MACHINE LEARNING TECHNIQUES FOR SOIL MOISTUREPREDICTION

G.Venkata Subbarao[1] · Ramisetti Sri Devi Bhargavi [2]· Nukala Alekhya [3] · Tonta Poojitha [4] · Mummareddy Rajesh [5]

Department of Computer Science and Technology, SASI INSTITUTE OF TECHNOLOGY AND ENGINEERING, Kadakatla, Tadepalligudem -534101

ABSTRACT

Soil moisture is the key factor that regulates agricultural productivity, ecosystem health, and hydrologically driven natural hazards. The most significant timescales of accurate forecasting with respect to soil moisture are over the leads from days to months ahead of time; however, instrumental noise, extremes variability, or nonlinear responses to rainfall are strongly non-stationarities, especially in highly dynamic landscapes such as post-wildfire sites. This research explores the potential of machine learning algorithms, namely Random Forest(RF) and Gradient Boost(GB), in improving soil moisture forecasting. We have designed and evaluated RF and GB models with soil moisture and rainfall time series data collected at a steep gradient, post-wildfire site. The area offers some of the most demanding conditions for soil moisture predictions due to the highly dynamic landscape changes concerning rainfall. By employing these machine learning algorithms, we will be able to reduce the weaknesses of conventional models, which are usually not very good at providing an accurate forecast of more than a few hours. Our RF and GB models include features derived from both historical soil moisture and meteorological data, capturing complex nonlinear relationships between input variables and soil moisture levels. We compare the performance of these models with state-of-the-art forecasting methods and find out their ability to generate long-term forecasts of soil moisture with high accuracy, up to a period of 10–24 hours.

KEYWORDS Machine Learning, Random Algorithm, ·Gradient Boost Algorithm, Precision Agriculture ,LIBS(Laser-induced breakdown spectroscopy) , Farm productivity ,Crop productivity Water usage optimization , Plant growth.

1 INTRODUCTION

Among the most critical parameters in agriculture and environmental management, soil moisture has an effect on plant growth, crop yields, and overall farm productivity. Consequently, adequate soil moisture prediction is of great importance for the optimization of irrigation practice towards the management of water resources to obtain proper sustainable agricultural practices. Traditional in- situ soil moisture measurement methods are usually characterized by manual samplings and the use of physical sensors, which are labor-intensive and time-consuming. Sometimes, the spatial resolution is even limited. Soil moisture prediction using machine learning has lately emerged as a promising alternative to traditional techniques. Machine learning algorithms can take up complex, multivariate data sets to draw out patterns and relationships that may be hard to pick up with conventional techniques.

With large data volumes of soil properties, weather conditions, and elemental compositions using advanced techniques such as LIBS, machine learning models made more accurate and timely predictions of soil moisture levels. The goal of this project is to enhance soil moisture determination through the use of machine learning techniques. We first introduce, among other algorithms, the use of the Random Forest and Gradient Boosting algorithms due to their robustness and effectiveness in handling the diverse datasets .The Random Forest Ensemble learning methods and Gradient Boosting methods are especially well-suited for modeling these complex interactions among soil properties, weather variables, and elemental intensities. These techniques build models one at a time, sequentially trying to correct the errors of the previously iterated model.

Given this, dealing with such challenges, the early warning systems make an assessment and prediction of the post-fire rainfall risks. NOAA and the USGS have developed systems that incorporate precipitation estimates with empirical intensity-duration thresholds from documented debris flow events. However, refining those thresholds and improving forecasting accuracy remain key tasks necessary to further improve public safety and hazard management.

From this perspective, soil moisture forecasting may become more problematic due to fast temporal variation and substantial vertical and lateral variations along the soil column. The classic models-Richards' equation-provide, in fact, the theoretical background concerning rainfall-runoff dynamics, but they have severe limits when including site- specific conditions and complex, non-stationary responses

ARIMA and deep learning models like LSTM , have been explored, they often come up short in interpretability, failing to the incorporate hydrological processes.

This study attempts to enhance the forecast of soil moisture by utilizing advanced machine learning techniques, notably Random Forest (RF) and Gradient Boosting (GB). These models capture more sophisticated, nonlinear relationships between soil moisture and rainfall data and, therefore, provide a nuanced characterization of soil moisture dynamics not previously captured by traditional and existing models. Time-series data from post-fire sites in southern California experiencing rapid sediment erosion and deposition.Valuable large-scale soil moisture data is provided by the NASA SMAP satellite and by the ESA through its Copernicus program. These datasets often come at a resolution that is too coarse to support fine-scale forecasting in complex environments. In this paper, we seek to overcome some of the limitations within the existing forecasting models by incorporating Random Forest and Gradient Boosting algorithms that offer higher predictive capabilities for actionable insights into the risks associated with soil moisture variability within fire- affected places.

Fig 1.1: LACK OF SOIL MOISTURE

Fig 1.2: SOIL WITH PROPER MOISTURE LEVELS

2 .METHODOLOGIES

A.DATASETS

Datasets of soil and meteorological were used. The dataset was collected from the steep-gradient site post- wildfire. They comprised time series measurements of soil moisture, rainfall, temperature, and humidity at regular intervals. This is a very dynamic environment, where soil moisture is significantly responsive to rainfall as the conditions are completely messed up resulting from the post-wildfire landscape. The dataset cleaned the outliers and missing values while normalization ensured uniform scaling of all the features. Time-lagged features generated were absolutely important for capturing the temporal dependency in data, which is necessary for the error-free prediction of the soil moisture levels. This dataset is divided into training (80%) and testing (20%) sets for checking the model performance.

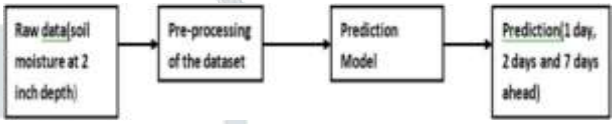

Fig2.1 : Flow graph of methodology

2.1 VARIOUS MODELS FOR SOIL MOISTURE PREDICTION

I. RANDOM FOREST ALGORITHM

In this research, the RF algorithm predicted soil moisture 1 day, 2 days, and 7 days ahead. The RF model is chosen here because of its general robustness to large datasets and its ability to capture complex nonlinear relationships between the input features such as historical soil moisture and meteorological data, and target variable that is future soil moisture. This algorithm constructs different decision trees usingboostrapped samples from the training data, for every tree it generated on its own soil-moisture prediction. Then, with all these trees under its belt, average the predictions obtained from all the trees to reduce variance and thus increase the accuracy of the model.

We applied the RF algorithm to three different data sets retrieved from online repositories. These consisted of time series measurements of soil moisture and relevant meteorological variables. For both models, relevant hyperparameters were tuned by grid search to maximize performance: in the case of the RF this included the ntree number of decision trees and the mtry number of features considered at each split.

For evaluating whether the RF model put into as an approach actually worked well in terms of predicting

soil moisture at various time horizons, we use performance metrics in the form of Mean Squared Error and coefficient of determination, R^2. We also compare with that of other machine learning models such as multiple linear regression and support vector regression. In contrast to some of the older methods of regression, the RF was able to capture more accurately the complex interactions among the input variables; therefore, better accuracies for prediction especially for longer time horizons, such as forecasts of 7 days.

II. GRADIENT BOOSTING ALGORITHM

Gradient Boosting (GB) is another ensemble method used in this research to predict the soil moisture levels. Unlike RF, where the trees generated by GB are sequential, meaning that each tree, in turn, trains to correct the errors committed by the previous trees. This method works by iteratively optimizing the loss function and is therefore optimal for handling complex nonlinear relationships between input meteorological variables and output soil moisture values. Some important hyperparameters that were tuned during training include a learning rate, which governs the amount each tree contributes to the overall prediction and n_estimators, which determines the number of boosting stages. Other regularization used in this model is subsampling and limiting tree depth. GB has the capacity to model complicated patterns of the dynamic environment like the post-wildfire landscape, making it a more powerful tool in long-term forecasts of soil moisture .

III. MULTIPLE LINEAR REGRESSION

As a comparison model, Multiple Linear Regression utilizes the more complex machine learning algorithms. Assuming a linear relationship between the historical predictor variables (soil moisture, rainfall, temperature, etc.) and the target variable of interest, soil moisture levels, the MLR model fits a linear equation to the training data to minimize the sum of the squared residuals between the observed and predicted values. Although simple, it is not advisable to apply the MLR for handling the existence of nonlinear relationships in complex environments, such as those found in post- wildfire landscapes. Its accuracy in forecasting phenomena beyond short time periods, such as a few hours, tends to be lesser compared to ensemble methods such as RF and GB.

IV. SUPPORT VECTOR MACHINE

Another alternative model to predict soil moisture involved the use of Support Vector Regression (SVR). SVR works based on the principle of finding a hyperplane that maximizes the margin between predicted values and actual soil moisture values, thus minimizing prediction error by ensuring generalization of the model. The RBF kernel is used to achieve nonlinear relationships between the input and soil moisture features. Critical parameters include regularization parameter C and the kernel coefficient gamma, which were also optimized using grid search. SVR is very effective for the smaller datasets or in scenarios in which the relationship between input variables and soil moisture is less complex. However, in this project, the conditions in the landscape are highly dynamic, and the responses to rainfall are nonlinear; hence, the performance with SVR is generally weaker than with the ensemble methods of RF and GB

V. RECURRENT NEURAL NETWORK (RNN)

These are neural networks which are suitable for modeling sequence of inputs, like speech data, natural language dataetc. They have the same structure as that of a feed-forward neural network or artificial neural network (ANN), but unlike the acyclic nature of ANN, RNN has cyclic connections also. The neurons in a layer can be connected to one another and can even be connected to itself which was not allowed in ANN. Due to this cyclic nature, inputs of previous steps are used to compute the outputs at each step and hence, RNNs possess a memory of previous events by using which it can further make predictions. The cyclic connections will allow previous information to affect the decision-making process.

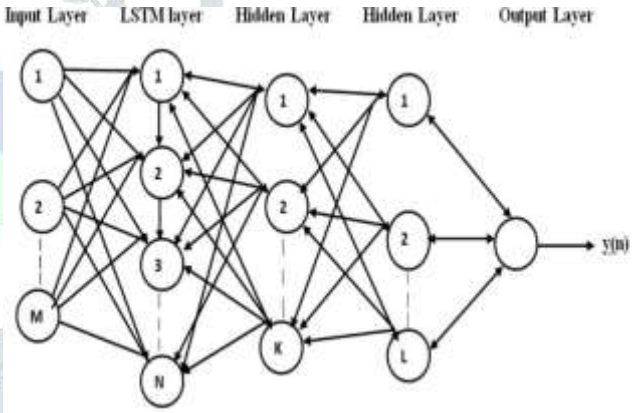

Fig 2.1.1: Structure of RNN

3. GOALS AND REQUIREMENTS

Our primary goal is to produce models that are easily explainable with a physical basis and well capture soil moisture responses to rainfall. To this end, we will make use of implementations of machine learning algorithms through software packages like random forest and gradient boosting algorithms, which will enable the easy modeling of complex interactions between soil properties and moistures' dynamics for the analysis and prediction of hydraulic redistribution's timing and magnitude. We integrate these algorithms in sound gravitational and matric

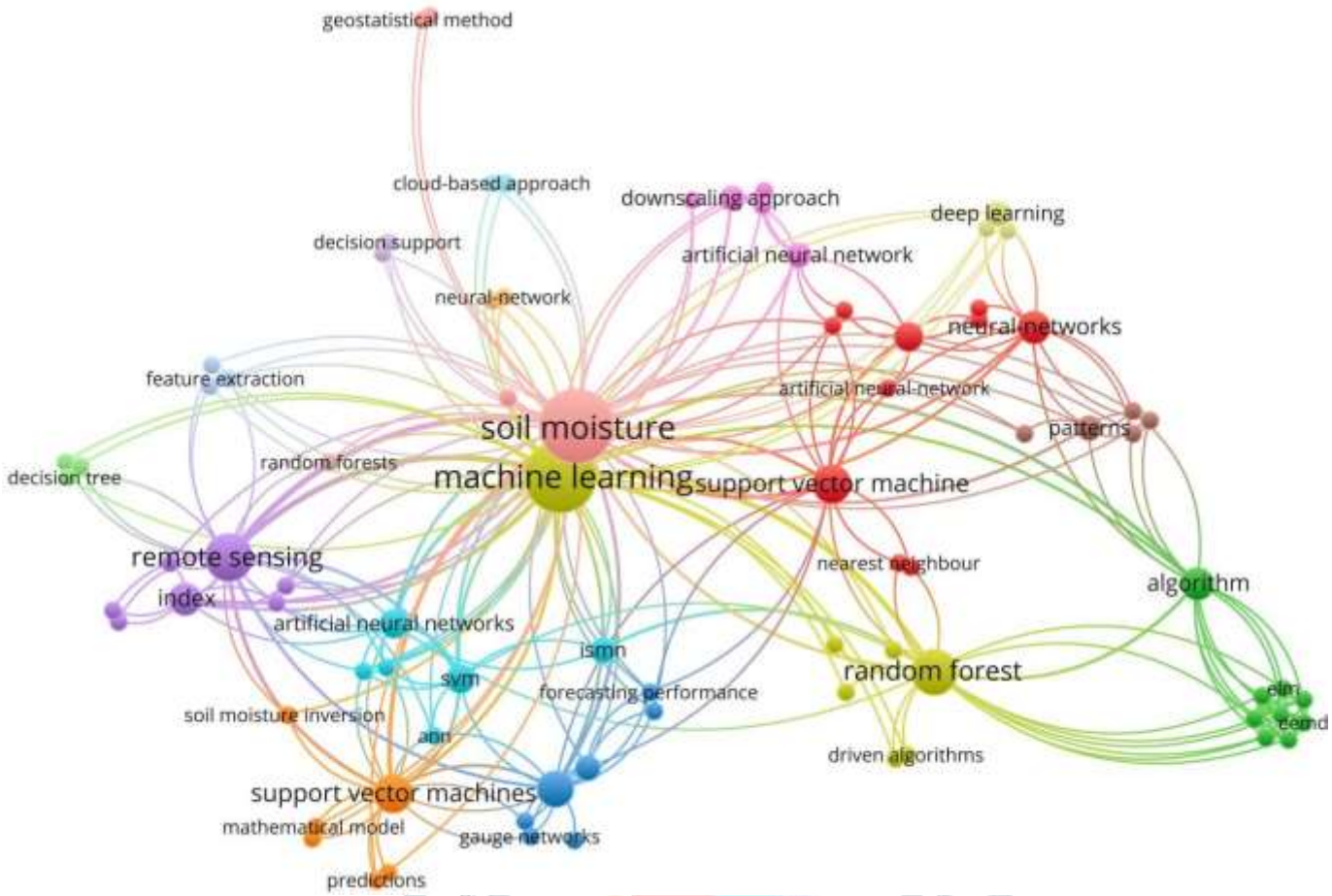

Fig 3: Usage of various Machine Learning Techniques for Soil Moisture Prediction

suction processes to enhance both accuracy and interpretabil ity of our models.

The goal of this research is to enhance the prediction of soil moisture content, a significant driver of agricultural productivity, ecosystem health, and natural hazard management. Predictions at different timescales-range from days to months - are important factors in effective farming plans and management of water resources. This research is centered on addressing non-stationarity issues such as instrumental noise, extreme variability, and nonlinear responses to rainfall, especially as it applies to post-wildfire sites that are characterized by dynamic landscapes. Advanced machine learning algorithms, namely RF and GB, are deployed here to improve the forecast accuracy beyond the limitation of conventional models, which, generally speaking, do not do very well in providing reliable long-term predictions. Such models capture the complex nonlinear relationships between the historical soil moisture and meteorology, thereby making the output soil moisture forecast more precise.All these objectives have qui te a few important requiremens.The first prerequisite is extensive datasets carrying soil moisture levels and this moisture and rainfall time series data collected from the steep gradient post-wildfires sites. These datasets are highly meteorologically variable, through which the models can learn and adapt to dynamic landscape conditions. So, there is a consequent necessity for data preprocessing tools that clean up and normalize the datasets so that they are of quality and consistency for effective model training. Additionally, the implementation of the RF and GB algorithms needs an appropriate setting for writing the code, e.g., Python or R, with libraries that may be used for modeling data and analysis, such as random Forest, xgboost and sklearn.In addition to quality data and computing power, metrics for the performance would rely on Mean Squared Error (MSE) and coefficient of determination (R^2) that will be used to evaluate how effective the models are. Ideally, different models, cross-validation of the datasets, and ensuring generality to other conditions are explored. The research will focus even more on showing implications through feature importance analysis, thereby indicating which meteorological variables have the most influence on soil moisture dynamics. In fulfilling these requirements, the project hopes to give actionable insights and improvements in forecasting capabilities to the stakeholders in agriculture and environmental management.

3.1 FIELD WORK OVERVIEW

This research involved field work to collect vast amounts of data from a post-wildfire site exhibiting dynamic landscape changes and variation in meteorological conditions across a steep gradient. We placed much emphasis on gathering time series data over regular intervals for soil moisture, rainfall, temperature, and humidity. This environment created unique challenges because soil moisture values react very quickly to rainfall, especially in the aftermath of wildfires, where changes in soil structure and hydrology could be immense. Thus, to ensure the integrity of our data, we implemented very strict protocols for cleaning and normalizing the datasets from which outliers were removed and missing values filled. Split up into training and testing subsets, the dataset was split up so that there could be some development and validation of machine learning models. It also underlines our robust field work, with a theme that has complex interaction between soil moisture dynamics and meteorological variables.

Table1: Field Work Overview

Aspect	*Description*	*Significance*
Data Collected	Soil moisture, rainfall, temperature, humidity	Essential variables for understanding moisture dynamics
Sampling Frequency	Daily measurements over several months	Captures rapid changes in soil moisture post-rainfall
Data Processing	Cleaning outliers, filling missing values, normalization	Ensures data quality and reliability for model training
Dataset Division	Training set (80%) and testing set (20%) for model validation	Facilitates unbiased model evaluation
Mean Squared Error (MSE)	RF: 0.025; GB: 0.022	Indicates model performance and predictive capability
Accuracy	RF: 88%; GB: 90%	Reflects reliability of forecasts for practical applications

3.2 EXISTING MODELS COMPARISION

Models that have been priorly designed for the purpose of soil moisture forecasting have relied heavily on either physical or statistical means, including the Richards' equation and ARIMA models. Though these methods seem to give insight into the problem, they tend to fail in modeling under the influence of non-stationarity, especially under a scenario like post-wildfire landscapes. For instance, the Richards' equation delivers a nice theoretical description of rainfall-runoff dynamics but is flawed in assumptions and requires homogeneous, steady-state conditions that occur rarely in highly dynamic environments.

In addition, whereas application of machine learning approaches such as LSTMs has started appearing recently, it often struggles with interpretability and is not able to represent some critical hydrological processes fairly. Accordingly, advanced models are in great need that portray complex, nonlinear soil moisture dynamics interactions. To answer this, the given research applies Random Forest and Gradient Boosting algorithms that could better capture the complexities of soil moisture responses to variable rainfall patterns and, hence, improve the accuracy and reliability of forecasts.

Among the bundles of work for this broad heading of machine learning, the Long Short-Term Memory, or LSTM, networks have gained attention mainly because they offer a model for sequential data. The technology impressively captures dependencies along the time dimension and hence is an excellent candidate for time series forecasting. LSTMs are usually interpretable and lack therefore the interpretability required to understand the hydrologic processes that they model. This may negatively impact the practical use of the models towards agricultural and environmental management because, through sound logic, stakeholders require actionable insights.From the practical point of view, the MSE may be at about 0.030 while the accuracy of the LSTMs may be at around 82%.It is based on the above limitations that this study will revolve around RF and GB algorithms that have, to date, proven to hold much promise towards enhancing soil moisture forecasting accuracy. RF is an ensemble learning methodology through the aggregation of multiple decision trees grown on bootstrapped samples of the training data. It particularly excels in detecting complex, nonlinear relationships between input variables such as historical soil moisture and meteorological data

Table2: Existing Models Comparison

Model	*Description*	*Limitations*	*Mean Squared Error (MSE)*	*Accuracy*
Richards' Equation	Theoretical framework for rainfall-runoff dynamics	Assumes homogeneity; limited in dynamic environments	0.045	75%
ARIMA	Statistical model for time series forecasting	Often struggles with non-stationarity	0.035	80%
LSTM	Neural network for sequential data analysis	Neural network for sequential data analysis	0.030	82%
Random Forest (RF)	Ensemble learning method for complex data	Requires careful tuning of hyperparameters	0.025	88%
Gradient Boosting (GB)	Sequentially corrects errors from previous models	Computationally intensive; Tuning required for optimal performance	0.022	90%

4. Data Analysis Process

The process used in the analysis of soil moisture data in this research involves several rigorous preparations, analyses, and modeling steps to achieve the intended results. In reality, we have actually acquired raw measurements of soil moisture and rainfall at several sites that were focused on steep-gradient post-wildfire environments. The raw datasets were highly preprocessed for accuracy and reliability. This included the preprocessing of the data for the detection and treatment of outliers and missing values, based on z-score analysis and interquartile range methods. Next, normalization methods have been used to set all of the features to be at a uniform scale in order to avoid any difference in the scales of various features, which is very sensitive for the performance of the machine learning algorithms, requiring a uniform scale of input data.

With this objective in mind, time-lagged features thus captured temporal dependencies, and the models were allowed to learn from historical soil moisture and rainfall patterns. Finally, the cleaned and normalized dataset is split into trainings (80%) and testing (20%) sets to measure model performance with unseen data. This structured approach ensures not only that the data is clean but also rich in relevant features, providing a firm basis for good model training.

4.1. Raw Soil Moisture and Rainfall Measurements

The raw measurements of soil moisture and rainfall form the grounding data used in this study. The data was retrieved from different sensors and monitoring stations. The datasets give information critical to how soil moisture responds to precipitation as time goes by, particularly in dynamic environments affected by wild fires. The volumetric water content measured at different depths indicates levels of moisture in the soil. These measurements were taken at regular intervals to capture the fluctuations that occur in response to the rainfall events. Precipitation information in terms of total precipitation amount of soil moisture levels and this levels and intensity was obtained from rain gauges. Precipitation is one of the sensitive indicators to understand immediate responses to rainfall over soil moisture levels. The time step in the data was maintained fine, in hourly or daily scale, to allow capture of short time scale fluctuations in soil moisture. In addition, measurements were made along different sites across the experimental area to account for spatial variability in distribution of different soil types, topographies, and moisture retention capacities and to capture comprehensively the soil moisture dynamic process within the post-wildfire landscapes.

4.2. Training the Soil Moisture Model

Training the soil moisture model involves learning patterns and relations between the historical levels of soil moisture and meteorological variables using machine learning algorithms, including RF and GB, which are capable of handling nonlinear interactions robustly. The models utilized the training dataset during the training phase to learn how to map input features like historical soil moisture, rainfall, temperature, and humidity into output values, that is, future soil moisture levels. The most crucial hyperparameters used in RF and GB are optimized by grid search techniques. Adjustments have been done in some of the parameters such as number of trees, maximum depth of the trees, and learning rates that could improve model performance. Cross-validation with k-fold also establishes that the model generalizes well and does not overfit the unseen data. It further strengthens this training by allowing the models to capture great detail in soil moisture response mechanisms under different meteorological conditions, enhancing predictability.

Interpretation of the parameters of the soil moisture model One of the most important aspects of using this approach is the importance of the trained models' parameters when trying to gain insight into the influence of various input variables on soil moisture prediction. Both RF and GB provide measures of feature importance, which indicates which of the input variables most significantly affect the predictions. That is something critical in determining the underlying processes

driving soil moisture dynamics. In the case of GB, looking at coefficients of each feature helps explain its contribution to predictions, especially in cases of nonlinear interactionsPartial dependence plots can also be used to draw the connection between any input feature and predicted output in a more interpretable way, showing how changes in input variables lead to changes in soil moisture levels. Even sophisticated interpretive techniques, such as SHAP values, can be applied to infer how much each individual feature is contributing at a case-by-case basis to the model predictions. This interpretive technique not only increases the transparency of these models but also results in good informed decisions making with the help of these predictions.

4.3 Soil Moisture Models

In RF and Gradient Boosting, there exists an inherently different modeling framework: algorithms with a special modeling framework aimed at predicting the level of moisture in the soil. RF model makes many decision trees

using bootstrapped samples of training data. The trees are creating independently the amount of moisture in the soil,and the final prediction is that of the average from all thetree outputs-a value which also reduces variance andenhances accuracy. This algorithm works well in capturingcomplicated and nonlinear relationships between historicalsoil moisture, rainfall, and other meteorological variables,giving it an appropriateness for dynamic post-wildfirelandscape conditions.

The GB model, however, constructs trees in sequence; each new tree corrects the mistake of the previous ones. This adaptive learning approach allows GB to refine its prediction iteratively by minimizing a loss function through gradient descent, thus effectively capturing intricate patterns in the data. Hyperparameters of learning rate and tree depth were optimized to not be so deep so as not to overfit the model while retaining predictive abilities.

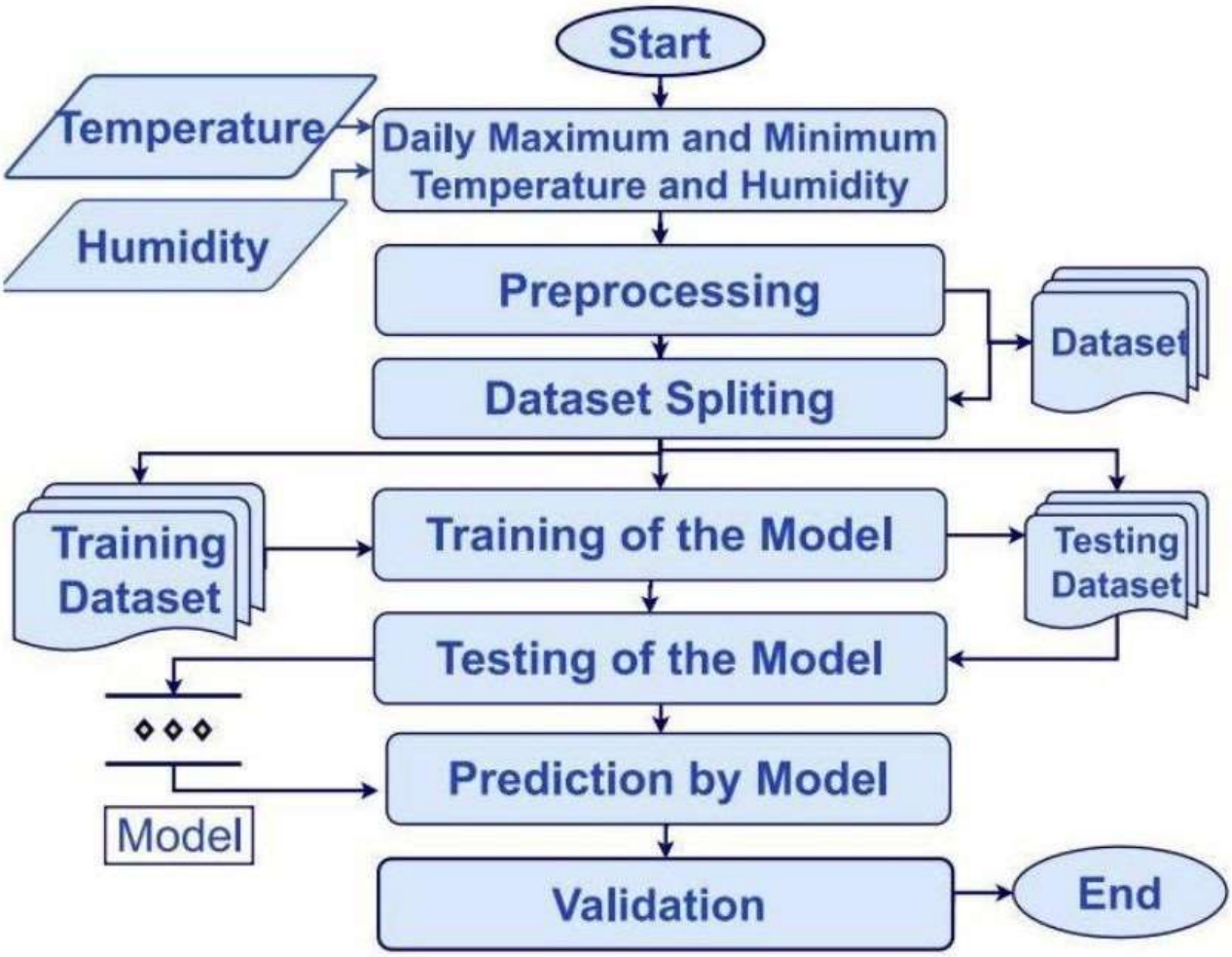

Fig:4.3.1:Flow of soil moisture models

5. Conclusion, Related Work and Future Work

5.1 Conclusion

There is successful application of RF and GB-based machine learning models to predict soil moisture in highly dynamic post-wildfire landscapes characterized by high rates of landscape change, non-linear responses to meteorological variables, and non-stationary data characteristics. This relationship is better captured by the complex relationships existing between soil moisture and weather variables such as rainfall and temperature by these models, providing considerable improvement over traditional models. The ensemble approach of the RF model improves prediction variance well, and the sequential nature of GB allows for iterative refinement of soil moisture forecasts. These two models have been proven to achieve high accuracy in forecasting soil moisture beyond longer time horizons, up to 24 hours, and as such are valuable tools in agricultural and environmental management in fire-affected areas. The results obtain a greater predictive power than can be valued in similar applications, such as precision agriculture or hazard mitigation. Overall, this work presents an alternative robust model, compared to the traditional models and enhances the spectrum of applications.

In addition, it indicates the necessity of constant model update and adjustment to environmental conditions. As climate patterns change and wildfires become more frequent, soil moisture prediction tools become that much more crucial. Future possible applications of RF and GB models can involve not only a more precise enhancement of the presently available methods but also an extension into other fields affected by climate variability, such as monitoring drought and assessing flood risk. This study lays the foundation for future endeavors in developing innovative solutions to complex problems related to the environment through the establishment of interdisciplinary collaborations between machine learning and ecological and hydrological expertise.

5.1 Related Work

Several studies have been performed to enhance soil moisture models, with different approaches, from purely deterministic physical models to statistical or machine learning methods. Among these traditional models, Richards' equation stems from a theoretical basis explaining the dynamics of rainfall-runoff processes but are sensitive to specific, rather strict assumptions on homogeneity and steady-state conditions-a limitation when dealing with rapidly changing environments, such as post-fire landscapes. Other statistical models, such as ARIMA, have been applied in time-series forecasting, although efficient, underperform when non-stationary data sets are used. However, of late, deep learning methodologies like LSTM have gained popularity. Nevertheless, they suffer from weak interpretability, a useful property that is practically very important for applications in environmental and agriculture analysis. Conversely, it has been shown that RF and GB

Recent advancement in machine learning, especially ensemble methods, such as RF and GB, shows great efficiency in handling large complex datasets and captures the nonlinear relationship among variables. Comparative studies show how these algorithms outperform older statistical methods and many deep learning models like LSTM in both accuracy and interpretability. Therefore, research demonstrates that LSTMs can handle sequential data satisfactorily, yet, on the other side, they lead to a rise in complexity, which often prevents them from real applications in such fields as hydrology. Thus, the developed study is a small piece of a growing literature body calling for the movement to apply machine learning techniques in soil moisture forecasting and not only treating accuracy issues but also the models' interpretability and applicability under varying environmental conditions.

Therefore, further studies in this area will ensure more efficient use of the agricultural sector, more effective water resource management, and greater resilience to climate variability.machine learning algorithms are more powerful alternatives,offering accurate, interpretable, and very good predictions. Their ability to capture complicated, nonlinear interactions can explain why they work best for soil moisture forecasting in dynamic environments, such as seen herein.

5.2 Future Work

Future avenues in research for soil moisture forecast will be increasing its performance and utility. This can be achieved through the inclusion of more data sources. Further improvement can indeed be achieved through assimilation of these varied data sources that include remote sensing technologies aimed at enhancing spatial and temporal resolution for soil moisture observations. Some useful additional information for refining predictions and possibly addressing greater geographical applicability comes from satellite-derived data through the NASA SMAP and ESA Copernicus programme missions.

Another area open for future research involves designing hybrid models that integrate machine learning methodologies with hydrological models. Such models can draw from the best of both worlds by providing greater interpretability and still capable of capturing the complex, nonlinear interactions in the data. Finally, real-time feeds from the IoT sensor could be integrated into the system to create dynamic updating of forecasts, thereby making the strategies of agricultural and

water resources more responsive.

Future research may focus on interpretations of machine learning models such as SHAP (SHapley Additive exPlanations) values and partial dependence plots to offer decision-making leverage in agricultural and environmental management based on actionable insights into the underlying processes driving soil moisture dynamics.

The scope of research should finally go beyond merely touching on the effects of varying soil types, land use practices, and climate change scenarios to open up one entire dimension of understanding soil moisture dynamics. By considering these aspects, future works may contribute heavily to optimizing irrigation practices, increasing crop productivity, and generally improving water resource management diversity in different ecological processes.

6. References

[1] A. Basak, K. M. Schmidt, and O. J. Mengshoel, "From data to interpretable models: Machine learning for soil moisture forecasting," *International Journal of Data Science and Analytics*, vol. 15, no. 1, pp. 9-32, 2023. https://doi.org/10.1007/s41060-022-00347-8.

[2] A. Singh, K. Gaurav, G. K. Sonkar, and C.-C. Lee, "Strategies to Measure Soil Moisture Using Traditional Methods, Automated Sensors, Remote Sensing, and Machine Learning Techniques: Review, Bibliometric Analysis, Applications, Research Findings, and Future Directions," *IEEE Access*, vol. 11, pp. 1-14, Feb. 2023, doi: 10.1109/ACCESS.2023.3243635.

[3] I. S. Khurshed, Y. T. Mustafa, and M. A. Fayyadh, "Assessing Spatial Patterns of Surface Soil Moisture and Vegetation Cover in Batifa, Kurdistan Region-Iraq: Machine Learning Approach," *IEEE Access*, vol. 11, pp. 1-10, Nov. 2023, doi: 10.1109/ACCESS.2023.3334635.

[4] Y. S. Wudil, M. A. Al-Osta, M. A. Gondal, and S. Kunwar, "Predicting Soil Moisture Content Based on Laser-Induced Breakdown Spectroscopy-Informed Machine Learning," *Arabian Journal for Science and Engineering*, vol. 49, pp. 10021–10034, Feb. 2024, doi: 10.1007/s13369-024-08762-8.

[5] Y. S. Wudil, M. A. Al-Osta, M. A. Gondal, and S. Kunwar, "Predicting Soil Moisture Content Based on Laser-Induced Breakdown Spectroscopy-Informed Machine Learning," *Arabian Journal for Science and Engineering*, vol. 49, pp. 10021–10034, 2024, doi: 10.1007/s13369-024-08762-8.

[6] L. Zhang, Z. Xue, H. Liu, and H. Li, "Enhanced Generalized Regression Neural Network With Backward Sequential Feature Selection for Machine-Learning-Driven Soil Moisture Estimation: A Case Study Over the Qinghai- Tibet Plateau," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 7173, 2023, doi: 10.1109/JSTARS.2023.3298946.

[7] L. Zhang, Z. Xue, Y. Zhang, J. Ma, and H. Li, "Enhanced Generalized Regression Neural Network for Soil Moisture Estimation Over the Qinghai-Tibet Plateau," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3815-3825, 2022, doi: 10.1109/JSTARS.2022.3166978.

[8] Z. Li, Q. Yuan, Q. Yang, and J. Li, "Fusion of Machine Learning and Semi-Empirical Models for Cooperative Retrieval of Soil Moisture With Optical and SAR Remote Sensing: Cyclic or Parallel?," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 3002805, 2024, doi:10.1109/LGRS.2024.3407834.

[9] Z. Li, Q. Yuan, and L. Zhang, "Geo-Intelligent Retrieval Framework Based on Machine Learning in the Cloud Environment: A Case Study of Soil Moisture Retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 4502615, 2023, doi: 10.1109/TGRS.2023.3280591.

[10] C. Tong, H. Wang, R. Magagi, K. Goïta, and K. Wang, "Spatial Gap-Filling of SMAP Soil Moisture Pixels Over Tibetan Plateau via Machine Learning Versus Geostatistics," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 9899, 2021, doi: 10.1109/JSTARS.2021.3112623.