more efficiently in agriculture [9], but, on the other hand, irrigation is one of the main ways to increase food production and efficient water use is a cornerstone of sustainability [9]. Accurate measurement and prediction of soil moisture can optimize irrigation practices, mitigate drought impacts, and improve water resource management. Soil moisture sensors ensure crops receive the amount of water needed, it helps conserve water while boosting crop yields.

This paper presents notable challenges that are generally involved in an Earth Observation small satellite mission and further challenges that are posed by several limitations remain in existing systems, which are generally involved in satellite data analysis and machine learning. Most approaches to data-driven weather prediction rely solely on historical data and do not incorporate real-time environmental conditions such as temperature, humidity, and wind velocity. Soil moisture is a crucial parameter influencing agricultural productivity and is highly dynamic, making accurate measurement and prediction essential for improving prediction accuracy. Many agricultural systems lack user-friendly interfaces and are designed primarily for research purposes, limiting their applicability in real-world agricultural scenarios, which can be addressed by enhancing the efficiency and user experience of agricultural user interfaces.

The proposed system refers to automated weather monitoring and nowcasting, which provides reliable estimations up to 2 h ahead of time, and integrates machine learning with real-time weather data, whereby data from several locations are used to predict the weather. Satellite data dominate soil moisture remote sensing analysis [1], and learning-based and multi-sensor fusion methods improve estimation accuracy [3], which is enhanced by a novel stacking ensemble learning framework to estimate soil moisture. This study proposes a system to optimize farmers' capability to make informed decisions based on real-time soil data and contribute to food security, offering results in the form of yield estimation, crop recommendation, fertilizer suggestions, and chatbot-based support. Agricultural living labs bridge theory and practice through transformative approaches, and this approach aims to ensure sustainable agricultural practices, building resilience to climate change while balancing productivity and environmental considerations.

III. PROPOSED SYSTEM

This framework provides a solid integration of explainable AI, real-time weather intelligence, and adaptive user interactivity, leveraging the combined power of IoT and machine learning, to support ML-based decision making and real-time environmental sensing, which enhances the precision, timeliness, and effectiveness of farm-level and policy-level decisions. The goal of this research is to make sophisticated models for precise crop production forecasting and thorough evaluation of soil health, by leveraging sensor data, satellite imagery, and environmental variables to predict soil moisture with high predictive performance and enhanced explainability.

SMAP measures the amount of water in surface soils using a radiometer mounted to a satellite orbiting the Earth, and the NASA Soil Moisture Active Passive satellite mission was launched to provide global mapping of high-resolution soil moisture, which is utilized at the core of the system. The raw data is in HDF5 format, which organizes data hierarchically, and an HDF5 dataset is an object composed of a collection of data elements, or raw data, and metadata that stores a description of the data elements, and it is processed and transformed into a structured dataset. Feature engineering techniques, such as creating lag features and rolling aggregates, are applied to capture temporal patterns, and lagged features are created for average moisture readings using different time intervals to capture temporal dependencies in soil moisture. Random Forest models are used to facilitate the prediction of soil moisture, and the Random Forest ensemble learning algorithm is tested to demonstrate the capabilities and advantages of machine learning for soil moisture estimation, which predicts soil moisture based on historical trends.

Integrating a weather API into your application is a crucial step in leveraging real-time weather data to enhance functionality. This project focuses on integrating real-time weather data using the Open Weather API. The system retrieves key weather metrics, including current temperature, humidity, and short-term forecasts, and dynamically displays them on the website. The application fetched parameters such as temperature, humidity, pressure, wind speed, and weather description almost instantly. Soil moisture is a crucial parameter influencing agricultural productivity, Irrigation scheduling and climate modelling. Accurate prediction of soil moisture content is crucial for agricultural systems as it affects hydrological cycles, crop growth, and resource management. In this paper, we provide a survey of location-aware recommendation systems in mobile computing scenarios. Geolocation services enable mobile apps to identify and use a device's location for enhanced user experiences, personalization.

The system incorporates a decision support layer that enhances system usability, which can be achieved by integrating it with Business Intelligence capabilities to analyze data, forecast trends, and optimize, thereby empowering organizations and enhancing the system's usability. Soil moisture conditions are represented and categorized into intuitive levels such as dry, moderate, and moist, which can be determined with a computer simulation model and soil moisture monitoring, a critical component of modern agriculture. The system provides crop recommendations by mapping moisture levels to suitable crops, using advanced machine learning and a cloud-based system, which gives real-time, data-driven recommendations on the best crops to cultivate in a particular farm environment. Irrigation advice is generated to guide users on whether watering is required, which helps optimize water usage by employing irrigation technologies and management strategies, such as adjusting schedules based on weather and detecting leaks.

The system estimates soil fertility using a rule-based approach that considers soil type and moisture conditions, and the accurate assessment of soil fertility is critical for guiding nutrient management and promoting sustainable agriculture in semi-arid agroecosystems, which involves understanding how soil-microbiome-plant processes contribute to soil health. Proper fertilization increases soil organic matter content, enhances soil structure, and improves the soil's capacity to retain water, and implementing Soil Health Management Systems can lead to increased organic matter, more diverse soil organisms, reduced soil compaction and improved nutrient, which allows users to understand the quality of their soil and take appropriate actions. The study aims to present a plant recommendation system using environmental data, such as temperature, humidity, relative humidity, rainfall, and soil data, and incorporating soil information with machine learning for crop recommendation to improve agricultural output, which considers soil characteristics, soil types, and crop yield data collection.

Streamlit is an open-source Python framework for data scientists and AI/ML engineers to deliver interactive data apps, and it is used to create an interactive app for your data or model. The platform is engineered to support advanced analytics and AI-driven insights, ensuring that users have the necessary tools to interpret results without requiring specialized knowledge. Soil colour serves as a critical indicator of its properties and conditions, and it has colour coding for different indicators alone or aggregated according to soil, which helps determine soil health.

The ML model showed that incorporating soil moisture data, along with carefully selected weather variables, can achieve accuracy, and this study aims to develop an IoT-based crop prediction system that utilizes real-time data on soil nutrients, pH, and weather conditions, combining various AI agents for computer-aided soil classification, crop suggestion, and fertilizer prediction. The ML model gives farmers personalized, context-aware recommendations, and by synthesizing data from diverse sources such as soil sensors, weather stations, and crop monitoring devices, farmers gain a holistic view, which helps them make data-driven decisions.

IV. ARCHITECTURE

The proposed system effectively integrates deep learning, ensemble learning, and explainability to improve farming methods and production and support sustainable agricultural applications through supervised machine learning models to derive crop yield predictions, pest detections, and resource management. This research work proposes an AI-powered system utilizing intelligent agents for carrying out functionalities such as automation, decision support, data analysis, and others, enabling informed decision-making, resource optimization, and task automation.

The system begins with the input layer, where data is collected from three primary sources, including an input layer, hidden layers, and an output layer, where inputs are inserted into the input layer. The first source is satellite-based soil moisture data obtained from the NASA SMAP dataset, which measures and maps Earth's soil moisture and freeze/thaw state to better understand, and SMAP measures surface soil conditions everywhere on Earth every two to three days. The second source is real-time environmental data acquired through a weather API, including parameters such as temperature, humidity, and rainfall, which provides real-time and historical accurate data, and the Weather API lets you request real-time, hyperlocal weather data for locations around the world. The third source is user-provided input, which includes the geographical location and soil type, and a global soil database, framed within a Geographic Information System, containing up-to-date information on world soil resources. These inputs collectively ensure that the system considers both historical patterns and current environmental conditions while making predictions, by utilizing historical and current data to project future states of environmental variables, and environmental predictions are the scientifically informed forecasts about the future state of the environment.

It turns raw inputs into governed, reliable data. Data transformation is the process of taking raw data that has been extracted from data sources and turning it into usable datasets. The goal of a data pipeline is to ensure that raw data is transformed into a usable format and delivered to a destination where it can be used. The data is available in H5 format. One could re-code missing/bad values as some numeric value, but this would be non-standard. The first byte is chosen as a non-ASCII value to reduce the probability that a text file may be misrecognized as an HDF5 file; also, it catches errors when trying to access the data with their path. Data aggregation is the process of combining datasets from diverse sources into a single format and summarizing it to support analysis and decision-making. Data aggregation refers to the process of collecting and compiling data from multiple sources and presenting it in a summarized form. Data integrity concentrates on maintaining consistent data across systems while preventing unauthorized changes or corruption of information during storage or transmission. Data consistency refers to the degree to which distributed copies of data are synced across a system, ensuring that all the end users and applications receive the same data values.

Feature engineering is the process of transforming raw data into relevant information for use by machine learning models, which involves creating new features, combining existing ones, and extracting relevant information to boost model performance. This study explores the spatiotemporal characteristics of soil moisture, and soil moisture changes can be predicted using various modeling methods. Deep learning enhances time series analysis by automatically learning patterns and dependencies directly from data, capturing both short-term and long-term dependencies. Time series forecasting refers to the practice of examining data that changes over time, then using a statistical model to predict future patterns and trends, and incorporating time-related features is essential to improve predictive accuracy.

In hierarchical Feature Learning, we extract multiple layers of non-linear features and pass them to a classifier that combines all the features, and machine learning follows a structured process, starting with data collection and preprocessing, then model selection and training, followed by, and Deep learning (DL), a branch of machine learning (ML) and artificial intelligence (AI) is nowadays considered as a core technology of today's Fourth. Random forest is a commonly-used machine learning algorithm that combines the output of multiple decision trees to reach a single result, and Random Forest Regression is used when you want to predict a continuous numeric outcome or target variable, as opposed to a categorical one, and Random forest regression is a supervised machine learning algorithm. The model leverages seasonal vegetation change patterns for retrieval instead of assuming minimal vegetation changes, and For soil moisture forecasting, our results demonstrate that the temporal modeling capability of long short-term memory (LSTM) is well suited, and Researchers developed an ML-based prediction tool using climate projections, soil moisture measurements, and decades of settlement records. Ensemble learning is essentially a form of meta-learning, where the goal is to leverage the collective knowledge of several models to achieve better results, and Ensemble learning combines the predictions from multiple neural network models to reduce the variance of predictions and reduce generalization, and Resilience to noisy data: Ensembles are less affected by outliers or random noise in the dataset since they rely on multiple models..

CV is used in data analysis to validate the implemented models where the main objective is prediction and to estimate the prediction performance, the techniques to evaluate the performance of a model can be divided into two parts: cross-validation and holdout. RMSE and MAE are widely used metrics for evaluating models, Root Mean Square Error in Machine Learning is a popular metric used to measure the accuracy of a model. Evaluation results can identify needed improvements to your program, guide program planning, demonstrate the effectiveness of your program, and justify funding, comprehensive analysis ensures a holistic assessment.

The system incorporates a hybrid convolutional neural network-temporal attention mechanism approach for real-time prediction of soil moisture, allowing precise, timely monitoring of soil water content, and helping farmers optimise irrigation strategies based on soil moisture sensors. Temperature, humidity, and precipitation are primary climate factors influencing soil moisture, and changes in these factors can greatly influence soil hydro physical properties, affecting soil moisture dynamics. Soil type influences water retention capacity, and accounting for this factor improves the predictive capability by comparing the prediction performance of published pedotransfer functions and integrating advanced machine learning models.

Prescriptive analytics focuses on providing actionable recommendations that guide decision-making, and it involves translating the model's output into actionable insights by integrating it with clinical knowledge, patient history, and current information, which is intended to transform raw algorithmic data into actionable insights that directly support the decision-making process. Soil moisture values are categorized into distinct classes, such as dry, wet, and moist, and a value of 0 means soil moisture is at the wilting point, while a value of 1 means soil moisture is at saturation. The system provides recommendations regarding suitable crops, irrigation requirements, and soil fertility status, and the quality of irrigation water being used is the most critical factor in predicting, managing, and reducing salt-affected soils. These innovations highlight how locally adapted; low-tech solutions can bridge the gap between advanced research and practical, on-ground applications, and Explainable AI enables informed decisions by providing transparency without compromising performance.

The output layer consists of class labels and their associated probabilities, and it outputs the final result, which can have a user-friendly interface. The system displays predicted soil moisture values and environmental conditions in a clear and easily understandable format using soil moisture sensors and AI. Visual elements such as graphs, indicators, and color-coded classifications are used to enhance user experience, and the interface is designed to be simple and accessible.

Data integration architecture is the framework that defines how data from various sources is collected, processed, and delivered to target systems, ensuring seamless integration of diverse data sources into a unified format, and supporting the integration of machine learning and AI with data platforms. By combining satellite imagery and remote sensing technologies, this research aims to enhance the accuracy of crop yield predictions, and by analysing historical data alongside real-time observations, these systems can provide farmers with actionable insights to improve yield predictions. This layered model is particularly well-suited for application scenarios where task complexity increases progressively across layers, and the integration of modern tools, digital innovations, and advanced machinery has improved efficiency, reduced waste, and boosted productivity, making it simpler for farmers to use.

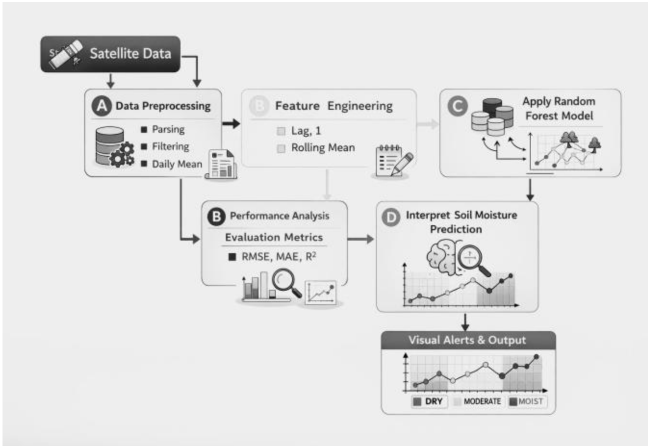

Fig. 1: Detailed architecture of the proposed system.

This section an Architecture is proposed as shown Figure 1 for smart Agriculture using internet of things (IoT) which provides numerous benefits, the technical architecture of the intelligent agricultural system is the main technical support for realizing the functions of the whole system, and it can visually represent smart farming systems, making it easier for architects and farmers to comprehend and work with, a reference architecture for data management in smart farming is also proposed A Data Flow Diagram (DFD) is a graphical tool used to represent how data moves through a system, it shows data inputs, outputs, data stores, and data flow diagrams (DFDs) are visual tools that map how data moves through systems, from sources to storage to outputs DSS is defined as a computer-based system that integrates and processes multi-source data, aiming at supporting farmers in their decision-making, the system enables users to upload soil images and input their location, automatically linking real-time weather data, and fertilizer prediction shows the Crop Recommendation interface, where users can input soil and environmental data These technologies can be utilized for tracking crop growth, identifying plant diseases, optimizing nutrient input, and many of the major components of digital agriculture can be decomposed into three features: (1) data and data collection systems, (2) decision support (DS) tools, with the analysis of different sets of information-such as soil composition, temperature, humidity, crop health, and weather forecasts-ML algorithms are able to process the data.

Feature engineering is the process of transforming raw data into relevant features for use by machine learning models, which involves selecting and creating input features such as lag values and averages. The Random Forest algorithm is used to predict soil moisture, as it can capture complex interactions between predictor variables and achieve high accuracy.

Predicted values can be refined using real-time weather conditions and soil characteristics, which are crucial to improving the accuracy of soil parameter predictions.

A decision support system is a computer-based information system that supports decision-making activities, providing evidence-based insights derived from data, and it transforms raw data into predictions and insights. Soil moisture status can be classified into different levels, and by selecting the right crop for given conditions, one can optimise the requirement of irrigation water and added fertiliser and increase yields. A user-friendly interface is one that feels intuitive, requires minimal learning, and helps people achieve their goals without frustration, making tasks simpler and navigation more intuitive.

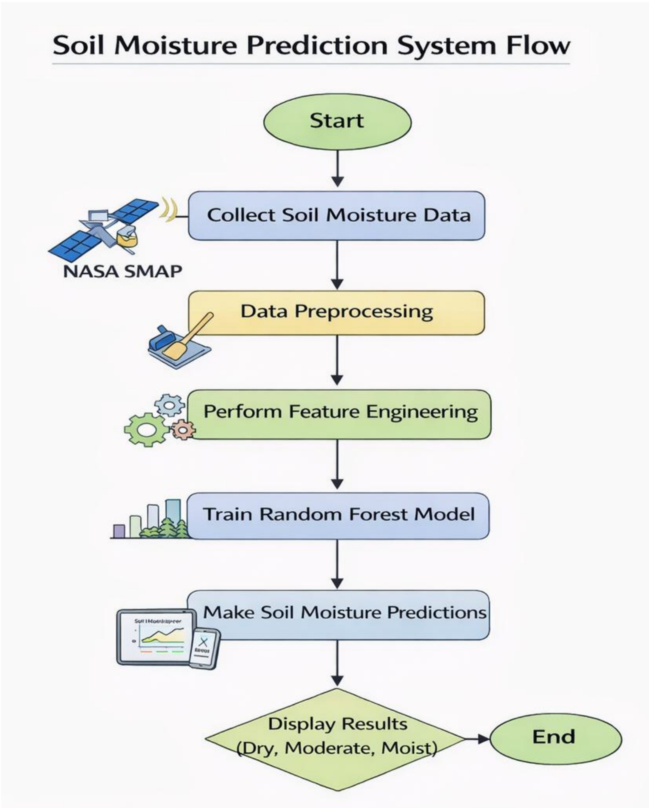

Fig 2: Flow Chart.

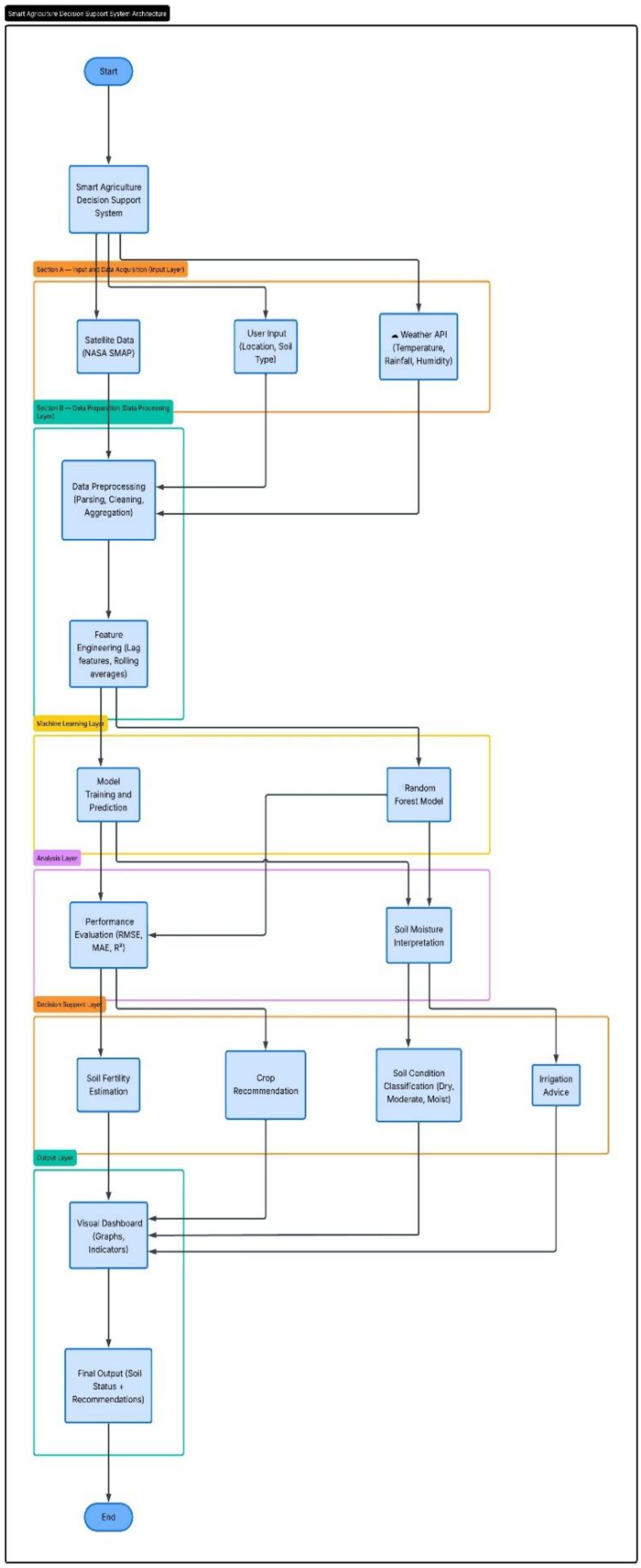

Fig 3: Detailed Flow Architecture diagram.

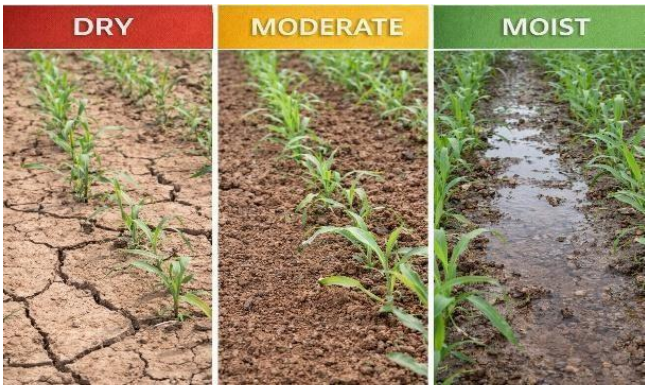

Fig 4: A visual representation of different soil conditions: dry, moderate, moist soil states.

Soil moisture plays an important role in ecology, hydrology, agriculture and climate change, and this study proposes a soil moisture prediction model, which allowed predicting SMC from the knowledge of PSD, parameterized by means of the closed-form van Genuchten model, and each type of soil moisture data set has its strengths and weaknesses. The Unified Soil Classification System classifies all soil into three categories: coarse grained, fine grained, and highly organic, and soil classification is the systematic categorization of soils based on their properties, such as texture, structure, and mineral composition. Soil classification is the systematic categorization of soils based on their properties, such as texture, structure, and mineral composition, and the USDA Soil Taxonomy System consists of six levels.

Soil drying and re-wetting alters root to shoot phytohormonal signalling and soil nutrient availability, and soils that are too dry can mean decreased crop yield and quality, so irrigation is required to support crop growth. Monitoring soil moisture levels in the field is a fundamental skill that helps producers optimize crop yields, conserve water and energy, and avoid soil erosion, and many commercial capacitive sensors can achieve 3%-5% accuracy after calibration for specific soil types and environments. The soil water content influences plant growth, soil temperature, transport of chemicals, and soil moisture and its availability to support plant growth is a primary factor in farm productivity, and adequate soil moisture ensures that plants receive the necessary water and nutrients.

Efficient water management is vital for sustainable agriculture, and an intelligent agricultural telemetry system provides users with data-driven real-time insights and promotes sustainable farming practices, and this system monitors real-time weather conditions, soil moisture, and plant health.

V. METHODOLOGY

The methodology adopts a multi-level data fusion strategy to integrate satellite, sensor, meteorological, and market data into meaningful agricultural insights through a series of well-defined steps utilizing modern computer vision technology to identify and map cultivable land for agricultural needs. It involves data collection, preprocessing, feature engineering, model training, and validation steps, including automated data preprocessing and feature engineering to transform raw data into a suitable format for training machine learning models. Each stage is designed to ensure accuracy, reliability, and practical usability, focusing on systematically applying engineering tools and analysis to ensure a product's reliability and validate the system to ensure accuracy, reliability, consistent intended performance.

The process begins with data collection, where soil moisture data is obtained from the NASA SMAP satellite dataset, which measures and maps Earth's soil moisture and freeze/thaw state to better understand. The data is available in HDF5 format, which is a general purpose library and file format for storing scientific data, and contains multiple parameters related to environmental conditions. For this project, the soil moisture parameter is extracted and used to build a time-series dataset, which can be used to generate Machine Learning-based high-resolution soil moisture data. In addition to historical data, real-time environmental data such as temperature, humidity, and rainfall is collected using a weather API, which delivers continuously updated weather information directly into applications.

Data preprocessing is the process of cleaning and preparing raw data so it can be used effectively for analysis and model. Data preprocessing enhances data quality, resolves discrepancies, and ensures that the data is correct, consistent, and reliable. This step was pivotal in identifying and rectifying any anomalies, such as missing or invalid values, which could compromise the integrity of the subsequent. Data in real-world scenarios often contains errors, inconsistencies, missing values, and outliers that can impact statistical analysis and. Data can then be aggregated in different ways in order to extract patterns from it. The dataset which is stored in a separated value (CSV) file format is loaded into a data frame with the help of the panda's library.

Feature engineering plays a critical role in improving the accuracy, efficiency, and interpretability of machine learning predictions, it involves selecting, creating, or modifying features to improve the model's ability to make accurate predictions or classifications. The characterization of temporal and spatial variability of soil moisture is highly relevant for understanding the many hydrological processes, and quantifying soil moisture is crucial for advancing modeling of energy balance at the land surface. A lag feature is simply a past value of a variable that has been shifted forward in time, and a rolling average takes a number of recent values and calculates their mean for every point in time. Soil moisture changes can be predicted using various modeling methods, and the logarithmic regression model enhances prediction accuracy by capturing nonlinear relationships in soil moisture variations.

Split the data into training and testing sets, then train a Random Forest model that constructs many decision trees during its training period, where a random forest algorithm is utilized to learn patterns from the dataset, it uses labeled data to learn how to classify unlabeled data. To build and evaluate a machine learning model, the dataset must be divided into two parts i.e. one for training the model and another for testing, it divides the data into training and testing sets to test the correctness of a model's predictions. Feature engineering allows the same model type to take on more complex formulas and capture essential patterns, which increases the model's predictive performance by capturing both linear and non-linear relationships in the data. Random Forest is chosen due to its balance between predictive accuracy, robustness, and computational efficiency, it works well with datasets that contain outliers or intrinsic noise because of its robustness and is especially well equipped to deal with noisy data.

Once the model is trained, predictions are generated for soil moisture values, which are based on the training phase that selects one model from all the available ones, and the model was trained using datasets that showed the percentage of water content correlated with the corresponding capacitance values for soil moisture content. These predictions represent the estimated moisture levels based on historical trends, and specifically, the model leverages historical observational data, including temperature, precipitation, evaporation, and soil water content, as well as current observational and mesoanalysis trends. To evaluate the performance of the model, standard metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2) are calculated, which are used to measure how well a machine learning model performs. These metrics provide insights into how well the model is performing and help in comparing different models or algorithms, allowing us to gauge the effectiveness of forecasting models.

The system integrates real-time weather data into the prediction process, like a real-time weather forecasting mobile application that integrates machine learning and IoT technology, and provides real-time conditions for anywhere in the world. The predicted soil moisture values are adjusted based on environmental parameters such as temperature, humidity, and rainfall, considering factors like air temperature, relative humidity, and wind speed to forecast SM. Higher temperatures may reduce soil moisture, while rainfall increases it, and soil type is also considered, as different soils have varying water retention capacities, like higher water retention capacity at depth compared with other layers.

The NCSMMN is a verifiable soil moisture dataset that can be used by operational decision-makers, providing value in their decision-making process, where soil moisture sensors provide a solution by allowing precise, timely monitoring of soil water content, helping farmers optimise irrigation strategies based on current soil moisture levels. The relative amount of water present in the soil sample should be described by an adjective such as dry, moist, or wet, and Drought & Dryness Categories include Abnormally Dry, Moderate Drought, Severe Drought, Extreme Drought, and Exceptional Drought. The system provides a crop suggestion system which suggests the farmer the most profitable crop they can grow with the available resources, and irrigation is vital to produce acceptable quality and yield of crops on arid climate croplands, where soil test levels for all nutrients have been put into categories labeled very low, low, medium, high, and very high. A friendly interface is defined as a user-friendly design that facilitates interaction and communication, characterized by visual elements, simple controls, and end-user-friendly explanation forms from technical literature, enabling users to easily interpret the results and take appropriate actions.

This targeted approach improves input use efficiency, and it uses advanced machine learning to give personalized crop recommendations, thus making it possible to automate and target precision at different stages, and by relying on precise and timely data, farmers can optimize their resource use, and it can assist in decision making by allowing the farmers to obtain correct information to make knowledgeable decisions.

VI. RESULT AND DISCUSSION

The performance of the proposed system was evaluated using standard regression metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2 score), which are used to evaluate the model. The obtained results indicate that the model achieves low prediction error and a high degree of correlation between predicted and actual soil moisture values, which demonstrates the model's ability to predict soil moisture dynamics. The Random Forest model is effective in capturing temporal patterns in the dataset because it makes no assumption about the residuals of the model and can deal with temporal autocorrelation. The inclusion of time-based features such as lag values and rolling averages enhances the model’s ability to predict short-term variations in soil moisture, which is essential for effective irrigation management.

This study aims to identify how organisations overcome the challenges to improving the quality of the services they deliver, which is an argument made by advocates of the agile approaches, and comparative analysis is a systematic approach used to evaluate and compare two or more entities, variables, or options to identify similarities, differences. The articles utilized machine learning models to predict target values of soil-related data such as Soil Moisture Deficit Index, Soil Moisture Agricultural, and this study evaluates the performance of five machine learning algorithms Decision Tree, -Nearest Neighbor, Random Forest, Light, which achieved the highest accuracy. Traditional approaches often rely on static or simplistic empirical models that fail to adapt to the complex and fluctuating conditions of, and existing smart home benchmarks do not model how device actions continuously affect environmental variables such as temperature and humidity. Soil moisture is a key variable for a range of hydrological and ecological processes, yet capturing its small-scale variability and preferential flow, and environmental factors like rainfall and air temperature variations weakened the relationship between canopy spectral data and soil moisture, increasing. Operational agencies face significant challenges related to the verification and evaluation of weather forecasts, and the accuracy and availability of historical project data can significantly impact the performance of prediction models, and even under normal conditions, the station's data do not fully reflect the weather on land or at sea due to its proximity to a 180-ft-tall cliff.

Several global satellite missions have advanced the availability of remotely sensed soil moisture products. The NASA Soil Moisture Active Passive (SMAP) satellite mission was launched to provide global mapping of high-resolution soil moisture. Satellite imagery has traditionally been valuable for large-scale environmental monitoring. Satellite technology plays a foundational role in modern environmental monitoring by offering consistent, scalable, and real-time observations. SMAP will provide information on soil moisture, which is critical for healthy plant growth. However, the overall objective of SMAP is to monitor global soil moisture mapping, and the data provided is often complex and not directly usable by farmers without further processing.

Combining satellite imagery with machine learning has the potential to address global challenges by remotely estimating socioeconomic and other factors, contributing to the ongoing efforts to employ ML and artificial intelligence, and this study offers current perspectives on existing satellite-based approaches. This integration allows for the development of predictive models that can account for various environmental factors and soil properties, and combining deep learning with dynamic model forecasts can substantially improve the skill of subseasonal predictions. AI can help with situational awareness, forecasting, damage assessment, and provide decision support, and the concept of an AI agent refers to a system or program that is capable of autonomously performing tasks on behalf of a user or another system. Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices, and the findings highlight the potential of ML and AI-based approaches in advancing soil fertility prediction and crop recommendation systems.

A system with effective usability can save time, reduce errors, and improve end-user satisfaction, and usability is a quality attribute that assesses how easy user interfaces are to use. Users of systems have been struggling to error-check their outputs successfully, and Expert Systems are often used to help non-experts when a human expert is too expensive. The use of soil moisture sensor data offers practical solutions for real-world problems, and soil moisture memory is demonstrated with data from a modeling system. It can help farmers optimize the use of resources, and Human–computer interaction is essential for optimizing smart greenhouse management and for fostering efficient and sustainable agricultural practices.

By integrating information from different modalities, they generate more comprehensive and accurate insights, and the integration of artificial intelligence technologies into managerial decision-making processes is transforming strategic planning across industries, which provides a robust capability and aids clinicians in making quicker and more accurate decisions. Machine learning models can be combined into a hybrid system to successfully predict out-of-distribution test instances and key ML algorithms such as object detection, semantic segmentation, pose estimation, and anomaly detection help enable these space applications, which can generalize across diverse prediction tasks. The integration of Big Data and Artificial Intelligence into agriculture has ushered in a new era of precision farming, enabling farmers to take a more organised approach to cultivation and adequately forecast the outcome with a high level of accuracy, which offers several key benefits such as enabling better crop health, yield management, and meeting sustainability targets.

Despite these limitations, satellite platforms have limitations such as reduced spatial resolution compared to UAVs, sensitivity to cloud cover, and limited penetration depth. Most soil quality measurements have been limited to laboratory-based methods that suffer from time delay, high cost, intensive labour requirement. However, there is significant interest in SM and the Mesonet does not directly measure SM. The estimate is produced using a constrained omniscient model that mandates only that identical observations receive identical predictions. We explore empirically the relationship between six data quality dimensions and the performance of 19 popular machine learning algorithms. This study compares the ways training dataset size and interactions affect the performance of those prediction models. Integrating real-time data monitoring with dynamic LCAs offers a promising approach to address these limitations and provide more comprehensive and accurate results. Sensor-based data collection of human behaviour enables real-time monitoring of behavioural and physiological markers. Future work could focus on extending the applicability of the framework by incorporating advanced predictive analytics and optimizing cross-modal consistency.

VII. CONCLUSION

Machine Learning-based Decision support systems provide data-driven insights to farmers for real-time decision- making toward improved productivity and estimating sub-daily resolution soil moisture using Fengyun satellite data and machine learning. This research proposes the development of an advanced AI-Driven Decision Support System for Smart and Sustainable e-Agriculture, leveraging machine learning, and some systems integrate weather data with soil moisture levels to provide irrigation recommendations. This study investigates soil moisture variability by integrating NASA's Soil Moisture Active Passive (SMAP) datasets and 'AGRO AI' has been developed by leveraging artificial intelligence (AI) combined with NASA's Earth Observing System (EOS) satellite data. In this study, the Random Forest (RF) ensemble learning algorithm is tested to demonstrate the capabilities and advantages of ML for RZSM estimation and Random Forest (RF) and Artificial Neural Network (ANN) are effective machine learning algorithms for predicting soil water content.

The proposed system is designed to monitor and measure various environmental parameters such as temperature, humidity, atmospheric pressure. This work integrates more of the predictive potential of climate-change models by exploring other environmental variables, such as humidity and temperature. Remote-sensing technologies have transformed soil-moisture estimation by enabling large-scale, high-resolution, and continuous monitoring. This approach reduces error compared to linear regression, making it more suitable for real-world applications where soil moisture behaviour is complex. The front-end developed interface offers results in the form of yield estimation, crop recommendation, fertilizer suggestions. Decision support tools have emerged as essential tools to help make accurate, evidence-based agricultural decisions aimed at enhancing soil health and fertility.

The study highlights the advantages and disadvantages of deep learning applications in this area. This system can ease the development of context-aware mobile applications, and enable context-aware mobile crowdsensing considering environmental, personal, and other factors, which increases organisation's adaptability to environmental changes. Unlike traditional AI systems that execute predefined algorithms within constraints, AI agents possess the capacity to autonomously perceive, analyse context, selecting actions, and executing decisions in real time. Machine learning is a subset of AI that involves training computer algorithms to learn patterns in data and make predictions or decisions based on the data. The IWeCASF-FEADM integrated approach for regional soil moisture estimation is developed in this work. Combining thermal and optical Remote sensing data can enhance the accuracy of soil moisture assessment by considering the weather conditions and soil and crop features to be essential to determine the soil moisture content.

This review paper presents a novel analysis of state-of-the-art AI and ML techniques, which have limitations, including reliance on historical satellite data, and this paper reviews how AI is being used to analyse extreme climate events. This study aims to enhance soil moisture prediction by integrating in situ observations with machine and deep learning models, and integrating machine learning models with ground sensors can enhance soil moisture prediction. The effectiveness of such a decision support system is scalable and effective in enhancing agricultural progress, and recent intelligent decision support systems that integrate crop mechanistic models with machine learning and deep learning have been presented.

REFERENCES

[1] D. Entekhabi et al., “The Soil Moisture Active Passive (SMAP) Mission,” Proceedings of the IEEE, vol. 98, no. 5, pp. 704–716, May 2010.

[2] N. Das, R. Entekhabi, and E. Njoku, “An Algorithm for Estimation of Soil Moisture Using Passive Microwave Observations,” IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no. 1, pp. 310–322, Jan. 2011.

[3] J. C. Shi et al., “Soil Moisture Retrieval Using Satellite Remote Sensing: Current Status and Challenges,” Remote Sensing of Environment, vol. 215, pp. 305–320, Sep. 2018.

[4] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[5] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed. Sebastopol, CA, USA: O’Reilly Media, 2019.

[6] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[7] M. Reichstein et al., “Deep Learning and Process Understanding for Data-Driven Earth System Science,” Nature, vol. 566, pp. 195–204, Feb. 2019.

[8] S. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” Advances in Neural Information Processing Systems (NeurIPS), pp. 4765–4774, 2017.

[9] FAO, “Water for Sustainable Food and Agriculture,” Food and Agriculture Organization of the United Nations, Rome, Italy, Tech. Rep., 2017.

[10] NASA Jet Propulsion Laboratory, “SMAP Level-3 Soil Moisture Product Specification,” [Online]. Available: https://smap.jpl.nasa.gov/