Indian Institute of Remote Sensing, ISRO

SOIL MOISTURE ESTIMATION USING MICROWAVE AND MACHINE LEARNING

APRIL 22

SUBMITTED TO:

Dr. Dipanwita Halder

Scientist/Engineer-SF

Department of Agriculture and Soils

SUBMITTED BY:

ANUSHA PARIDA

ID-TS90729

M.Sc. Agriculture Analytics

http://www.iirs.gov.in

ACKNOWLEDGEMENTS

I would like to give my immense regards and appreciation to all the members, organisations for their continuous support and guidance that has helped me to complete this report. A special thanks to my guide, **Dr. Dipanwita Haldar**, Scientist/Engineer-SF, Department of Agriculture and Soils whose constant support and guidance has helped me to gain better insights of my research project. Dr. Halder provided me with invaluable guidance and expertise, steering me in the right direction and providing access to the necessary datasets for my research. Her mentorship has been crucial in shaping my understanding and approach to the project, and I am truly grateful for her support throughout this journey. Her expertise in SAR data has been instrumental in my research, providing valuable insights and guidance on how to effectively utilize and analyse SAR data for agricultural applications. Her deep understanding of the subject has not only enriched my project but has also broadened my knowledge in this field.

I am also thankful of our Course Co-ordinator **Dr. N.R. Patel** who guided me through this project and whose supervision, constant encouragement has been a source of motivation, and I am grateful for the opportunity to work for this amazing project.

I would also like to extend my gratitude towards **MTech students of IIRS** who provided me with the ground truth data. Their contribution has been invaluable, and without their data, this project would have been much more challenging. I appreciate their support and collaboration in this endeavour.

And lastly, I would like to convey my sincere thanks to all my seniors, friends especially to my fellow co-associates who were attached with me during this project to make the demanding time joyful but always efficient.

In the words of Henry David Thoreau, ‘I am grateful for what I am and have. My thanksgiving is perpetual’. This sentiment truly encapsulates my feelings of gratitude towards all those who have supported me on this journey.

CONTENTS

List of tables:

List of Figures:

ABBREVIATIONS USED:

SAR: Synthetic Aperature Radar

RVI: Radiation Vegetation Index

MPDI: Microwave Polarization Difference Index

GEE: Google Earth Engine

ML: Machine Learning

ANN: Artificial Neural Network

1.ABSTRACT

This research significantly contributes to the advancement of soil moisture estimation techniques, particularly in agricultural and environmental applications. The study's findings hold substantial promise for improving sustainable land management practices and enhancing climate change adaptation strategies. By integrating Synthetic Aperture Radar (SAR) data and machine learning algorithms, such as decision trees, the research achieves notable success in accurately estimating soil moisture levels.

The decision tree model stands out as a particularly effective tool in this study, demonstrating superior performance compared to other machine learning models. With a low Mean Squared Error (MSE) of 9.61 and high R-squared value of 0.89, the decision tree model consistently outperforms alternative methods. This superior performance can be attributed to the decision tree's ability to capture complex relationships between SAR data and soil moisture content. The model's interpretability and simplicity further enhance its practical utility, making it a valuable tool for soil moisture estimation in diverse landscapes.

Future research directions should focus on incorporating additional variables and exploring the temporal dynamics of soil moisture using SAR time series analysis. By expanding the scope of variables considered and examining how soil moisture levels change over time, researchers can further refine soil moisture estimation models. These advancements will contribute to more robust and accurate soil moisture predictions, with profound implications for agricultural sustainability and environmental resilience.

Overall, this study underscores the value of SAR data and machine learning algorithms in improving soil moisture estimation accuracy. By leveraging these tools, researchers and practitioners can make informed decisions regarding land management and climate change adaptation, ultimately fostering sustainable agricultural practices and environmental stewardship.

Keywords: Soil Moisture Estimation, Synthetic Aperture Radar (SAR), Machine Learning, Radar Vegetation Index, Microwave Polarization Difference Index, Decision trees, Google Earth Engine.

2. INTRODUCTION

Soil moisture is an integral quantity in hydrology that represents the average conditions in a finite volume of soil. The moisture content of a soil is an indicator of the degree of saturation of the specimen and is represented by the ratio of the mass of water to the mass of solids in the soil sample.

This estimation is essential in monitoring farming activities, predicting natural disasters, managing water supply. It may be a signal for future flood or water deficit ahead of other indicators.

Soil moisture affects factors like:

- content of air, salinity and number of toxic substances
- ground structure and thickness
- temperature and heat capacity of the ground
- prevents weathering and determines the field's readiness for agricultural processing.

These are the following possibilities that demonstrate the importance of soil moisture measurement.

Soil moisture can affect global climate change and even global warming. In addition, surface-soil-moisture content is an important parameter in micrometeorology and hydrology studies. It has a guiding significance for agricultural production and drought-disaster prevention and reduction.

The methods for determining or predicting soil moisture is divided into two types: the conventional method and the modern method. Generally, the conventional method includes thermo-gravimetric technique and the calcium-carbide technique where field sampling and laboratory determination are needed. In the thermo-gravimetric technique, the soil samples are dried in an oven at 105 °C for 24 hrs. Next their weights are recorded and the finally the soil-moisture content is obtained by calculating the fraction of the oven-dry weight.

The dielectric technique is only suitable for measuring the moisture of single soils. Although conventional methods have many advantages such as high accuracy, short duration, they also have disadvantages like their complex sampling processes, large number of repeated experiments, the limitations of various soil samples for laboratory measuring. The TDR, neutron-scattering-probe method and FDR methods are too expensive and complicated. The heat-flux technique and the resistive method are simple to use, affordable, but they also need soil-specific calibration and have high response times.

Compared with these conventional methods, remote-sensing technology has been one of the most important tools for monitoring and estimating large-scale surface soil moisture. Remote-sensing technology has some advantages and has been applied for soil-moisture estimation since the 70s. Different types remote-sensing sensors are utilized to inverse the soil-moisture content. Currently, the three main types of remote-sensing sensors are used to obtain soil-moisture content are: optical, thermal, and microwave.

2.1 SAR data and its significance in Soil Moisture Estimation:

With the advances in technology, microwave remote sensing has become an alternative to soil-moisture monitoring. Microwave-remote-sensing technology can extract information by identifying the wide disparity in the dielectric permittivity of water, air and solids. Small changes in the soil-moisture content can affect the emissivity and backscattering of microwaves on the soil surface. By analysing the changes in this complex permittivity of the soil, the soil-moisture information can be extracted from microwave remote sensing. There are a variety of frequency microwave data such as P-band, L-band, C-band and X-band data which are commonly used to retrieve soil moisture.

Microwave remote sensing has a wide range of applications, high stability and good adaptability with high accuracy. The passive microwave signal offers several advantages over other methods for remote sensing soil moisture; it can penetrate cloud, it has a direct relationship with soil moisture through the soil dielectric constant, and it has a reduced sensitivity to land surface roughness and vegetation cover. Within the microwave spectrum, lower frequencies respond to a deeper soil layer and are less attenuated by vegetation, and are best suited for soil moisture remote sensing.

Table 1: Various bands of microwave along with their wavelengths

Bands	**Wavelength(cm)_**
P	133.76.9
L	76.9-19.3
S	19.3-7.1
C	7.1-5.2
X	5.2-2.7
K	2.7-0.83
K_u	2.7-1.36
K_a	1.36-0.83
Q	0.83-0.65
V	0.65-0.53
W	0.53-0.30

Here, soil moisture estimation is done using the C-band of Sentinel-1. The analysis of angular behaviour of SAR imagery for estimating soil moisture includes SAR data for a lower incidence angle as compared to high incidence angle.

Machine learning has become a new tool for estimating soil moisture using SAR data where various algorithms are used and compared to review which models predicts soil moisture at better accuracy. There is scope and need to further enhance our studies related to this for better approach, faster result and also to know it’s potential in near future.

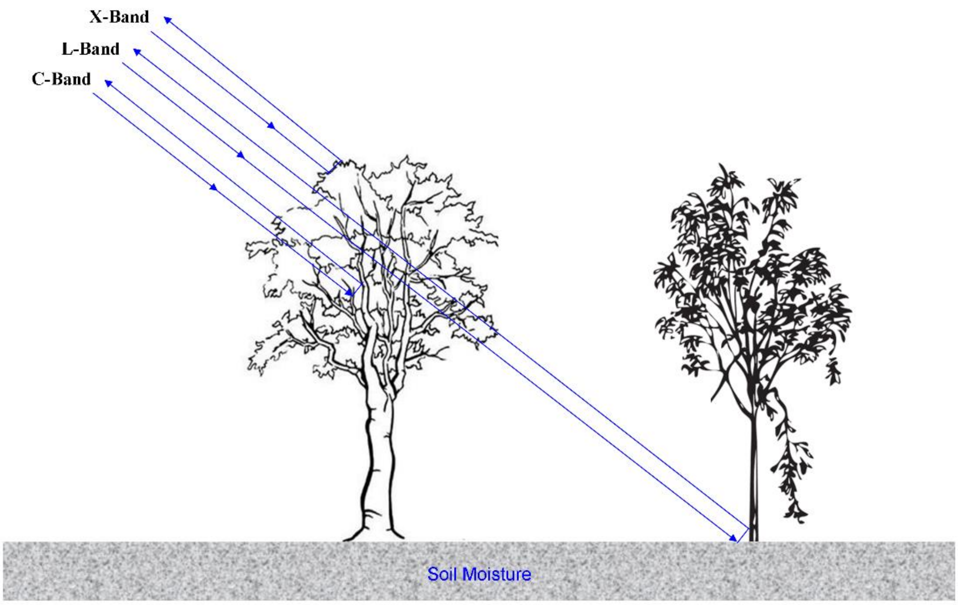

Figure 1: SAR band behaviour for soil moisture estimation

Source: https://gathacognition.com/article/gca1/space-borne-active-microwave-remote-sensing-of-soil-moisture-a-review?show=cited_by

2.2 PROJECT OBJECTIVES:

- Evaluate the effectiveness of microwave and SAR data for soil moisture estimation.
- Develop models and algorithms for integrating microwave and SAR data to improve soil moisture estimation accuracy.
- Assess the feasibility of using SAR data for monitoring soil moisture over large areas.
- Validate the developed models using ground truth data and compare the results with other soil moisture estimation methods.
- To develop a model or method for accurately predicting soil moisture content using mathematical equations in Python.
- To evaluate the performance of different interpolation methods for estimating soil moisture content at unmonitored locations within a study area.

2.3. Aim: To explore the potential applications of the model in:

- Improving agricultural practices.
- Optimizing irrigation scheduling.
- To develop a decision support system for farmers that uses soil moisture data to optimize irrigation scheduling and improve water use efficiency.

3. LITERATURE REVIEW:

1.**Machine Learning Models for Enhanced Estimation of Soil Moisture Using Wideband Radar Sensor**

-Uthayakumar, A., Mohan, M. P., Khoo, E. H., Jimeno, J., Siyal, M. Y., Karim, M. F.

- In this paper, they have mentioned about using ML algorithms and short-range wide band radar sensor for soil moisture estimation.
- It is measured in volumetric units using short-range off-the-shelf radar sensor operating at 3–10 GHz.
- The radar data are post processed to determine the soil moisture which is mapped to the input features extracted from the reflected signals for the training of the machine learning models.
- Various machine learning models, including neural network, support vector machine, linear regression, and k-nearest neighbour, were assessed. Model performance was evaluated using root mean square error, coefficient of determination, and mean absolute error out of which neural network gave better result.
- This research work has been carried out with an intention to develop cost-effective solutions for common users such as agriculturists to monitor the soil moisture conditions with improved accuracy.

2. Soil Moisture Content Estimation Based on Sentinel-1 SAR Imagery Using an Artificial Neural Network and Hydrological Components.

-Chung, J., Lee, Y.-G., Kim, J., Jung, C.,Kim, S.

- This paper gives idea to estimate SMC using Sentinel-1A/B C-band image and ANN over a 40 × 50-km 2 area located in the Geum River basin in South Korea.
- The hydrological components characterized by the antecedent precipitation index (API) and dry days were used as input data as well as SAR (cross-polarization (VH) and co-polarization (VV) backscattering coefficients and local incidence angle), topo-graphic (elevation and slope), and soil (percentage of clay and sand)-related data in the ANN simulations.
- An optimal ANN architecture was constructed in terms of the number of hidden layers, hidden neurons, and activation function. The comparison of the estimated SMC with the observed SMC showed that the Pearson's correlation coefficient (R) and the root mean square error (RMSE) were 0.85 and 4.59%, respectively.

3. ANALYSIS OF SENTINEL-1 DERIVED SOIL MOISTURE MAPS OVER OCCITANIE, SOUTH FRANCE

-Baghdadi, N., Bazzi, H., El Hajj, M., Zribi, M.

- This paper introduces an approach for mapping surface soil utilizing Sentinel-1 (S1) and Sentinel-2 (S2) satellite data.
- It couples S1 and S2 images and the inversion of the Water Cloud Model (WCM) using neural networks. The use of S1 and S2 data offers advantages such as high spatial resolution (10 m x 10 m) and high revisit time (6 days over Europe), making them suitable for operational SSM mapping.

- The proposed approach builds on this by incorporating a priori information on soil moisture conditions based on weather forecasts. This additional information helps constrain the range of possible soil moisture values estimated by the neural networks, leading to improved accuracy.

4. Estimation of Soil Moisture from Optical and Thermal Remote Sensing: A Review

- Zhang, D., Zhou, G.

The authors present a comprehensive review of soil moisture (SM) estimation using optical and thermal remote sensing, focusing on the physical basis and status of estimation methods. They highlight advancements in using temporal information for SM estimation, emphasizing the relationship between SM and surface reflectance or vegetation index for optical methods, and surface temperature for thermal methods. The review compares different approaches, discusses their strengths and weaknesses, and proposes key research directions for future work in SM estimation using remote sensing.

5. Analysis of microwave polarization difference index characteristics about different vegetation types in northeast of China

-Yang-yang Li, Kai Zhao, Xing-Ming Zheng, Jian-hua Ren.

The study analyses the Microwave Polarization Difference Index (**MPDI**) characteristics for different vegetation types in northeast China, focusing on its potential for soil moisture content and biomass inversion. Using monthly AMSR-E microwave brightness temperature data and IGBP classification standard data for 2010, the research examines MPDI characteristics and seasonal variations for various surface types. The results indicate that MPDI is influenced by frequency, with most MPDI values decreasing with increasing frequency. Additionally, MPDI shows seasonal differences for different vegetation types, correlating with vegetation density and canopy characteristics. Snow cover is also identified as an important factor in the region.

MPDI has potential advantages for estimating SM in the areas where surface cover types vary from bare soil to densely vegetated surfaces. It provides information about evaporation of moisture in soil and conditions of vegetation on the ground. The greater the value of MPDI is, the more arid the surface is, and vice versa (Liang, L.; Zhao).

The canopy optical depth is derived using the Microwave Polarization Difference Index (MPDI), the single scattering albedo and the dielectric constant of the soil [Meesters et al., 2005]. The MPDI is defined as:

$$\mathbf{MPDI= (Tb, V-Tb, H)/ (Tb, V+Tb, H)}$$

Using revised equation of MPDI as ($\sigma°$ VV- $\sigma°$ VH)/ ($\sigma°$ VV+ $\sigma°$ VH) was calculated.

The Radar Vegetation Index is an alternative to NDVI. In crop fields it ranges from 0-1. It increases with crop growth and decreases with decrease in plant water content after crop enters the reproductive phase. Nasirzadehdizaji et al.,2019; Gururaj et al.,2019 gave formulation to calculate RVI using Sentinel Dual-pol data (VV-VH) as:

$$\mathbf{RVI= (4^* \sigma° VH)/ (\sigma° VV+ \sigma° VH)}$$

6. SOIL MOISTURE ESTIMATION USING MICROWAVE REMOTE SENSING - A LITERATURE REVIEW

-Kumari S.

The paper reviews the role of soil moisture in environmental studies and discusses microwave remote sensing approaches for its estimation. It highlights the importance of soil moisture in controlling water and energy exchange at the land-atmosphere interface and its applications in various fields. The review compares passive and active microwave sensors, noting their advantages and limitations. Active sensors, such as Synthetic Aperture Radar (SAR), are considered effective but face challenges related to surface roughness, crop cover, and soil texture variation over large areas. The paper emphasizes the need for different approaches to address these challenges and improve soil moisture estimation using SAR data.

7. **Soil moisture estimation underneath crop cover using high incidence angle C-band Sentinel-1 SAR data**

-Hari Shanker Srivastava, Thota Sivasankar, Madhuri Dilip Gavali, Parul Patel.

The study investigates the potential of high incidence angle C-band Sentinel-1 SAR data for soil moisture estimation under crop cover. SAR's sensitivity to dielectric constant and penetration through vegetation makes it suitable for large-scale soil moisture studies. The research incorporates crop effects on radar signals using a water cloud model, considering factors like Leaf Area Index (LAI) and plant water content. A two-layer feedforward neural network is used to develop soil moisture retrieval models, achieving promising results with correlation coefficients and RMSE values. The study concludes that high incidence angle SAR data, with proper consideration of crop effects, can be effective for soil moisture estimation under crop cover, and could potentially be extended to extract crop biophysical parameters from Sentinel-1 SAR data.

4.STUDY AREA OVERVIEW AND DATASET DESCRIPTION:

My study area encompassed regions within Uttarakhand, along with border areas of Himachal Pradesh and portions of Uttar Pradesh

Figure 2:Study area of the project

Co-ordinates: 77.59181170141844, 30.42523830461037

Area: 362112.66 ha.

Topography: It includes a mix of flat to gently sloping terrain, typical of agricultural land and orchards, with some areas near water bodies or canals possibly having more varied or undulating terrain. In contrast, locations like Badaban, Himachal Pradesh, and Tea Garden, Prem Nagar, have hilly or mountainous terrain, characteristic of the region.

Crops grown:The crops cultivated in the areas can vary based on local agricultural practices and climate conditions. However, some common crops grown there were wheat, rajma, berseem, sugarcane, mango orchards etc.

The ground truth data was collected when the crops were at a medium height, with less bushiness, indicating that they had recently been cultivated.

Rivers: The area is close to the Ganges River (also known as the Ganga), which is one of the major rivers in India. Other smaller rivers or tributaries of the Ganges could also be nearby.

4.1. Datasets used:

1. SENTINEL-1: Sentinel-1 is a satellite mission developed by the European Space Agency (ESA) as part of the Copernicus program. It is a constellation of two satellites, Sentinel-1A and Sentinel-1B, launched in 2014 and 2016, respectively.

Specifications:

- **Satellite:** Sentinel-1
- **Spatial resolution:** 5m
- **Swath:** 400 m
- **Incidence angle:** 30.89-46.04 degrees
- **Frequency:** 5.405 Hz (C-band)
- **Coverage**:
    1. Global coverage
    2. Revisit time of 6 days in IW mode and 12 days in EW mode
- **Transmit Polarization(T_x):** Vertical(VV)
- **Receive Polarization(R_x):** Vertical(VV)
- **Acquisition Mode:** Interferometric wide (5 m x 20 m)
- **Processing level:** Single look complex
- **Acquisition dates:** 31 January 2024 -9 February 2024

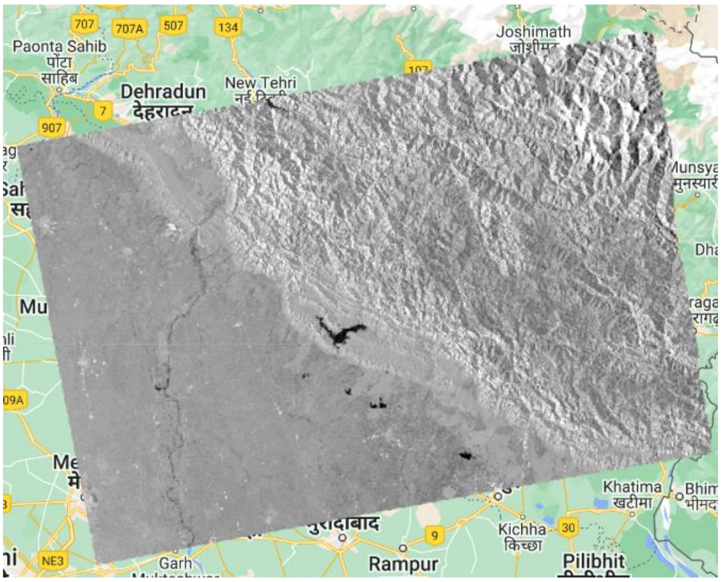

Figure 3: Sentinel-1 image for five dates of the region of interest

5. METHODOLOGY:

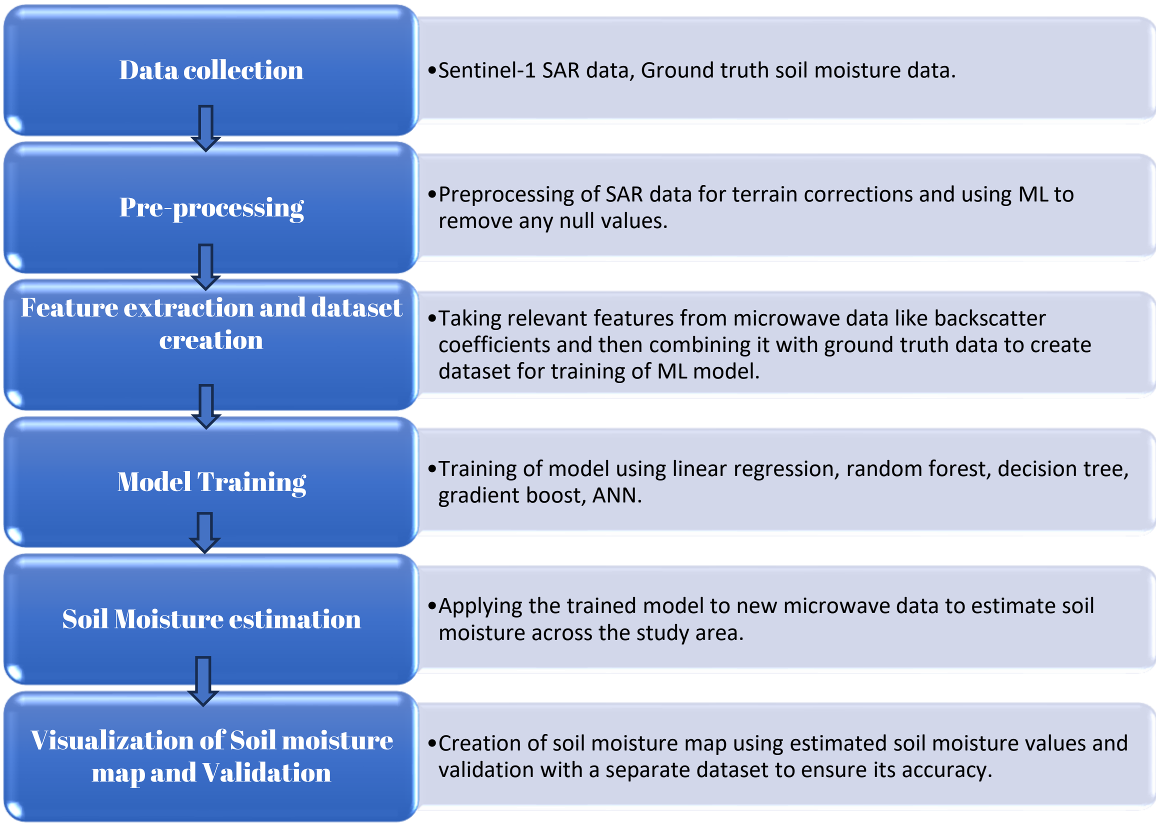

Figure 4: Workflow of the project

5.1. Ground truth collection and Study area analysis:

Soil samples were collected to ensure uniformity in soil moisture, crop cover, and surface roughness across the sampling areas. Moisture levels were measured using a Soil Moisture Theta Probe a type of soil moisture sensor used to measure the volumetric water content in soil. It works on the principle of capacitance measurement, where the dielectric constant of the soil is related to its water content. The probe consists of a rod that is inserted into the soil, and it measures the dielectric constant of the soil around it. This measurement is then converted into a volumetric water content reading, which indicates the percentage of the soil volume that is occupied by water.

Additionally, laboratory analysis was conducted to determine the gravimetric soil moisture content. Gravimetric soil moisture reading is taken by collecting soil samples from various points in the study area. The collected soil samples are weighed to determine their initial wet weight. Subsequently, the samples are dried in an oven at a specific temperature (usually around 105°C) to remove all moisture. After drying, the samples are weighed again to determine their dry weight. The gravimetric soil moisture content is then calculated as the difference between the initial wet weight and the final dry weight, expressed as a percentage of the dry weight. This method provides a direct measurement of the water content in the soil by mass, allowing for accurate assessment of soil moisture levels.

The study area included a mix of flat to gently sloping terrain, typical of agricultural land and orchards, with some areas near water bodies or canals possibly having more varied or undulating terrain. In contrast, locations like Badaban, Himachal Pradesh, and Tea Garden, Prem Nagar, have hilly or mountainous terrain, characteristic of the region. It included crops like wheat, sugarcane, berseem, rajma, mango orchards, forest areas, ploughed fields and bridges etc.

Figure 5: Soil moisture theta probe

5.2. Data collection:
Dual polarized Sentinel-1 data was used for data acquisition.

To cover my entire study area data was collected from 31st January to 9th February which included these following dates:

- 2024-01-31 12:47:39

- 2024-01-31 12:48:04
- 2024-02-05 12:55:49
- 2024-02-05 12:56:14
- 2024-02-08 00:44:01

5.3.Data Preprocessing:

The SAR data was preprocessed to remove any geometric and radiometric errors using Google Earth Engine like terrain corrections, speckle filtering then the image was clipped using roi in GEE to get my area of interest. The VV and VH values of groundtruth areas was extracted using Google Earth Engine.

5.4.Radar Vegeation Index:

The Radar Vegetation Index (RVI) is a metric used in radar remote sensing to characterize vegetation cover. It is calculated from the backscatter coefficients obtained from SAR (Synthetic Aperture Radar) images. The RVI is computed as the ratio of the radar backscatter values in two different polarization modes, typically VH (Vertical-Horizontal) and VV (Vertical-Vertical) or HV (Horizontal-Vertical) and HH (Horizontal-Horizontal). The index is particularly useful for monitoring changes in vegetation structure, biomass, and health, as different vegetation types and conditions exhibit distinct RVI values. RVI is often employed in conjunction with other SAR-derived indices and datasets for various applications in agriculture, forestry, and ecosystem monitoring.

The RVI was calculated using Sentinel Hub EO Browser evalscript where VV,VH values were fed along with my data mask to give quantitative measure of scattering randomness as RVI4S1.

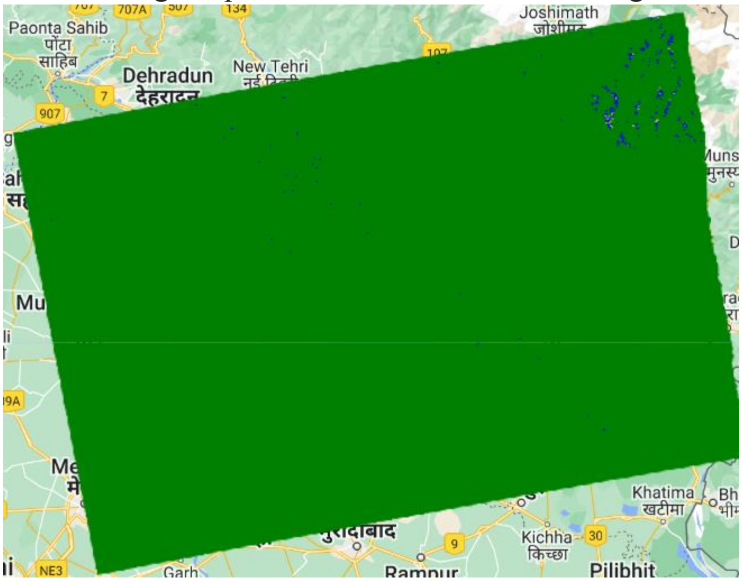

Figure 6: RVI map of the area

5.5.Microwave Polarization Difference Index:

The Microwave Polarization Difference Index (MPDI) is a vegetation index used in remote sensing, particularly in radar imagery analysis. It is calculated using the difference between the HH (Horizontal-Horizontal) and VV (Vertical-Vertical) polarized backscatter coefficients from radar data.

The index is formulated as follows:

$$MPDI= (\sigma° VV- \sigma° VH)/ (\sigma° VV+ \sigma° VH)$$

The HH polarization is more sensitive to the structure of the canopy, while the VV polarization is more sensitive to the volume of the canopy. By subtracting VV from HH, the MPDI emphasizes structural differences within vegetation. Areas with dense and complex vegetation structures tend to have higher positive MPDI values, while areas with less dense or less structured vegetation tend to have lower or negative MPDI values.

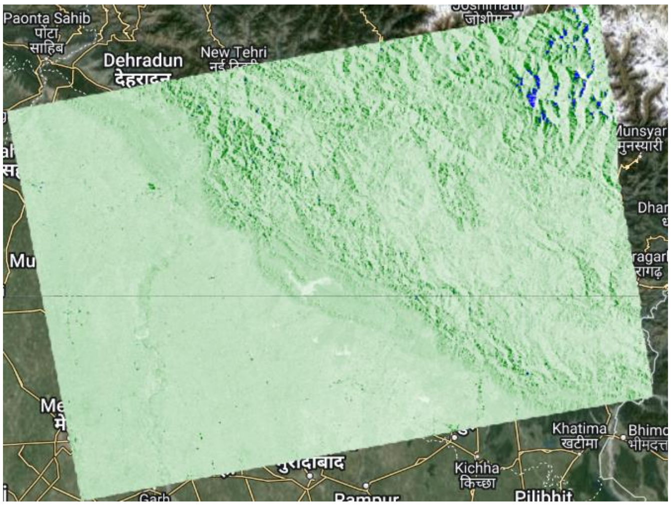

Figure 7: MPDI map of the area

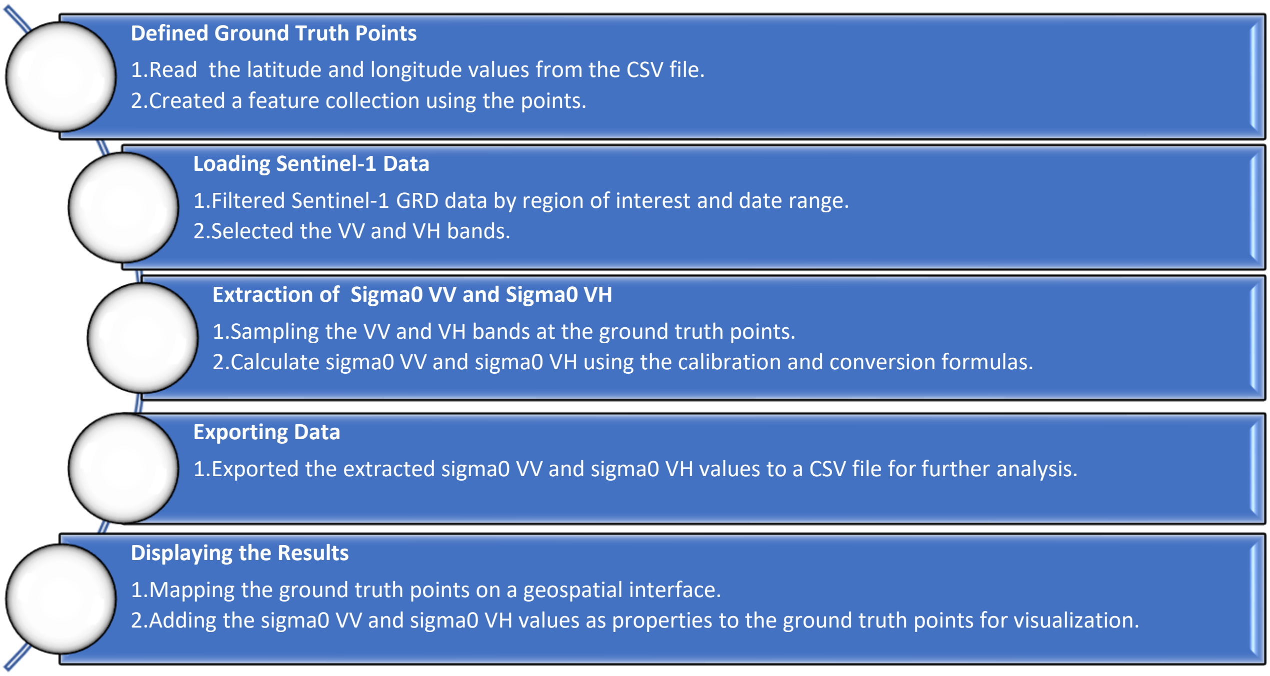

Figure 8: Methodology flowchart of SAR data processing

6. MODELS USED FOR ESTIMATION:

1. **Decision trees:** Decision trees are a popular machine learning technique used for both classification and regression tasks. In the context of this project, it has been used for training model keeping soil moisture as target variable to estimate soil moisture for a new dataset. Decision trees are a type of model that uses a tree-like graph of decisions and their possible consequences. Each internal node represents a "test" or "decision" on an input feature, and each branch represents the outcome of the test. The leaf nodes represent the final decision or prediction.
   In a regression decision tree, for soil moisture prediction, each leaf node represents a predicted numerical value. The decision tree algorithm learns the optimal decision rules from the training data, splitting the data into subsets based on the input features to minimize the variance of the target variable (soil moisture) within each subset.
   A single decision tree was used for training because it provides an interpretable model that can be easily understood and visualized.
   ***Estimation of soil moisture using decision tree regression:(Pekel, E.)**
   The paper applies decision tree regression to estimate soil moisture (SM) using parameters such as air temperature, relative humidity, and soil temperature. The method efficiently generates a decision tree from instances, offering high accuracy (R2), low error (MSE, MAE), with a tree depth of five being optimal. Overall, decision tree regression proves effective in handling SM estimation with satisfying fitness criteria.

2. **Random forest:** Random Forest is utilized as a powerful ensemble learning technique, combining the predictions of multiple decision trees trained on SAR and moisture data. By aggregating the predictions of 100 trees, Random Forest enhances prediction accuracy and robustness, particularly beneficial for complex, non-linear relationships in soil moisture estimation. The model's ensemble approach helps mitigate overfitting and captures the intricate interactions between SAR data and soil moisture, resulting in more reliable and accurate predictions.
   ***Decomposition-Based Soil Moisture Estimation Using UAVSAR Fully Polarimetric Images (Heather N., HasanlouM., Hosseini M., Akhavan Z.):** This study explores the use of polarimetric decomposition methods to estimate soil moisture over agricultural fields, focusing on soybean, wheat, and corn. Random Forest and neural network regression algorithms are employed, with feature selection playing a crucial role in enhancing model performance. The research highlights the significance of utilizing polarimetric UAVSAR data and advanced machine learning techniques for accurate soil moisture estimation, showcasing a correlation of determination (R2) of 0.86 for soybeans.
3. **Gradient boosting:** Gradient Boosting is a machine learning technique used for both regression and classification tasks. It builds a predictive model in the form of an ensemble of weak prediction models, typically decision trees. Gradient Boosting trains the model in a stage-wise manner, where each new tree helps correct errors made by the previous set of trees. This iterative process allows for the creation of a strong predictive model by combining the output of multiple weak models.
   ***Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau (Wei Z., Meng Y., Zhang W., Peng Z.,**

Meng L.): This paper presents a new downscaling approach, the Downscaling based on gradient boosting decision tree (DENSE) method to replace high-resolution soil moisture products lost due to the SMAP satellite's radar failure. It uses 26 soil moisture indices derived from MODIS and a digital elevation model to downscale SMAP observations at a coarse scale to a fine scale over the Tibetan Plateau. The method effectively captures spatio-temporal soil moisture variability and improves spatial resolution from 36 km to 1 km, except in areas with dense vegetation where further improvements are needed.

4. **Artificial Neural Network:** Artificial Neural Networks (ANNs) are a class of machine learning models inspired by the structure and function of the human brain. They consist of interconnected nodes, called neurons, organized in layers.
   In a typical ANN, there are three types of layers: input layer, hidden layers, and output layer. The input layer receives the initial data, which is then passed through one or more hidden layers where the neurons perform mathematical transformations on the input. Each neuron in a layer takes input from the neurons in the previous layer, applies a weighted sum of these inputs along with a bias term, and then passes the result through an activation function to introduce non-linearity.
   The final output is generated by the neurons in the output layer, which typically represent the predictions or classifications made by the model. During training, the network adjusts the weights and biases of the neurons based on the difference between the predicted output and the actual target values, using an optimization algorithm like gradient descent.
   In this project, Artificial Neural Networks (ANNs) are utilized to predict soil moisture levels using SAR (Synthetic Aperture Radar) data and ground truth observations. The ANN model architecture comprises two hidden layers, with the first layer containing 100 neurons and the second layer containing 50 neurons. SAR data, such as backscatter coefficients, are fed into the network as input features, while ground truth soil moisture values serve as the target output. Through training on a dataset split into training and validation sets, the ANN learns to map the complex relationships between SAR data and soil moisture through iteration of about 200, enabling it to predict soil moisture levels accurately. It uses reLu activation function and adam solver optimization algorithm to give better predictions.
   ***Combined Use of Sentinel-1 SAR and Landsat Sensors Products for Residual Soil Moisture Retrieval over Agricultural Fields in the Upper Blue Nile Basin, Ethiopia (Gessesse B., Melesse A., Yigrem Y.):** This study explores the use of Sentinel-1 SAR data and soil roughness parameters to estimate residual soil moisture in the Upper Blue Nile basin, Ethiopia. The research integrates Landsat data and the Water Cloud Model (WCM) to estimate crop residue water content. An Artificial Neural Network (ANN) is employed to translate SAR backscattering and surface roughness data into soil moisture values. Results show that the ANN model performed well, particularly when incorporating SAR VV data alongside surface roughness parameters, demonstrating the potential of SAR data for accurate soil moisture estimation in agricultural areas with crop residues.

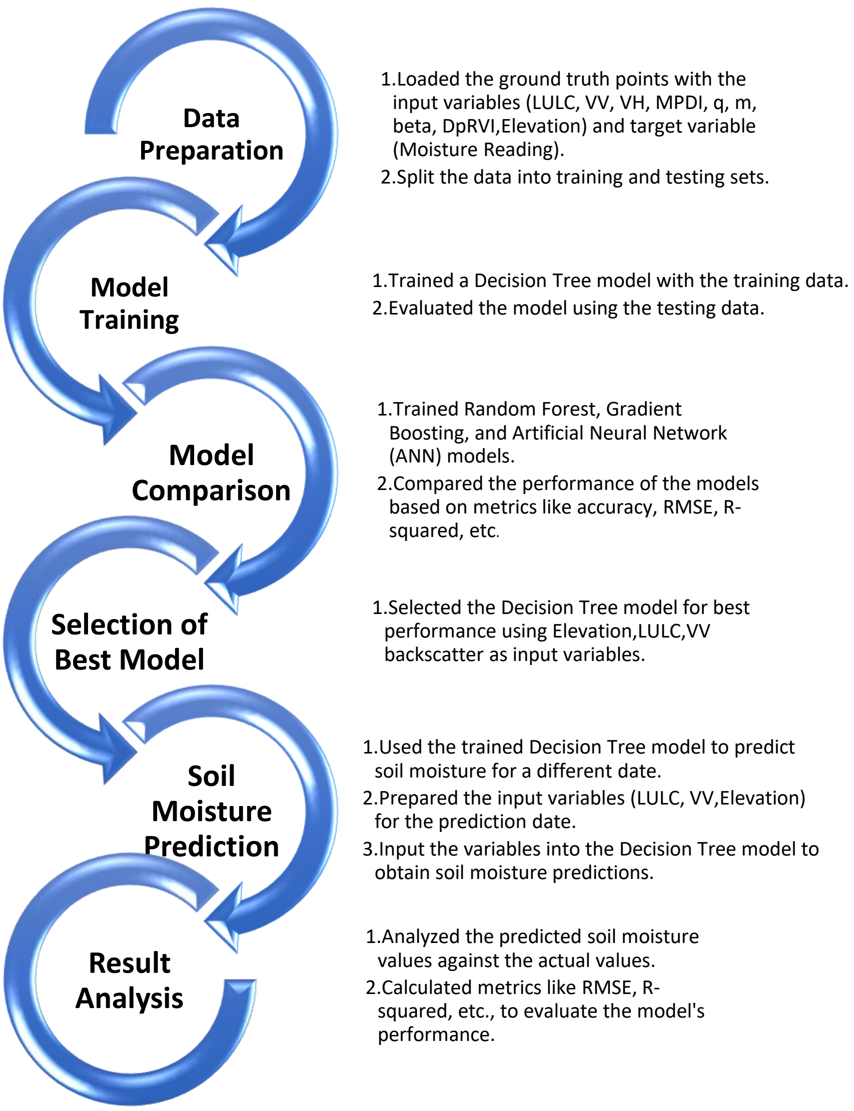

Figure 9: Model Training and Soil Moisture Prediction

7. Visualization of Soil moisture map and Validation:

After predicting soil moisture values for different dates, the data was exported to a CSV file. Using QGIS, a point shapefile layer was created to match the moisture readings with their respective latitudes and longitudes, representing the locations where ground truth data was collected. Additionally, a polygon shapefile outlining the study area was extracted using Google Earth Engine. Both the point and polygon shapefiles were then imported into ArcGIS Pro. The IDW (Inverse Distance Weighted) interpolation tool from the 3D Analyst toolset was utilized to generate the predicted soil moisture map.

Validation of the generated map was done with the Sentinel map using rasterio library in python.

Inverse Weighted Distance: Inverse Distance Weighted (IDW) interpolation is a method used in GIS and spatial analysis to estimate values at unknown locations based on values from known locations. It assumes that the value at an unmeasured location is more similar to nearby measured values than to those further away. The IDW method calculates the interpolated value using a weighted average of the measured values, where the weights are inversely proportional to the distance between the known and unknown points. In essence, closer points have a greater influence on the interpolated value than those farther away. The IDW method is commonly used for spatial interpolation in various fields, including geology, hydrology, environmental science, and geography.

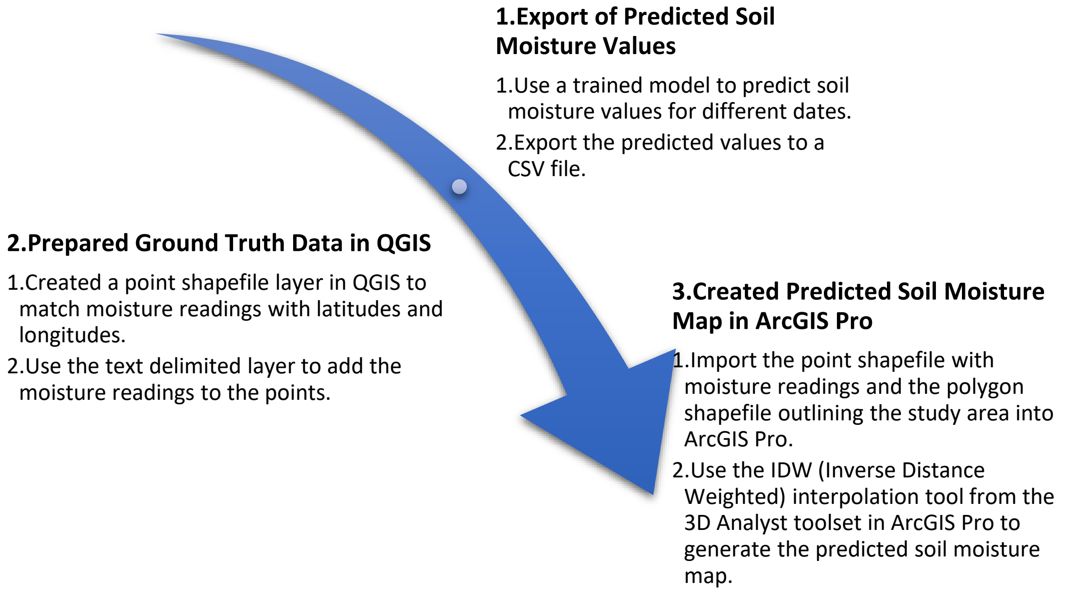

Figure 10: Flowchart of Soil Moisture Mapping Using IDW Interpolation

8. RESULTS AND DISCUSSION:

8.1. Co-relation between the variables using Scatter plot

1. **Scatter plot between VV backscatter and Soil Moisture:**
   A 2nd order polynomial trendline fitted the data to give R^2 value of 0.8173.

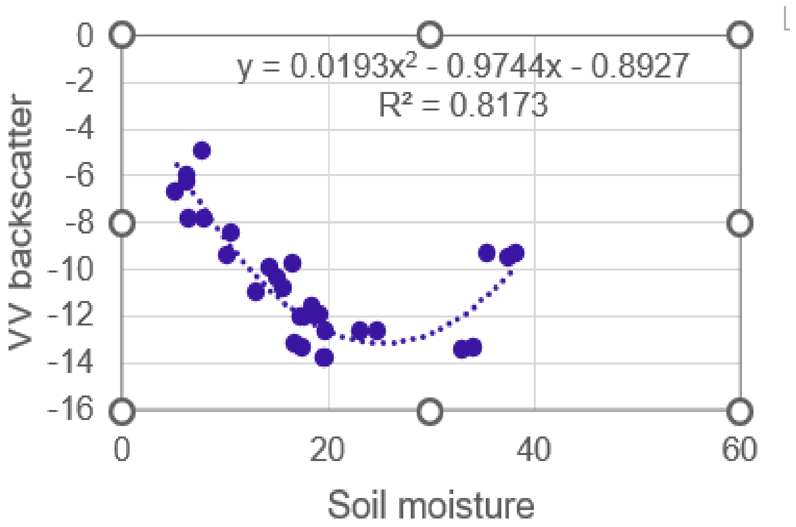

Figure 11: Relation graph between VV v/s Soil Moisture

2. Scatter plot between VH backscatter and Soil Moisture:

A 2nd polynomial trendline fitted the data to give R^2 value of 0.8014

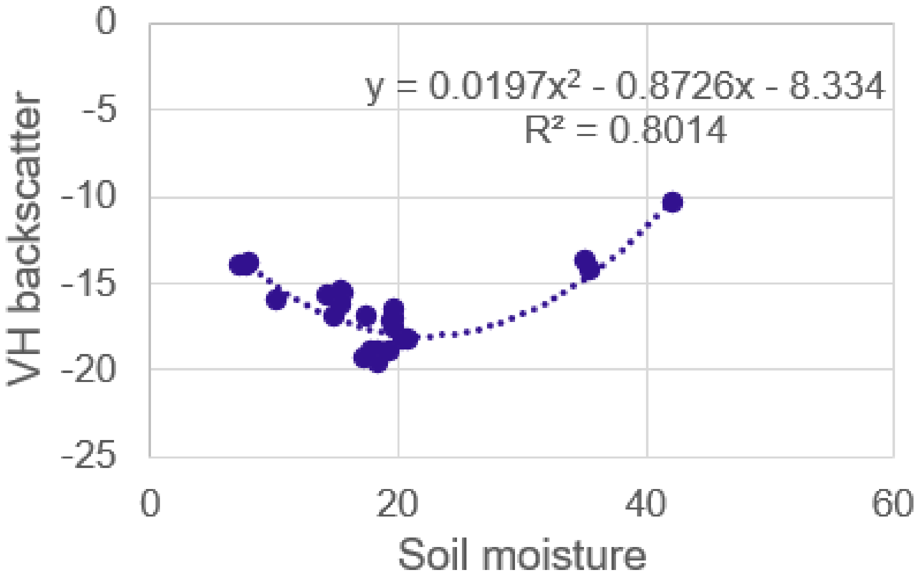

Figure 12: Relation graph between VH v/s Soil Moisture

3. Scatter plot between MPDI and Soil Moisture:

A 2nd polynomial trendline fitted the data to give R^2 value of 0.762.

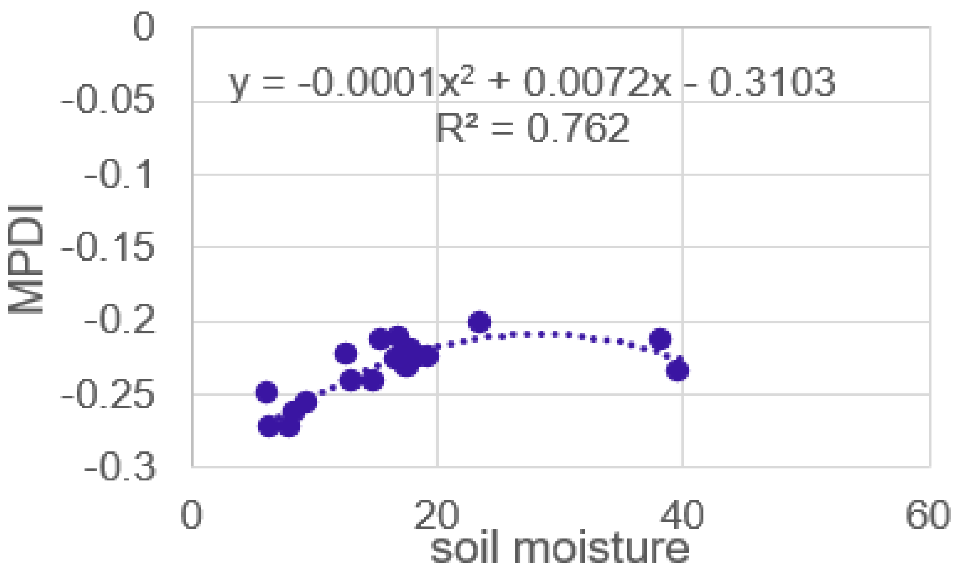

Figure 13. Relation graph between VV v/s Soil Moisture

4. **Scatter plot between VH backscatter and DpRVI:**
   A 2nd polynomial trendline fitted the data to give R^2 value of 0.7072.

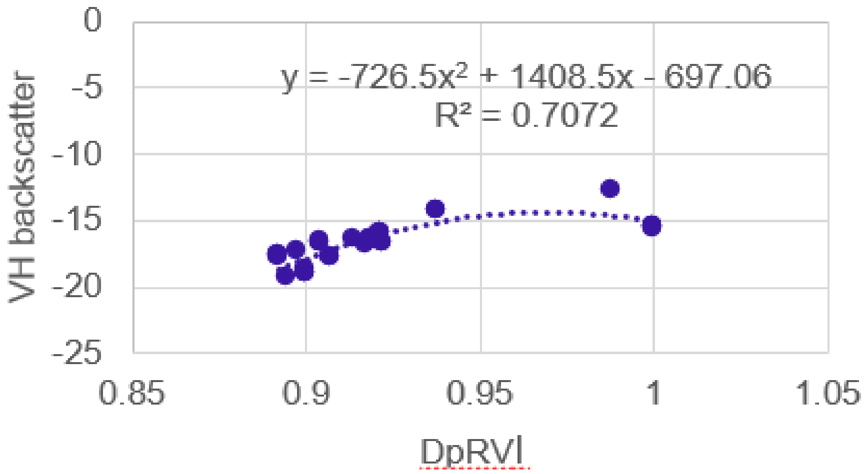

Figure 14. Relation graph between VH v/s RVI Index

5. **Scatter plot between m (co-pol purity index) and Soil Moisture:**
   A 2nd order polynomial trendline fitted the data to give R^2 value of 0.8389.

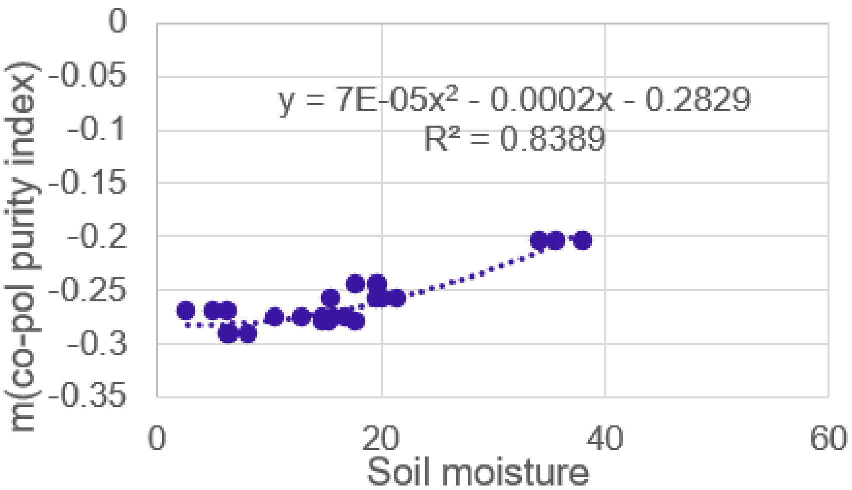

Figure 15.Relation graph between m v/s Soil Moisture

From the analyses conducted, it is evident that VV backscatter and Co-pol purity index exhibit a stronger relationship with soil moisture compared to other variables. Among these, the relationship between VV backscatter and soil moisture is particularly noteworthy. This relationship is significant because it allows for the development of a predictive model that can estimate soil moisture for different dates. By inversing the equation derived from the VV backscatter and soil moisture relationship, it becomes possible to predict soil moisture values for dates beyond the ones covered by the dataset. This finding underscores the importance of VV backscatter as a key factor influencing soil moisture and highlights its potential for practical applications in soil moisture prediction.

Comparison of RVI Across Landcovers:

The graph below displays the Radar Vegetation Index (RVI) values, which are used to differentiate between vegetation and bare soil. A value of 0 on the RVI scale indicates bare soil, while a value of 1 represents dense vegetation. The RVI serves as an indicator of the vegetation state, with values between 0 and 1 indicating varying degrees of vegetation cover. This information is crucial for assessing the vegetative health of an area and monitoring changes over time.

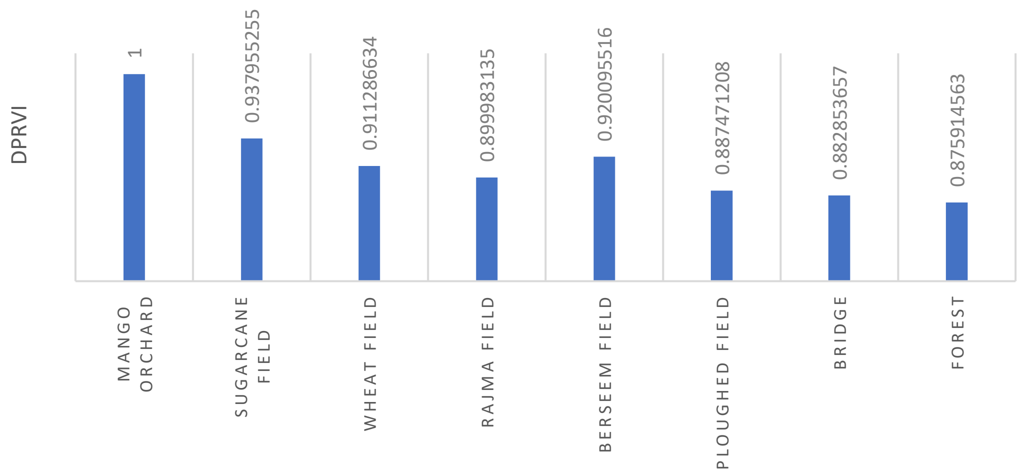

Figure 16. DpRVI comparison across different landcovers

8.2. ANALYSIS OF BACKSCATTER CO-EFFICIENTS USING MACHINE LEARNING ALGORITHMS

Machine learning algorithms such as decision trees, random forests, gradient boosting, and artificial neural networks (ANNs) are being utilized. These algorithms are applied to predict soil moisture levels based on various input variables such as land use/land cover, backscatter coefficients from SAR images, and other relevant factors. Decision trees are used for their simplicity and interpretability, while random forests leverage the power of ensemble learning for improved accuracy. Gradient boosting further enhances predictive performance by combining multiple weak learners. ANNs, on the other hand, offer the ability to capture complex relationships in the data. Overall, these machine learning techniques are instrumental in deriving insights and predictions for soil moisture estimation, leveraging the power of data-driven models.

Figures 17. Backscatter v/s Moisture reading plots of different models

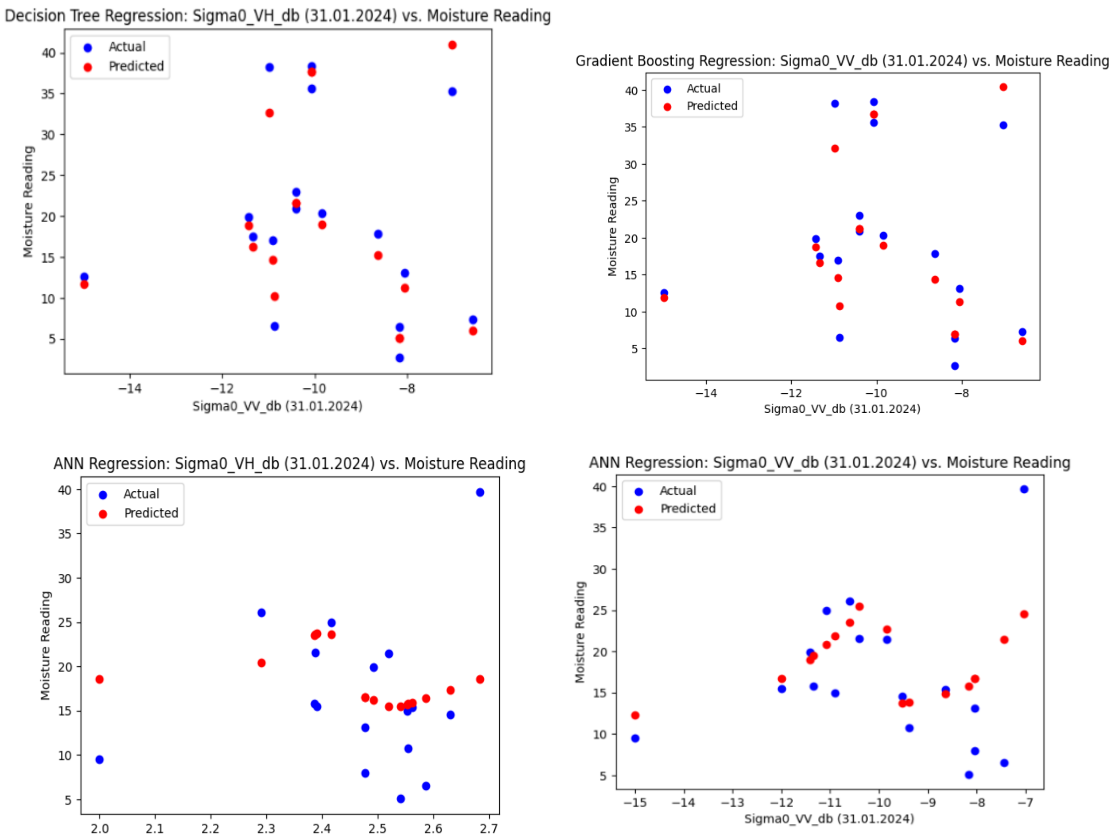

Table 2: Backscatter(VH) v/s Moisture reading plots validation and analysis result

Models	MAE(Mean squared error)	R^2
Random forest	31.12	0.75
Decision tree	6.68	0.95
Gradient Boost	8.18	0.9338636213025683
Artificial Neural Network	62.96	0.09

Table 3: Backscatter(VV) v/s Moisture reading plots validation and analysis result

Models	MAE(Mean squared error)	R^2
Random forest	13.08	0.89
Decision tree	6.69	0.95
Gradient Boost	7.89	0.94
Artificial Neural Network	45.57	0.34

Based on the analysis, it is evident that the decision tree algorithm exhibits a low Mean Absolute Error (MAE) and a high R-squared value, indicating a strong fit for the data. These results suggest that the decision tree model is effective for both training the model and making predictions.

8.3.MODEL TRAINING AND VALIDATION RESULT OF DIFFERENT MODELS:

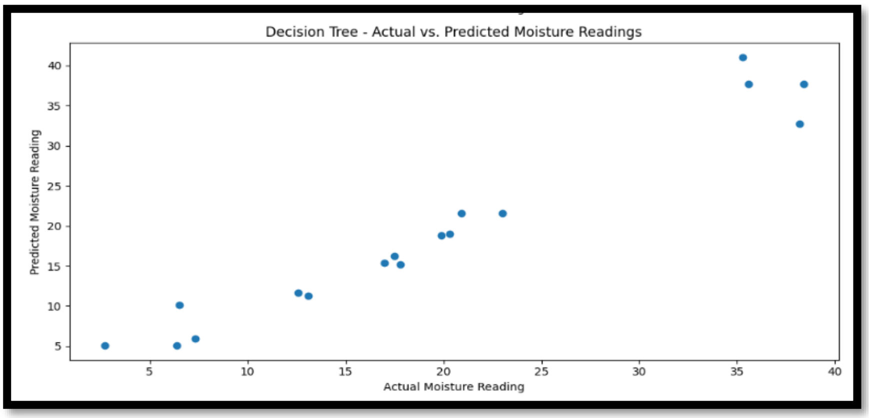

Figure 18. Validation plot of decision tree

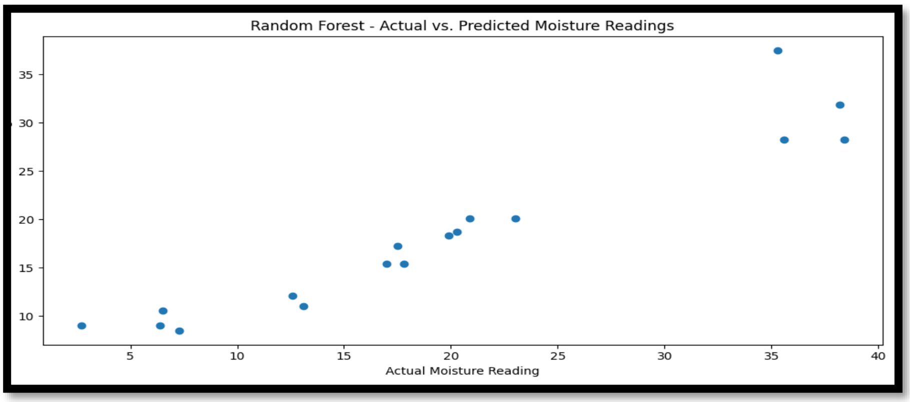

Figure 19. Validation plot of random forest

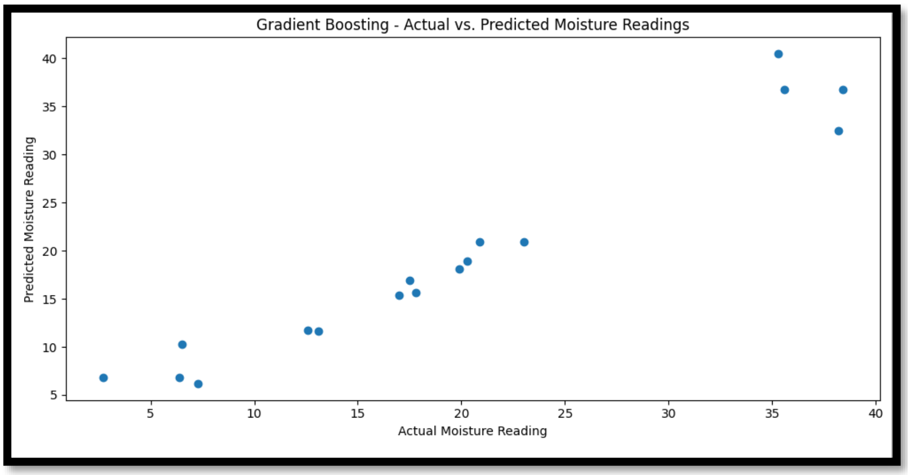

Figure 20. Validation plot of Gradient boost regression

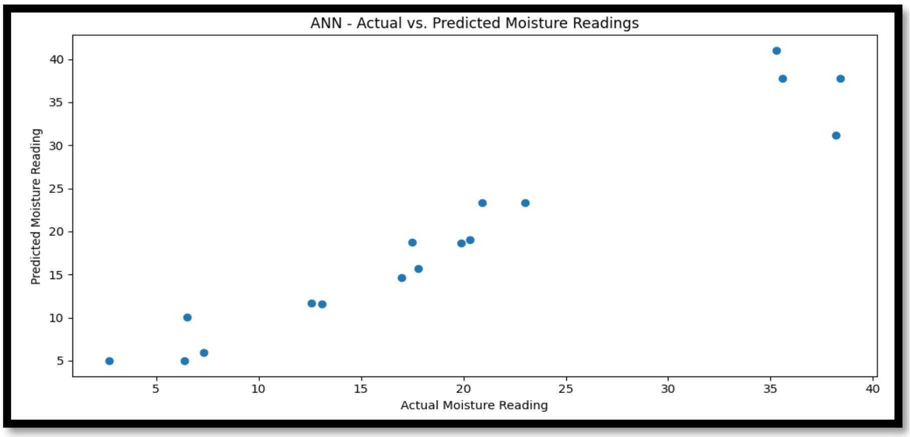

Figure 21. Validation plot of ANN

4.Table: Model Validation results

Models	MAE(Mean squared error)	R^2	RMSE
Random forest	20.25	0.77	4.50
Decision tree	9.62	0.89	3.10
Gradient Boost	18.85	0.78	4.34
Artificial Neural Network	10.04	0.78	3.17

Based on the validation analysis, it is evident that the decision tree algorithm exhibits a low Root Mean Square Error (RMSE) and a high R-squared value, indicating a strong fit for the data.

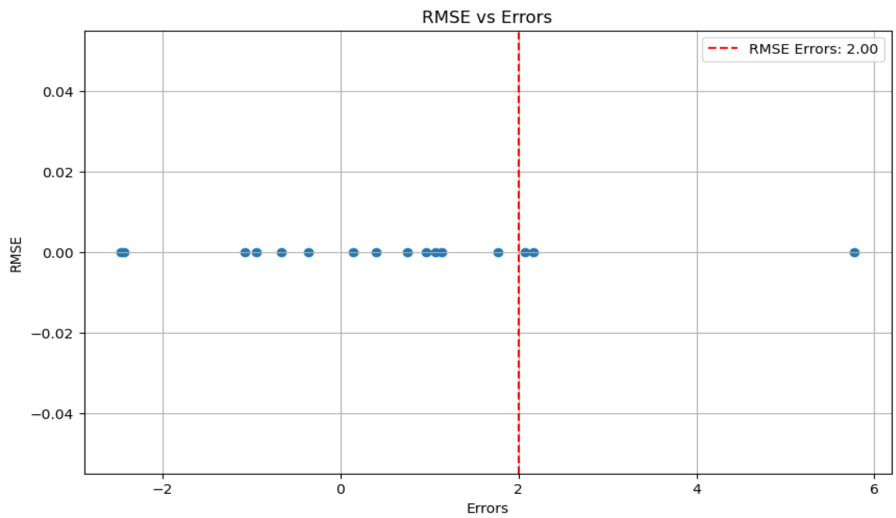

Figure 22: RMSE v/s Error graph of decision tree

9. GENERATED SOIL MOISTURE MAP BY PREDICTIVE MODELLING

1. **Soil moisture map for 31st January 2024, the date for which ground data is available**

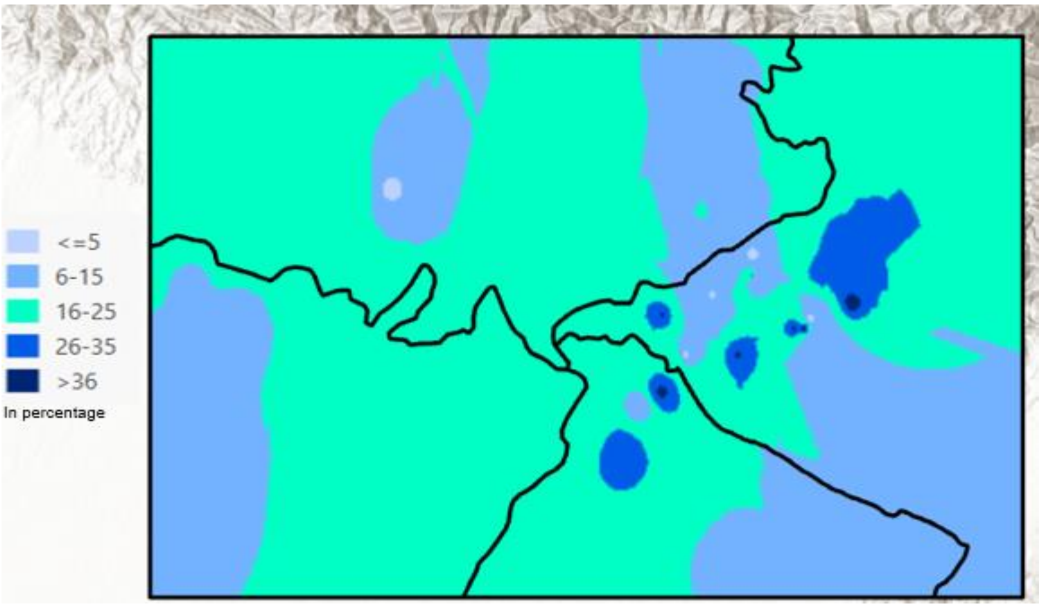

Figure 23. Soil moisture map of 31.1.2024

2. **Predicted Soil moisture map for 19th January 2024**

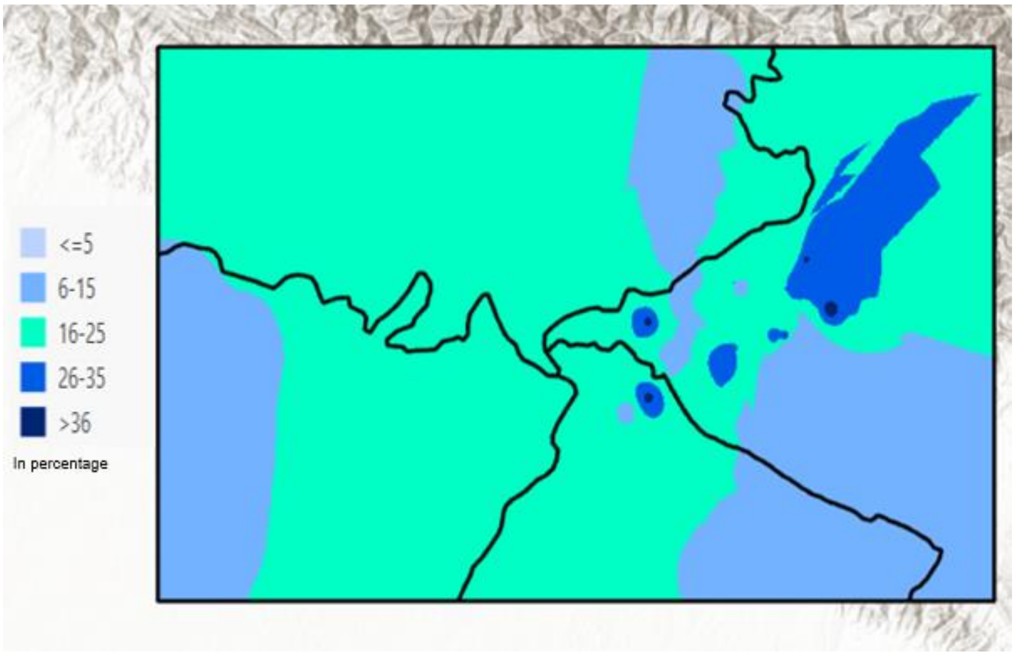

Figure 24: Soil moisture map of 19.1.2024

10. CONCLUSION:

In conclusion, this study has demonstrated the effectiveness of using SAR data and machine learning techniques for soil moisture estimation in the study area encompassing regions within Uttarakhand, Himachal Pradesh, and Uttar Pradesh. The analysis revealed a strong relationship between VV backscatter and soil moisture, highlighting the significance of SAR data in predicting soil moisture levels. The decision tree algorithm emerged as a robust model for training and prediction, exhibiting low Mean Absolute Error (MAE) and high R-squared values.

The findings of this research have practical implications for agriculture, water resource management, and environmental monitoring in the study area. Accurate soil moisture estimation can help optimize irrigation scheduling, improve crop yield predictions, and facilitate sustainable land management practices. Additionally, the study provides valuable insights for future research directions, such as integrating additional variables and exploring temporal dynamics of soil moisture using SAR time series analysis.

While this study has contributed valuable insights, it is important to acknowledge its limitations. Future research could address these limitations and further enhance our understanding of soil moisture dynamics in agricultural and environmental contexts.

Overall, this research underscores the importance of SAR data and machine learning techniques in advancing soil moisture estimation capabilities, and it lays the foundation for further research in this field. By improving our ability to predict and monitor soil moisture levels, we can contribute to more sustainable agricultural practices and better water resource management strategies in the study area and beyond.

11. REFERENCES:

- Uthayakumar, A., Mohan, M. P., Khoo, E. H., Jimeno, J., Siyal, M. Y., & Karim, M. F. (2022). Machine Learning Models for Enhanced Estimation of Soil Moisture Using Wideband Radar Sensor. *Sensors (Basel)*, 22(15), 5810. https://doi.org/10.3390/s22155810
- Chung, J., Lee, Y., Kim, J., Jung, C., & Kim, S. (2022). Soil Moisture Content Estimation Based on Sentinel-1 SAR Imagery Using an Artificial Neural Network and Hydrological Components. *Remote Sensing*, 14(3), 465. https://doi.org/10.3390/rs14030465
- Baghdadi, N., Bazzi, H., El Hajj, M., & Zribi, M. (2019). Analysis of Sentinel-1 Derived Soil Moisture Maps Over Occitanie, South France. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium* (pp. 1-4). https://doi.org/10.1109/IGARSS.2019.8899078
- Zhang, D., & Zhou, G. (2016). Estimation of Soil Moisture from Optical and Thermal Remote Sensing: A Review. *Sensors (Based)*, 16(8), 1308. https://doi.org/10.3390/s16081308
- Li, X., Zhao, K., Zheng, X., & Ren, J. (2013). Analysis of Microwave Polarization Difference Index Characteristics about Different Vegetation Types in Northeast China. *Proceedings of the 2013 International Conference on Remote Sensing, Environment and Transportation Engineering*, 9. https://doi.org/10.2991/rsete.2013.9
- Kumari, S. (2021). Soil Moisture Estimation Using Microwave Remote Sensing - A Literature Review. *SGVU Journal of Climate Change and Water*, 8, 55-72. ISSN: 2347-7741.
- Srivastava, H. S., Sivasankar, T., Gavali, M. D., & Patel, P. (2023). Soil moisture estimation underneath crop cover using high incidence angle C-band Sentinel-1 SAR data. *Kuwait Journal of Science*, 51(2). https://doi.org/10.1016/j.kjs.2023.07.007
- Pekel, E. (2020). Estimation of soil moisture using decision tree regression. *Theoretical and Applied Climatology*, 139(1). https://doi.org/10.1007/s00704-019-03048-8.
- Akhavan, Z., Hasanlou, M., Hosseini, M., & McNairn, H. (2021). Decomposition-Based Soil Moisture Estimation Using UAVSAR Fully Polarimetric Images. *Agronomy*, 11(1), 145. https://doi.org/10.3390/agronomy11010145.
- Wei, Z., Meng, Y., Zhang, W., Peng, J., & Meng, L. (2019). Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau. *Remote Sensing of Environment*, 228, 111-123. https://doi.org/10.1016/j.rse.2019.02.022.
- Ayehu, G., Tadesse, T., Gessesse, B., Yigrem, Y., & Melesse, A. M. (2020). Combined Use of Sentinel-1 SAR and Landsat Sensors Products for Residual Soil Moisture Retrieval over Agricultural Fields in the Upper Blue Nile Basin, Ethiopia. *Sensors*, 20(11), 3282. https://doi.org/10.3390/s20113282