PROJECT REPORT

ON

BIOPHYSICAL PARAMETER ESTIMATION USING MICROWAVE DATA AND

MACHINE LEARNING

****

**COURSE - M.Sc. Agriculture Analytics**

**SUBMITTED TO: SUBMITTED BY:**

**Dr. Dipanwita Haldar** **Akriti Dhyani**

(Scientist/Engineer - SF) ID-TS90723

Dept. of Agriculture and Soils

**Manisha Shetty**

**Mr. Prasun Kumar Gupta** ID-TS90728

(Scientist/Engineer - SE)

Dept. of Geo Informatics

**ACKNOWLEDGEMENT**

We would like to express our sincere gratitude to Dr. Dipanwita Haldar, for her invaluable guidance, mentorship and unwavering support throughout the journey. Her expertise, insights, practical advice, and encouragement has helped us tremendously. This research is successful only because of her constant support and guidance. We are deeply indebted to her for always motivating us whenever we came across challenges.

We are grateful to our guide, Mr. Prasun Kumar Gupta, for helping us deal with all the technical issues. His patience and insightful comments were instrumental in shaping the direction and scope of our research.

We would also like to thank our course director, Dr. N. R. Patel, for giving us the opportunity to work on this wonderful project. His vision and leadership have been integral to the success of this project. We are grateful for his support and encouragement throughout this journey.

Additionally, we extend our heartfelt appreciation to our seniors, Mr. Soumya Ranjan Sahoo and Ms. Steena for sharing their knowledge gained by working on this project before us. Their inputs in this project were of great help. We want to thank our friends and family for constantly supporting and encouraging us.

Finally, we would like to acknowledge the support of the Indian Institute of Remote Sensing, for providing resources and infrastructure necessary for the completion of this project.

Table of Contents

|  |  |  |
| --- | --- | --- |
| Sr. No. | Title | Page No. |
| 1. | Abstract | 1 |
| 2. | Introduction | 2 |
| 2.1. | Gap in Research | 2 |
| 2.2. | Objectives of Project | 2 |
| 3. | Optical Vegetation Indices | 3 |
| 3.1. | NDVI | 3 |
| 3.2. | NDRE | 4 |
| 4. | SAR | 4 |
| 4.1. | SAR Polarisation | 4 |
| 4.2. | SAR Interaction with Vegetation | 5 |
| 5. | Literature Review | 5 |
| 6. | Study Area | 7 |
| 7. | Dataset | 8 |
| 7.1. | Sentinel-1 | 8 |
| 7.2. | Sentinel-2 | 9 |
| 8. | Methodology | 10 |
| 8.1. | Remote Sensing Variables and Biophysical Parameter Correlation | 13 |
| 8.2. | ML Models Used for the Study | 16 |
| 8.2.1. | Random Forest Regressor | 16 |
| 8.2.2. | Extreme Gradient Boost | 16 |
| 8.2.3. | Artificial Neural Network | 16 |
| 9. | Results and Discussion | 18 |
| 9.1. | ML Models | 18 |
| 9.2. | Regression Analysis | 21 |
| 9.3. | Final Maps | 23 |
| 10. | Conclusion | 25 |
| 11. | References | 26 |

**List of Figures**

|  |  |
| --- | --- |
| Fig 1. | NDVI Values for Different Objects |
| Fig 2. | SAR Wavelength Interaction with Vegetation |
| Fig 3. | Different SAR Scattering Mechanisms |
| Fig 4. | Study Area: Siyana, U.P. |
| Fig 5. | Workflow |
| Fig 6. | Correlation Heatmap |
| Fig 7. | LAI vs VV Scatter Plot |
| Fig 8. | DBH vs VV/VH Scatter Plot |
| Fig 9. | Orchard Height vs VV Scatter Plot |
| Fig 10. | Canopy Circumference vs VV |
| Fig 11. | Effect of n-estimators on RMSE in RF for Canopy Circumference with input DBH and VH |
| Fig 12. | Effect of n-estimators on RMSE in RF for Canopy Circumference with input DBH and VV |
| Fig 13. | Effect of Epoch on Loss in ANN for DBH with input VH and Canopy Circumference |
| Fig 14. | Effect of Epoch on Loss in ANN for LAI with input VV and VH |
| Fig 15. | LAI Actual vs Predicted Values |
| Fig 16. | DBH Actual vs Predicted Values |
| Fig 17. | Orchard Height Actual vs Predicted Values |
| Fig 18. | Canopy Circumference Actual vs Predicted Values |
| Fig 19. | LAI Actual vs Predicted |
| Fig 20. | DBH Actual vs Predicted |
| Fig 21. | Orchard Height Actual vs Predicted |
| Fig 22. | Canopy Circumference Actual vs Predicted |
| Fig 23. | LAI Predicted Map |
| Fig 24. | Predicted Map of Canopy Circumference |
| Fig 25. | Predicted Map of DBH |
| Fig 26. | Predicted Map of Orchard Height |

**List of Tables**

|  |  |
| --- | --- |
| Table 1. | Sentinel-1 Information |
| Table 2. | Sentinel-2 Information |
| Table 3. | Biophysical Parameters with Remote sensing parameters derived using Sentinel 1 and sentinel 2. |
| Table 4. | Univariate Statistical Models for Deriving Different Biophysical Parameters |
| Table 5. | MAE and RMSE for different Biophysical Parameters |

**1. Abstract:**

Mangoes, the king of fruits form an integral part of the Indian households and the agricultural trade market. For obtaining a good yield, constant monitoring of mango orchards is required to protect the trees from instances of pest and disease. This study attempts to estimate biophysical parameters of mango orchards using SAR and optical data.

Accurate estimation of biophysical parameters such as leaf area index (LAI), diameter at breast height (DBH), plant height, and canopy circumference is crucial for effective agricultural monitoring and management.

In this study, we explore the potential of combining optical data from Sentinel-2 and synthetic aperture radar (SAR) data from Sentinel-1 for estimating these parameters in mango orchards situated in Siyana, Uttar Pradesh, India. A comprehensive analysis was conducted, including the derivation of various vegetation indices (VIs) from Sentinel-2 imagery and the extraction of VV and VH polarizations from Sentinel-1 SAR data. Correlation analysis and univariate statistical models were employed to examine the relationships between these remote sensing variables and field-measured biophysical parameters. Furthermore, machine learning models, including Random Forest (RF), Extreme Gradient Boosting (XG Boost), and Artificial Neural Networks (ANN), were utilized for estimation purposes. Our findings indicate that integrating SAR data, particularly VV and VH polarizations, with optical VIs significantly improves estimation accuracy compared to using optical data alone. The ANN model demonstrated superior performance in estimating LAI and DBH, while RF yielded the best results for plant height and canopy circumference estimation. This study underscores the synergistic benefits of multi-sensor data fusion for precise retrieval of biophysical parameters in mango orchards, facilitating enhanced orchard monitoring and management practices.

**2.Introduction**

India is the second largest producer of fruits and vegetables globally and occupies the first position in producing fruits like mango, banana, papaya, sapota, pomegranate, acid lime, and aonla and vegetables like peas and okra. India takes the 2nd position in the world in total horticultural production.

Advancement in technology makes it possible to get timely crop data over an area by using Remote sensing data from different satellites. In recent years, there has been a significant advancement in remote sensing technology, particularly in the field of earth observation. Optical and Synthetic Aperture Radar (SAR) sensors provide valuable information about the biophysical parameters of crops, such as crop type, growth stage, and yield estimation. Machine learning (ML) algorithms have been widely used to analyze remote sensing data for crop monitoring and management.

Machine learning algorithms are used to analyze and interpret the data obtained from these sensors to estimate various biophysical parameters such as vegetation indices, biomass, and crop yield. The estimation of biophysical parameters using Optical and SAR data using Machine Learning has the potential to revolutionize the way farmers manage their crops, leading to more sustainable and efficient agricultural practices.

**2.1. Gap in Existing Knowledge**

When it comes to mango orchards, very few studies have been carried out to study these trees and their biophysical parameters. There are certain studies conducted to estimate biophysical parameters using SAR and optical data on citrus or apple orchards. Combination of Sentinel-1 and Sentinel-2 can be used for various purposes which will provide great results. Red Edge bands, for the use of classification and biophysical parameters, need to be explored more and the same goes for the microwave data. Most of the studies focus on crop classification, monitoring phenology or seasonality. There is no existing relationship of the parameters such as plant height or DBH with the optical and SAR data for estimating the biophysical parameters. The convergence of these factors lead to the critical need to conduct research on mango orchards and its biophysical parameters.

**2.2. Objectives of Project:**

* To assess the sensitivity of biophysical parameters of mango trees with optical and SAR indices.
* To develop ML models for biophysical parameter retrieval.

**3. Optical Vegetation Indices**

The combinations of surface reflectance at two or more wavelengths that highlight a particular property of vegetation are termed as Vegetation Indices(VIs). Each index emphasizes different vegetation properties.

More than 150 VIs have been published in scientific literature but only some of them are systematically tested. These Indices are widely used to study vegetation dynamics. They are very useful in the determination of healthy vegetation, water bodies, bare soil, etc.

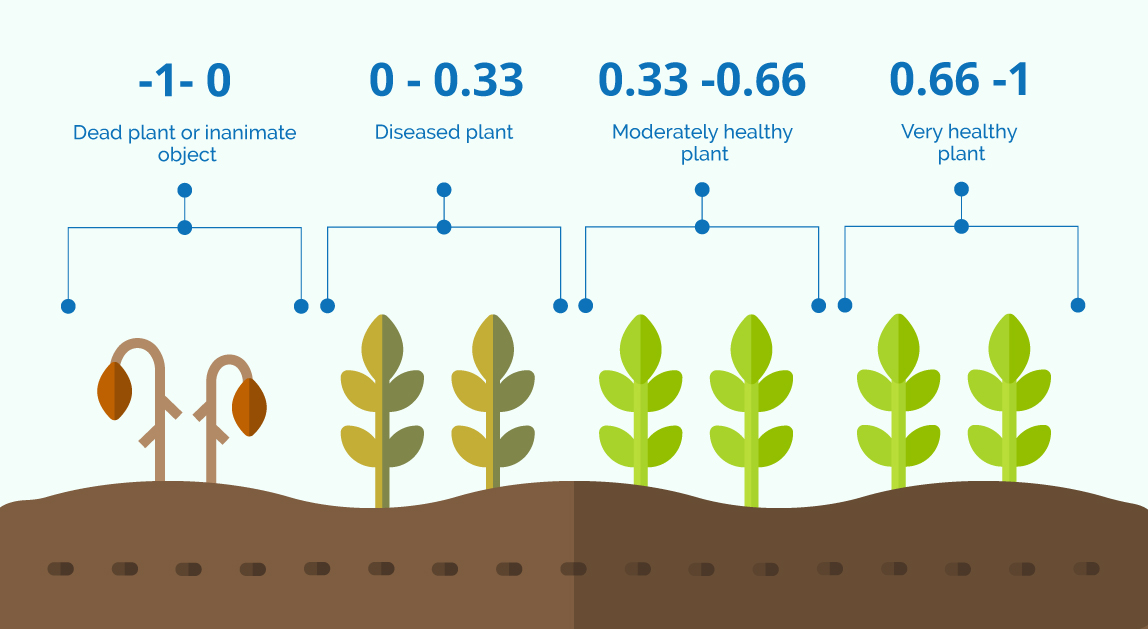
**3.1. Normalized Difference Vegetation Index**

This is also known as NDVI or greenness index. It is the most commonly used index in remote sensing which shows the health condition, density and greenness of the vegetation. NDVI is a dimensionless indicator that uses near-infrared and red bands of electromagnetic spectrum. The values ranges from -1 to 1 where:

* Values close to 1 indicate healthy, vigorous vegetation.
* Values close to 0 indicate less vegetation, bare soil or non-productive areas.
* Negative values indicate presence of water , snow or clouds.

The standard formula to calculate NDVI is:

*NDVI = (NIR - Red)/(NIR + Red)*



***Fig 1.* NDVI Values for different objects**

Source: <https://www.auravant.com/wp-content/uploads/2021/07/NDVI-values-by-crop-condition.jpeg>

**3.2. Normalized Difference Red Edge**

NDRE or Normalized Difference Red Edge is very much similar to NDVI. This is preferred over NDVI while sensing intense canopy as it uses red-edged light (instead of red light used in NDVI) that can penetrate leaves with much more intensity. Changes in chlorophyll content can also be detected using this indicator which is the leading indicator of nitrogen inside the leaves. The value of this indicator has its range from -1 to 1.

The standard formula to calculate NDRE is:

*NDRE = (NIR - Red Edge)/(NIR + Red Edge)*

**4. Synthetic Aperture Radar**

One of the powerful tools of remote sensing is Synthetic Aperture Radar(SAR). It is used to create two-dimensional images or three-dimensional reconstructions of objects, such as landscapes (NASA). The major difference between SAR and other earth observation instruments is that it can penetrate through clouds. SAR uses doppler effect generated by the forward motion of spacecraft to synthesize a large antenna. Data obtained from the SAR system depends upon the system (azimuth resolution) and specific parameters(wavelength, polarization and incident angle of the transmitted signal.

**4.1. SAR Polarisation**

Radar signals can transmit horizontal (H) or vertical (V) electric field vectors, and receive either horizontal (H) or vertical (V) return signals, or both. The orientation of the plane in which the transmitted electromagnetic wave oscillates is referred to as polarization. When signals are emitted in vertical and received in horizontal polarization, it would be indicated by VH. The same way when signal is emitted in horizontal and received in vertical polarization, it would be indicated by HV. The signal strength from these polarizations holds the information about the target surface.

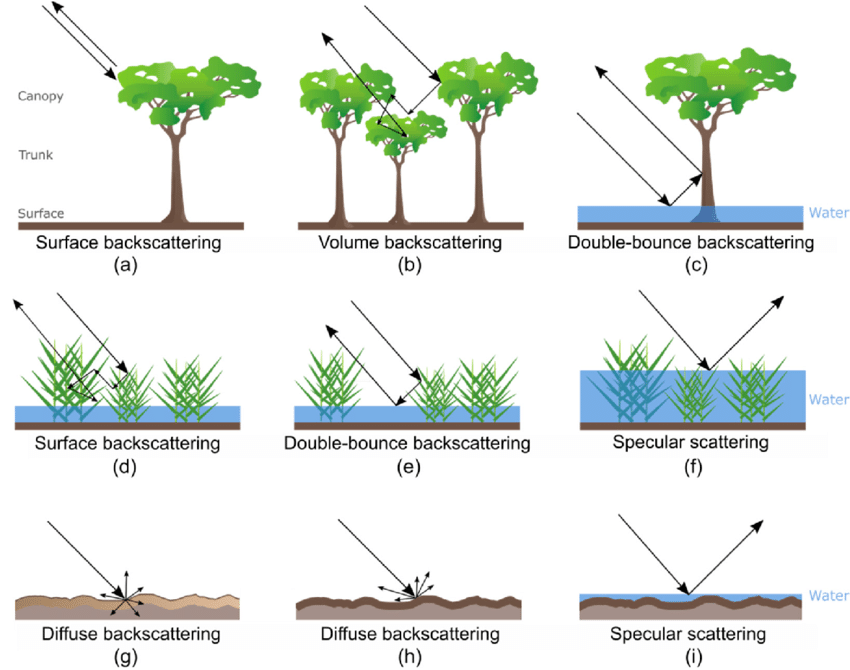
**4.2. SAR Interaction with Vegetation**

The interaction or microwave with vegetation canopy itself is heterogeneous. Volumetric scattering occurs in the vegetation canopies. Longer wavelengths penetrate dense canopies as the incident wavelength is larger than the canopy components.



***Fig 2.* SAR wavelength interaction with vegetation**

Source: <https://www.earthdata.nasa.gov/learn/backgrounders/what-is-sar>



***Fig 3.* Different SAR scattering mechanisms**

Source: <https://www.earthdata.nasa.gov/learn/backgrounders/what-is-sar>

**5. Literature Review**

A paper on **“Deep Learning-Based Estimation of Crop Biophysical Parameters Using Multi-Source and Multi-Temporal Remote Sensing Observation” (**Bahrami et al., 2021). This paper aims at modeling biophysical parameters of the crop, for example, Leaf Area Index(LAI) and biomass, using radar and optical data. The study was conducted for canola, soybean and corn over an agricultural area near Winnipeg, Canada. In situ measurements were collected during the SMAPVEX-12 campaign over Manitoba, Canada. Several polarimetric features were extracted from UAVSAR data. Besides, various spectral VIs were extracted from RapidEye optical data.Machine learning models used in the estimation were the regression models, that are Random Forest (RF), Gradient Boost (GB), Extreme Gradient Boost (XGB), and Support Vector Regression (SVR) and also Artificial Neural Networks (ANN) using Keras package. The results showed that the integration of SAR polarimetric and spectral VIs better estimate dry biomass and LAI. XGB showed great potential in assessing crop biophysical parameters. Also, GB and XGB have great potential in parameter estimation and remarkably improved accuracy.

**“Time series potential assessment for biophysical characterization of orchards and crops in a mixed scenario with Sentinel-1A SAR data”** (Haldar et al., 2020). In this study, potential for Sentinel-1A SAR data was assessed for the time-series analysis of orchard biophysical parameters and crop system. Biophysical parameters like DBH, canopy radius and visual height showed promising relationship with backscatter coefficients. The study area was Saharanpur, U.P. Sentinel-1A C-band IW GRD SAR data of nearly 24-35 days and six dates were used for the study. SAR data was processed in SNAP. Statistical analysis and multiple regression was performed. VV polarization was found to perform better than VH and ratio of VV and VH polarization. Significant relationship with biophysical parameters was observed in VV polarization than VH and VV/VH. This study reveals high potential of Sentinel-1A SAR data.

**“Leaf Area Index Estimation in a Heterogeneous Grassland Using Optical, SAR, and DEM Data”** (Lu et al., 2019) was published in 2019. This research integrated WorldView-2, Sentinel-1, and DEM images for retrieving vegetation LAI in a heterogeneous grassland area. 121 optical variables, 13 SAR variables, and 7 DEM variables were extracted from these 3 types of images, respectively. Correlation analysis was performed to evaluate correlations of these variables with LAI. Four combinations of the variables, including optical, optical + SAR, optical + DEM, and optical + SAR + DEM, were designed, aiming to evaluate the contribution of each type of data to the estimation of LAI. Four random forest models using these 4 combinations of variables were established for estimating LAI. Results show that many optical variables, including VIs, PCs, and textural variables, have stronger correlations with LAI than the SAR or DEM variables. However, SAR and DEM variables were rated more important than many optical variables in the random forest models and some optical variables that have strong correlations with LAI show low importance values.

**“Multi-Temporal Sentinel-1 and Sentinel-2 Data for Orchards Discrimination in Khairpur District, Pakistan Using Spectral Separability Analysis and Machine Learning Classification”** (Rehman et al., 2024). This study focused on examining the effectiveness of Sentinel-1 (S1) and Sentinel-2 (S2) satellite data for discriminating major orchards in the Khairpur district of the Sindh province, Pakistan using Machine Learning models. The effective combination of optical and SAR datasets has shown advantages in vegetation classification. For the study, 70% of the samples were collected during the field survey while the remaining 30% were collected from high-resolution google earth images. Four categories of LC (banana, mango, dates and other LC) were used for training. RF and SVM were the ML models used in the study for classification. The study concludes that multi-temporal fusion of S1 and S2 data, coupled with ML models, offer a reliable approach for orchard classification. The findings will be useful for orchard monitoring, improving yield estimation and precision based agricultural practices.

**“Evaluation of Multi-Orbital SAR and Multisensor Optical Data for Empirical Estimation of Rapeseed Biophysical Parameters”** (A. Allies et al., 2021). This article aims to evaluate the potential of multitemporal and multiorbital remote sensing data both in the microwave and optical domain to derive rapeseed biophysical parameters (crop height, dry mass, fresh mass and plant water content). NDVI, fCover, GAI were derived from dense temporal series of 98 Landsat-8 and Sentinel-2 images and backscattering coefficients and RVI were obtained from 231 images acquired by Sentinel-1. The study area chosen was southwestern and central France. The relationship between SAR and optical RSI, and the ground measurements were evaluated for the duration of 2017-2018 crop cycle. Firstly, evaluation was performed for SAR dataset and afterwards evaluation was performed for concurrent acquisitions of optical and SAR data. Further in the study, orbital effects on SAR data were analyzed and GAI, fCover and NDVI were compared with each other for all monitored fields (i.e., blue and red fields). The results revealed the complementarity of SAR and optical data throughout the phenological cycle of rapeseed crop. Promising results were observed by assimilating optical and SAR data into crop models.

**6. Study Area**

****

***Fig 4.* Study area: Siyana, U.P.**

The study area chosen for the study is Siyana block in Bulandshahr dist., U.P., India.

Location: Siyana Block is in the southeastern part of Bulandshahr district in Uttar Pradesh, India. It is approximately 30 km southeast of the city of Bulandshahr.

Coordinates: The approximate geographical coordinates of the Siyana block are latitude 28.1541° N and longitude 78.2463° E.

Area: The total area of the Siyana block is around 200 square kilometers, although the exact location may vary based on administrative boundaries and updates.

Topography: Siyana block is part of the Gangetic plain and has a relatively flat topography with fertile agricultural land. Small rivers, canals, and drains crisscross the region.

Climate: Siyana block experiences a hot semi-arid climate characterized by hot summers and cool winters. The summers, from April to June, are hot, with temperatures ranging from 30°C to 45°C, while the winters, from December to February, are excellent, with temperatures ranging from 8°C to 20°C. From July to September, the monsoon season brings moderate to heavy rainfall to the region.

**7. Dataset**

**7.1. Sentinel 1**

Sentinel-1 is a space mission funded by the European Union and carried out by the European Space Agency (ESA) within the Copernicus Programme. It captures C-band SAR imagery with different resolution (down to 5 m) and coverage (up to 400 km). Dual polarization , short revisit times and rapid product delivery is provided. It offers reliable, repeated wide area monitoring.

| **Specification** | **Sentinel-1** |
| --- | --- |
| Launch Date | April 3, 2014 (Sentinel-1A) and April 25, 2016 (Sentinel-1B) |
| Mission Duration | Seven years (planned) |
| Orbit Altitude | 693 km |
| Orbit Type | Sun-synchronous, polar orbit |
| Inclination | 98.18 degrees |
| Repeat Cycle | 12 days |
| Antenna Type | Synthetic Aperture Radar (SAR) |
| Frequency Band | C-band (5.405 GHz) |
| Polarization | Dual-polarization (HH and HV) |
| Swath Width | Up to 400 km |
| Spatial Resolution | Up to 5 m |
| Data Access | Free and open to all users |

**Table 1. Sentinel-1 Information**

**7.2. Sentinel-2**

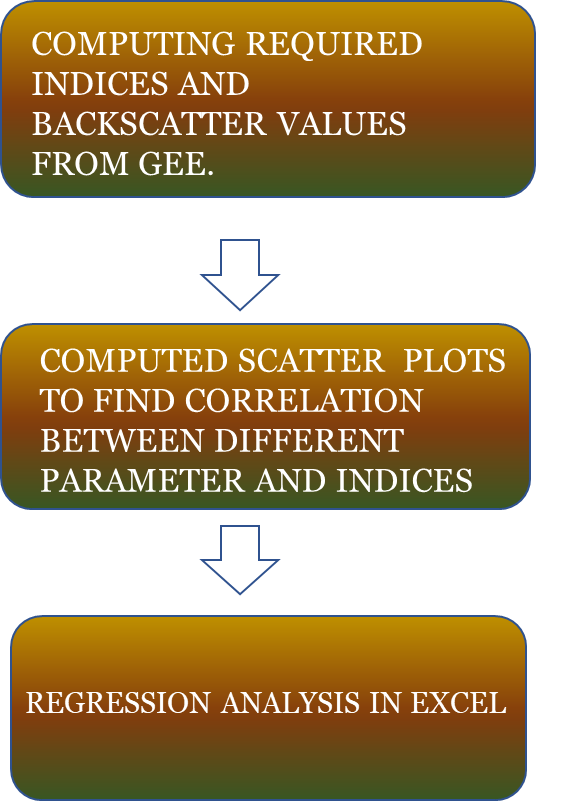
The Copernicus Sentinel-2 mission consists of a constellation of two polar-orbiting satellites in the same sun-synchronous orbit, phase at 180° to each other. It aims at monitoring variability in land surface and its wide swath width (290 km) and high revisit time (10 days at the equator with one satellite and five days with two satellites under cloud-free conditions, which results in 2-3 days at mid-latitudes) will support monitoring of Earth’s surface changes.

The Multispectral Instrument (MSI) works passively by collecting reflected sunlight from the Earth. The incoming light beam is split via a beam-splitter and focused onto two separate focal plane assemblies within the instrument; one Visible and Near-Infra-Red (VNIR) bands and the other one for Short Wave Infra-Red (SWIR) bands. The mission provides information for agricultural and forestry practices and helps manage food security.

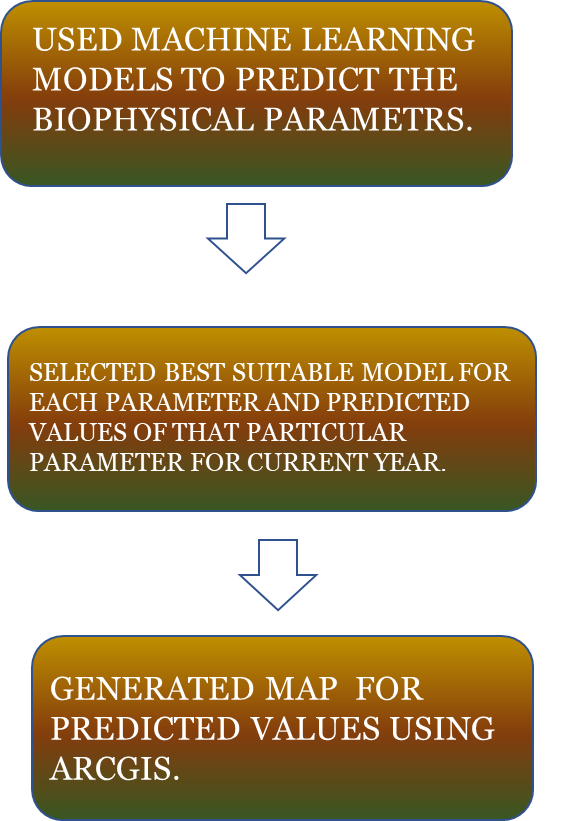
| **Specification** | **Sentinel-2** |
| --- | --- |
| Launch Date | June 23, 2015 (Sentinel-2A) and March 7, 2017 (Sentinel-2B) |
| Mission Duration | 7.25 years (planned) |
| Orbit Altitude | 786 km |
| Orbit Type | Sun-synchronous, polar orbit |
| Inclination | 98.6 degrees |
| Repeat Cycle | 10 days |
| Spectral Bands | 13 bands (4 visible/near-infrared and 9 shortwave infrared) |
| Spatial Resolution | 10 m (4 visible/near-infrared bands) and 20 m (9 shortwave infrared bands) |
| Data Access | Free and open to all users |

**Table 2. Sentinel-2 Information**

**8. METHODOLOGY:**

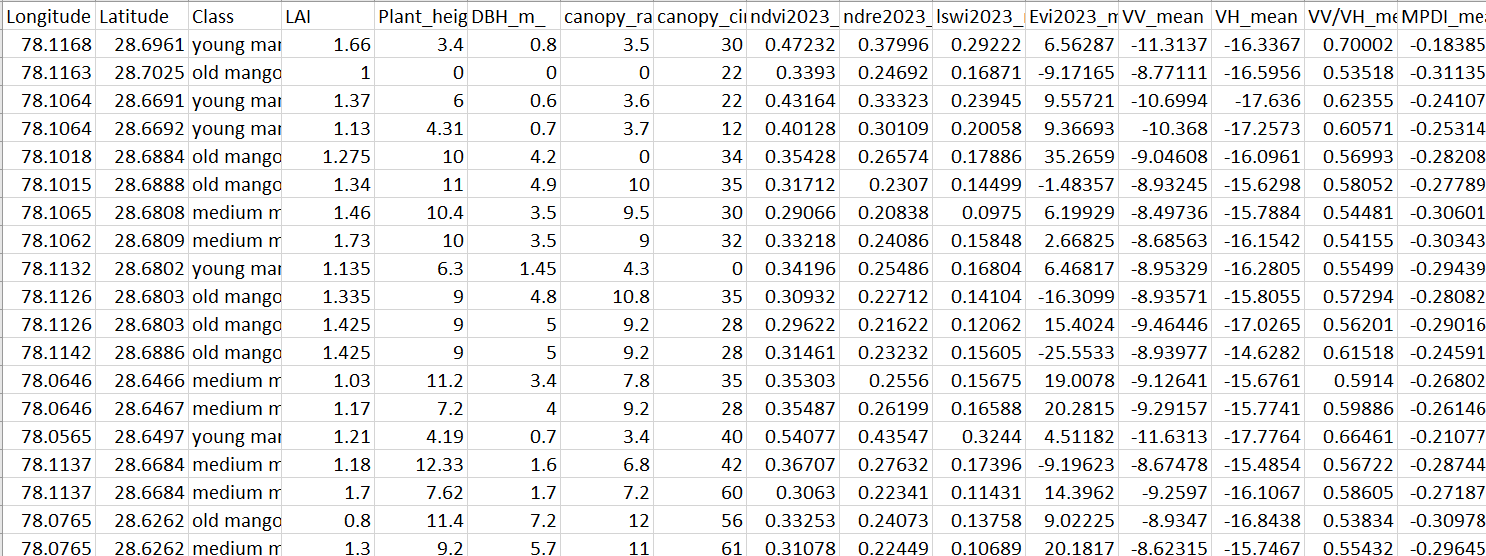






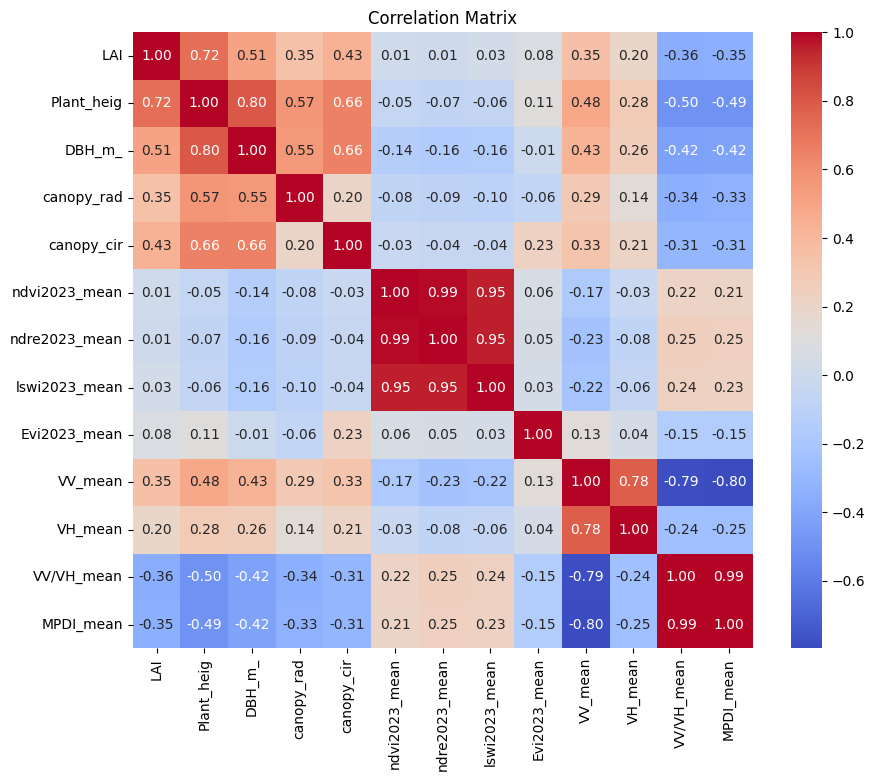
***Fig 5.* Workflow**

The study was carried out at Siyana Block in Uttar pradesh. Geographically, Siyana is located at gps coordinates of 28° 37' 40.4796'' N and 78° 3' 39.2868'' E. The various biophysical parameters LAI, orchard height, Diameter Breast height, Canopy Circumference were measured. The locations of all the plots were marked with the help of Global Positioning System (GPS) for the data collection. Remote Sensing variables such as NDVI, LSWI, EVI, NDRE, VV, VH, VV/VH, MPDI were calculated using Sentinel 1 and Sentinel 2 data in Google Earth Engine for the month of February 2023. February was selected since it is the peak flowering time for mango, for calculating indices values for Sentinel-1 and Sentinel-2 data in Google Earth Engine for Siyana's mango orchards. During this period, high backscatter ensures better differentiation of mango orchards from surrounding land cover types, reducing pixel mixing and improving accuracy in estimating biophysical parameters. Unlike optical data, Synthetic Aperture Radar (SAR) data used from Sentinel-1 does not require preprocessing in GEE. SAR penetrates clouds and provides reliable data even under cloud cover, eliminating the need for preprocessing and ensuring continuous monitoring of orchard dynamics regardless of weather conditions. Table 3 shows the data we used for further analysis.



**TABLE 3. Biophysical Parameters with Remote sensing parameters derived using Sentinel 1 and sentinel 2.**

The relationship of biophysical parameters with Sentinel-1A SAR backscatters (𝜎0) and their ratio is correlated using Pearson correlation coefficient. It was noticed that with the use of Pearson correlation coefficient ,the response of each polarization was different with Mango biophysical parameters. Also the SAR parameters correlated more with the biophysical parameters than the optical indices.



**Fig 6. Correlation heatmap.**

It is clear (from Fig 5) that the Pearson correlation coefficients obtained for 𝜎0VV (db) and 𝜎0VH (db) backscatter values are relatively strong as compare to sentinel 1 indices and their relationship with mango biophysical parameters. The most effective relationship found is the relationship between VV,VH and biophysical parameters itself. VV polarization reflects mainly from within the canopy, providing insights into canopy structure and volume scattering characteristics. Meanwhile, VH polarization interacts with both the canopy and the ground, offering information on surface roughness and moisture content. This dual-polarization approach enables a more comprehensive understanding of mango orchard characteristics, including tree density, canopy biomass, and soil moisture, thereby facilitating precise orchard management and monitoring.

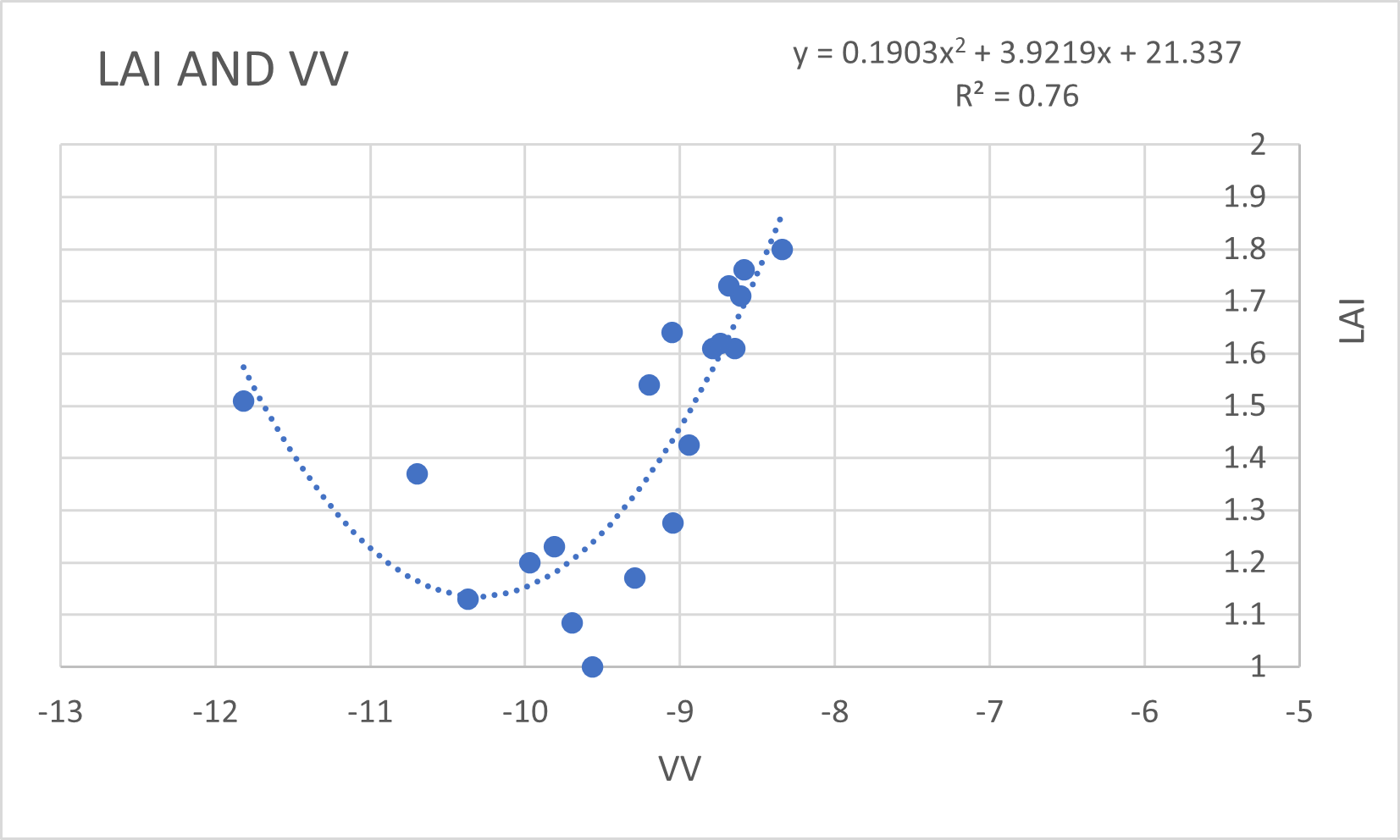
Simple univariate statistical models were developed in excel by creating a scatter plot between Remote sensing variables and Orchard biophysical parameters of Mango(LAI, DBH, orchard height, Canopy circumference). The scatter plots show that there is consistently a good relation between only certain biophysical parameters and Remote sensing variables. The resultant univariate models for different biophysical parameters with the most R square value was selected and calculated. These calculated Equations are shown in Table 4.

| **BIOPHYSICAL PARAMETER** | **EQUATION** | **R²** |
| --- | --- | --- |
| 1. LAI | y = 0.1903\*(VV)2 + 3.9219\*(VV) + 21.337 | R² = 0.76 |
| 1. DBH | y = 15.276ln(VV/VH) + 11.661 | R² = 0.76 |
| 1. PLANT HEIGHT | y = 1.1634(VV)2 + 25.571x + 144.63 | R² = 0.81 |
| 1. CANOPY CIRCUMFERENCE | y = 14.924(VV)2 + 303.48(VV) + 1559.4 | R² = 0.70 |

**Table 4. Univariate statistical models for deriving different biophysical parameters.**

**8.1. REMOTE SENSING VARIABLES AND BIOPHYSICAL PARAMETER CORRELATION**

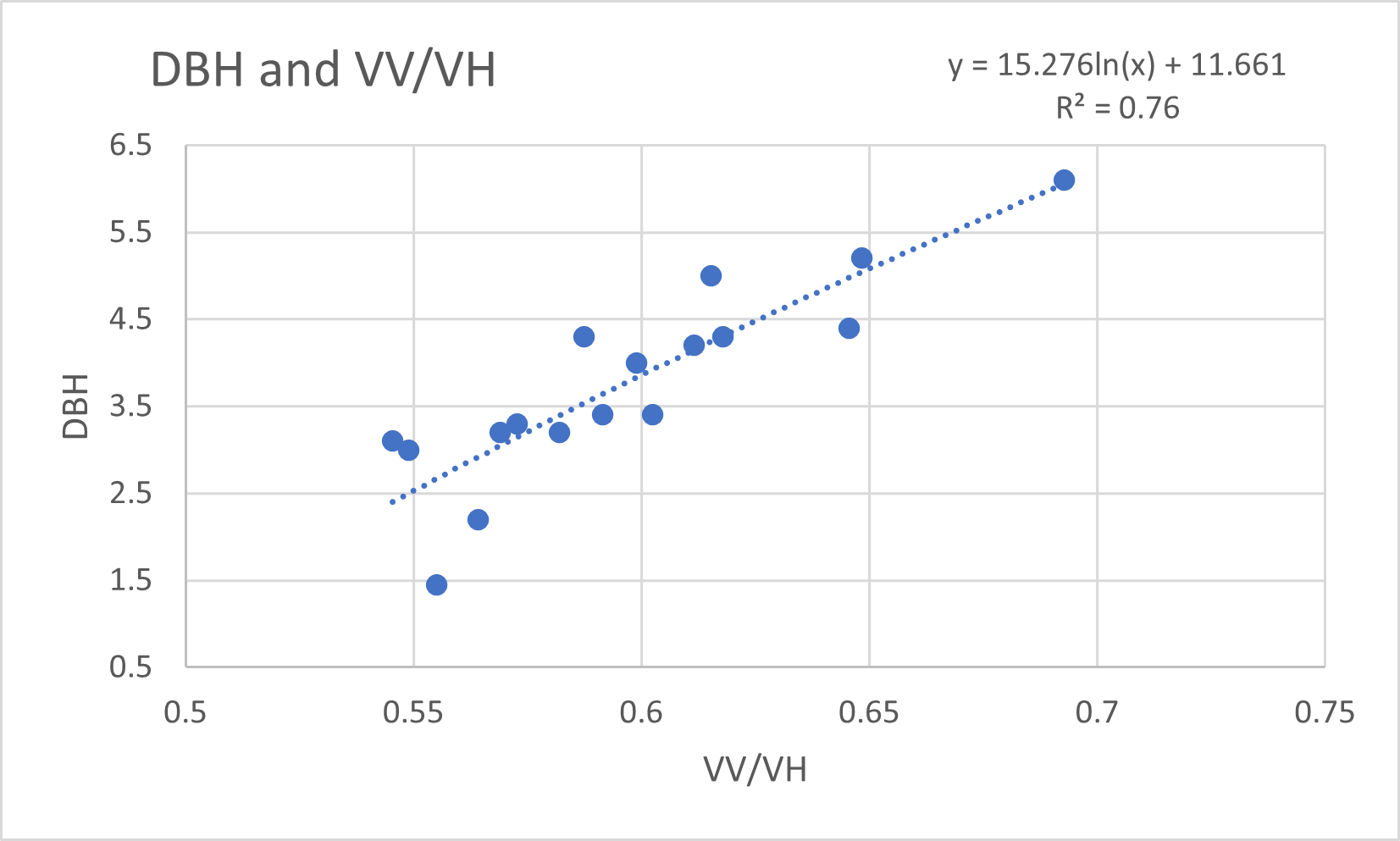
Numerous correlation plots were computed for different Remote sensing variables and Orchard Biophysical parameters. The Biophysical parameters showed a weak or moderate correlation with the Optical Indices. On the contrary to that the SAR parameters showed much better correlation with the Biophysical Parameters. Following are the Scatter plots that show a good correlation between the Remote Sensing Variables and Biophysical Parameters.



***Fig 7.* LAI vs VV scatter plot**

Leaf Area Index (LAI):

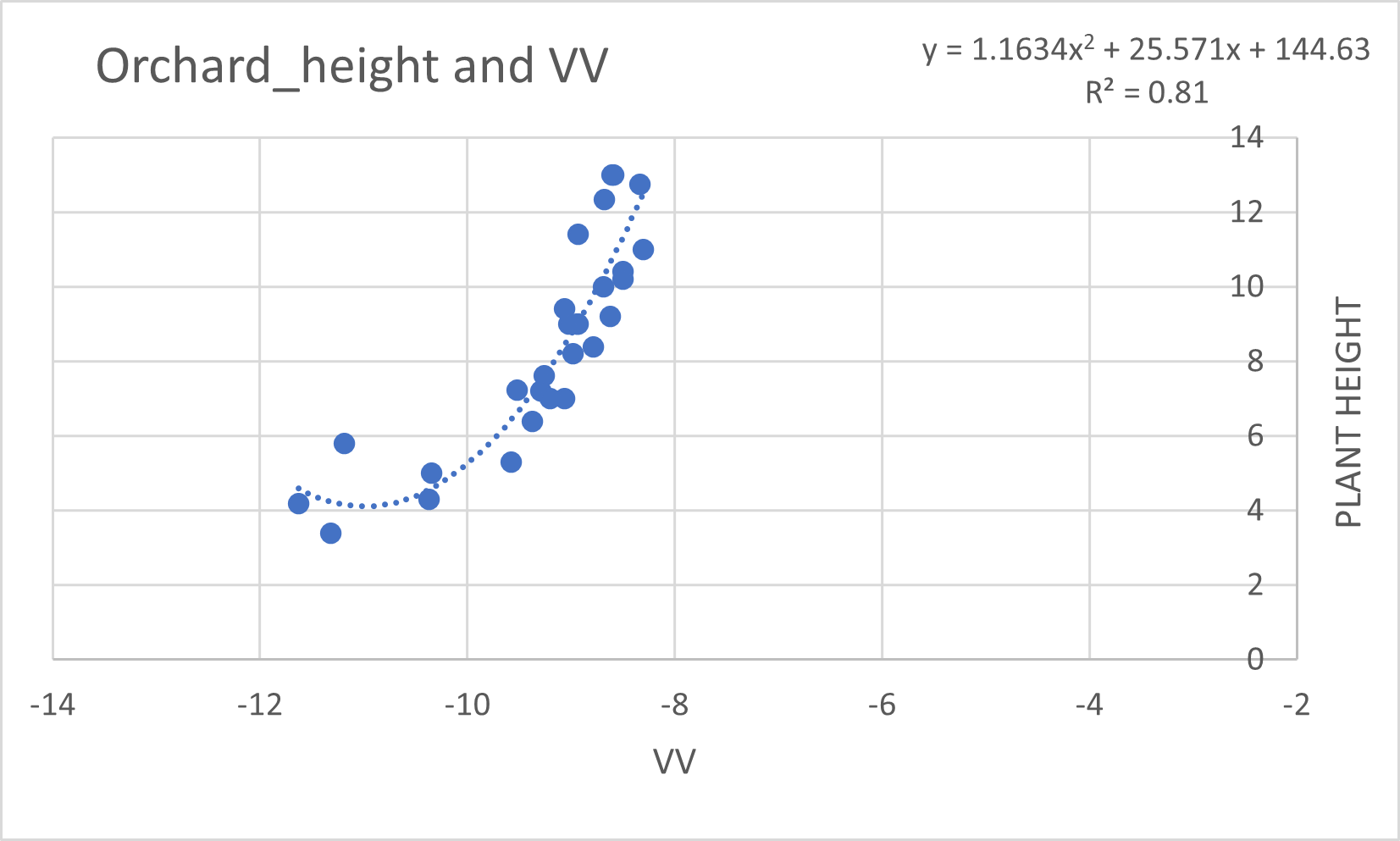
* LAI, representing leaf density, influences radar backscatter by affecting volume scattering within the canopy.
* VV polarization is sensitive to canopy structure and volume scattering providing additional information on LAI variations.
* The quadratic equation captures these interactions, resulting in a relatively high R² value of 0.76.



***Fig 8.* DBH vs VV/VH scatter plot**

Diameter at Breast Height (DBH):

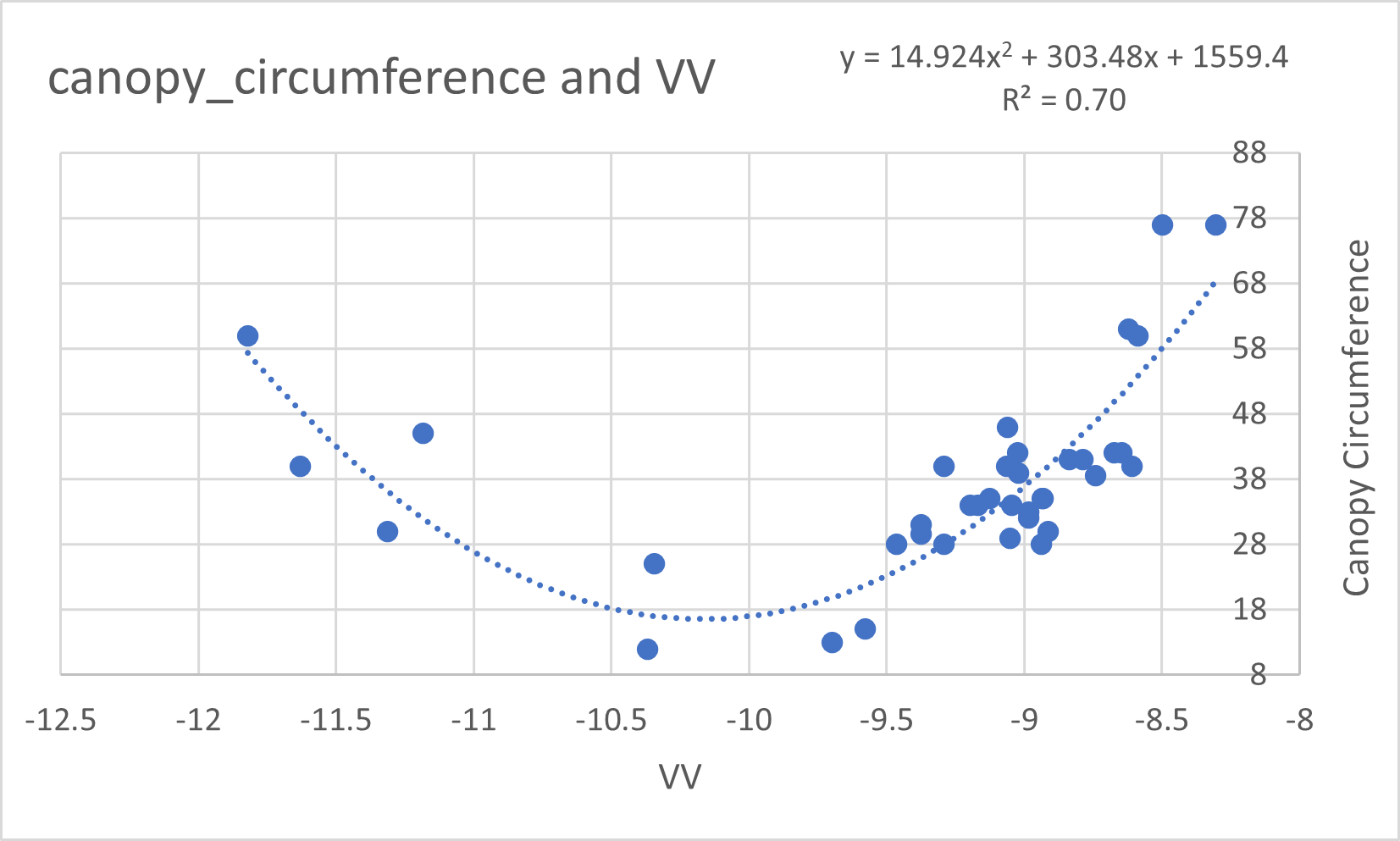
* DBH influences radar backscatter through its impact on canopy density and structure.
* The equation's logarithmic relationship with the ratio of VV to VH reflects the sensitivity of radar signals to variations in canopy structure and moisture content.
* This sensitivity, captured by the equation, yields a strong correlation (R² = 0.76) between DBH and radar backscatter.

****

***Fig 9.* Orchard Height vs VV scatter plot**

Orchard Height:

* Orchard height affects radar backscatter by influencing the scattering properties of the vegetation layer.
* The quadratic relationship with VV indicates the sensitivity of radar signals to variations in plant height and canopy structure resulting in a high R² value of 0.81.



***Fig 10.* Canopy circumference vs VV**

Canopy Circumference:

* Canopy circumference reflects the overall size and structure of the tree canopy, impacting radar backscatter.
* The quadratic relationship with VV suggests the sensitivity of radar signals to variations in canopy size and structure.
* The equation's high R² value of 0.70 indicates that VV polarization effectively captures variations in canopy circumference, possibly due to its sensitivity to canopy volume scattering.

**8.2. ML Models used for the study**

For the estimation of biophysical parameters, three models viz. Random Forest (RF), Extreme Gradient Boost (XGB) and Artificial Neural Network (ANN) were used. Every parameter was estimated using these models with different input variables.

**8.2.1. Random Forest Regressor (RF)**

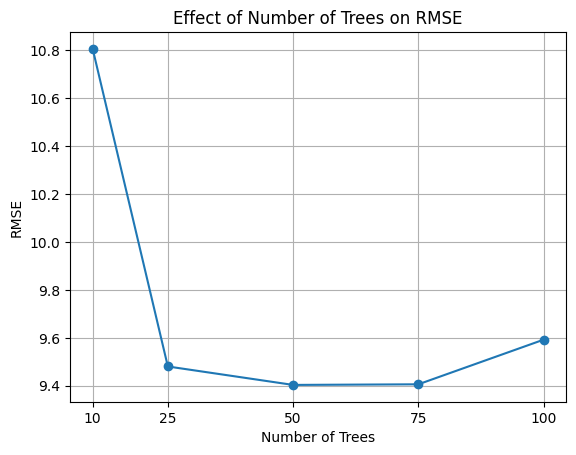
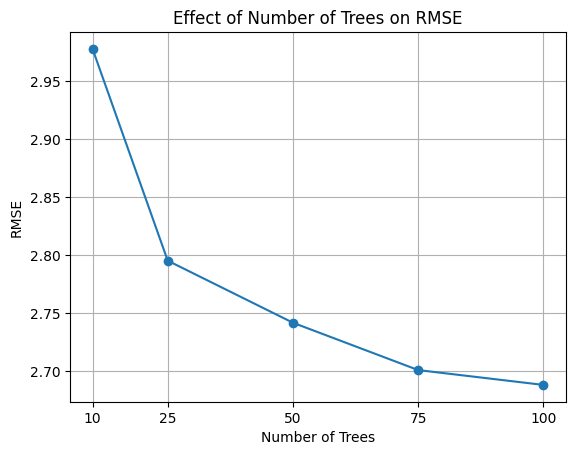
In RF, each tree is trained on a random subset of the input data and a random subset of the input features. This improves the model’s accuracy and reduces overfitting. Random forest regressor has several advantages over other regression models. For example, it is robust to noisy data, can handle missing values, and is relatively insensitive to the choice of hyperparameters. It also has good interpretability, as the importance of each input feature can be easily determined from the model.

**8.2.2. Extreme Gradient Boost (XGB)**

XGB is a scalable and efficient implementation of the gradient boosting algorithm, which is an ensemble machine learning technique. It works by sequentially adding weak learners (typically decision trees) to the model, where each new tree aims to correct the errors made by previous trees. This method has built-in regularization techniques (like L1 and L2) to prevent overfitting and improve generalization,

**8.2.3. Artificial Neural Network**

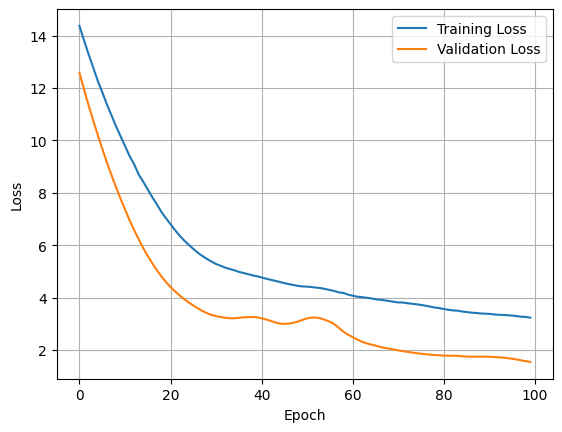
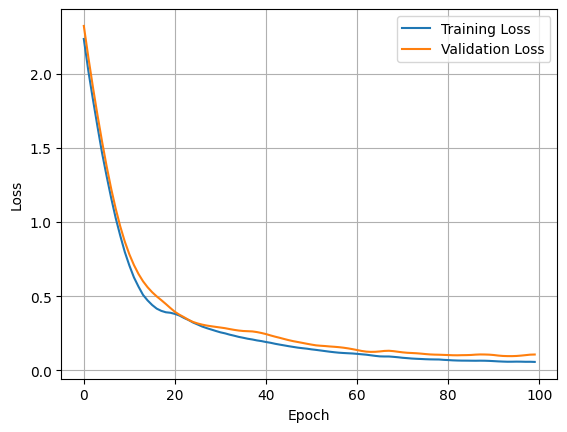
ANNs are a type of machine learning model inspired by the structure and function of biological neural networks in the human brain. They consist of interconnected nodes (neurons) arranged in layers, with each connection having a weight that determines the strength of the connection. ANNs can learn complex nonlinear relationships between input features and target variables through a process called backpropagation.

***Fig 11.* Effect of n-estimators on RMSE in RF *Fig 12.*Effect of n-estimators on RMSE**

**for Canopy Circumference with input in RF for Orchard Height with**

**DBH and VH. input DBH and VV.**

***Fig 13.* Effect of epoch on loss in ANN for  *Fig 14.* Effect of epoch on loss in ANN for**

**DBH with input VH and Canopy LAI with input VV and VH.**

**Circumference.**

**9. Result and Discussion**

**9.1. ML Models**

The values of all the Biophysical parameters were predicted for the year 2024 using the Following are the plots that signify how far the predicted values are from the target values. While estimating the biophysical parameters in ML, XGB did not show good results while RF and ANN gave significant results. Each parameter performed and gave different RMSE and MAE for different models with different outputs.

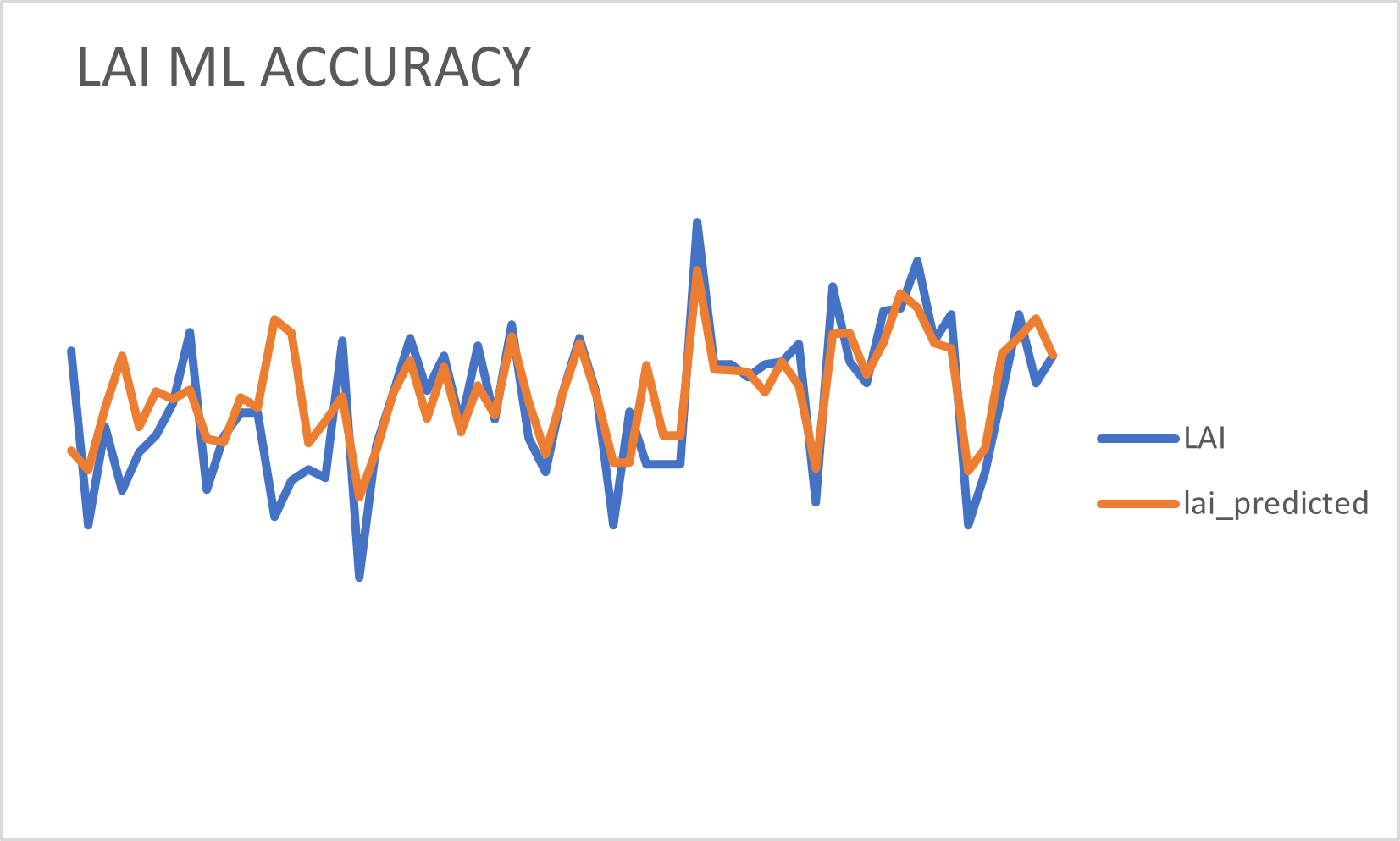
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **MODEL** | **Estimation** | **Input** | **MAE**  **(unit)** | **Percent** | **RMSE**  **(unit)** | **Percent** |
| **XGB** | **LAI** | VV, VH | 0.33 | 16.92% | 0.42 | 21.03% |
|  |  | NDVI, NDRE | 0.31 | 15.76% | 0.38 | 19.14% |
|  | **DBH** | VV, VH | 1.48 m | 29.65% | 2.05 m | 41.12% |
|  |  | VH, Circumference | 1.62 m | 32.41% | 2.04 m | 40.93% |
|  |  | VH, Height | 1.23 m | 24.66% | 1.72 m | 34.4% |
|  | **Height** | DBH, VV | 3.02 m | 22.99% | 3.70 m | 28.11% |
|  |  | VV, VH | 3.06 m | 23.34% | 3.73 m | 28.37% |
|  |  | DBH, VH | 3.09 m | 23.49% | 3.67 m | 27.91% |
|  | **Canopy Circumference** | VV | 11.52 m | 20.95% | 14.21 m | 25.84% |
|  |  | DBH, VH | 8.95 m | 16.28% | 10.88 m | 19.78% |
|  |  | VV, VH | 12.5 m | 22.76% | 16.45 m | 29.92% |
| **RF** | **LAI** | VV, VH | 0.31 | 15.71% | 0.38 | 19.37% |
|  |  | NDRE, NDVI | 0.28 | 14.36% | 0.36 | 18.2% |
|  | **DBH** | VV, VH | 1.57 m | 31.54% | 2.25 m | 45.1% |
|  |  | VH, Circumference | 1.16 m | 23.38% | 1.56 m | 31.37% |
|  |  | VH, Height | 0.87 m | 17.46% | 1.21 m | 24.35% |
|  | **Height** | DBH, VV | **2.24 m** | **17.01%** | **2.68 m** | **20.41%** |
|  |  | VV, VH | 2.89 m | 21.9% | 3.54 m | 26.88% |
|  |  | DBH, VH | 2.60 m | 19.76% | 2.99 m | 22.77% |
|  | **Canopy Circumference** | VV | 14.36 m | 26.11% | 15.56 m | 28.3% |
|  |  | DBH, VH | **8.10 m** | **14.73%** | **9.59 m** | **17.44%** |
|  |  | VV, VH | 10.40 m | 18.92% | 13.58 m | 24.7% |
| **ANN** | **LAI** | VV, VH | 0.28 | 14.45% | 0.33 | 16.5% |
|  |  | NDVI,  NDRE | **0.28** | **14.38%** | **0.34** | **17%** |
|  | **DBH** | VV, VH | 1.63 m | 32.6% | 2.25 m | 45.15% |
|  |  | VH, Circumference | **0.93 m** | **18.7%** | **1.34 m** | **26.86%** |
|  |  | VH, Height | 1.04 m | 20.88% | 1.29 m | 25.93% |
|  | **Height** | DBH, VV | 3.09 m | 23.5% | 3.85 m | 29.24% |
|  |  | VV, VH | 3.3 m | 25.1% | 4.07 m | 30.9% |
|  |  | DBH, VH | 2.56 m | 19.46% | 3.42 m | 26.01% |
|  | **Canopy Circumference** | VV | 12.79 m | 23.26% | 16.82 m | 30.59% |
|  |  | DBH, VH | 8.86 m | 16.1% | 11.24 m | 20.43% |
|  |  | VV, VH | 12.06 m | 21.93% | 16.41 m | 29.85% |

**Table 5. MAE and RMSE for different Biophysical Parameters**

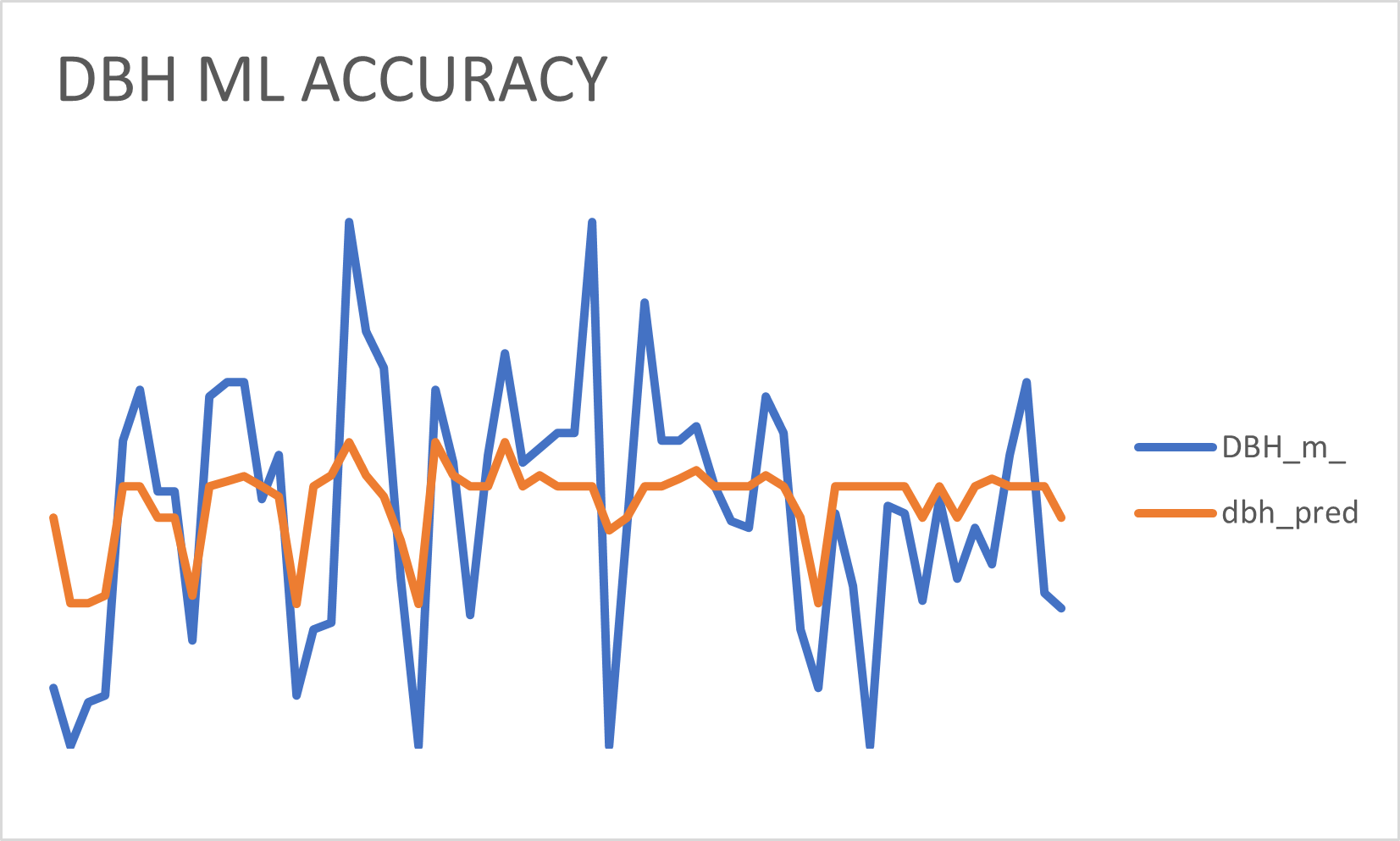
As shown in the table above, LAI and DBH gave better results in ANN. LAI with the input variables, NDRE ,NDVI and VV, VH in ANN with RMSE 0.34, MAE 0.28 and RMSE 0.33 and MAE 0.28 respectively, did not show much difference. In ANN, hyperparameter tuning was done which improved the accuracy. DBH, on the other hand with the input VH and Canopy Circumference gave best results with RMSE 1.34 m and MAE 0.93 m. This could be because of the relationship between DBH and Canopy Circumference. More the Circumference, more will be the DBH. Also, VH is most sensitive to volume scattering that is caused by the leaves and branches of the orchard canopy that means it is directly related with the Canopy Circumference. These could be the possible reasons for better estimation of DBH with VH and Canopy Circumference as inputs.

For Orchard Height and Canopy Circumference, RF was the best model. Orchard height was estimated best with the inputs, DBH and VV giving MAE 2.24 m and RMSE 2.68 m. VV backscatter is best suited for vertical objects and gives better values for them. Height being a vertical parameter gives better results with VV as input. DBH is also related with height, when height increases the canopy structures tends to increase with it and therefore DBH also increases. VV and DBH in ML also proved their relationship with Orchard Heights. In RF, with the number of trees the error gets reduced and thus provides good results. RMSE 9.59 m and MAE 8.10 m was observed for Canopy Circumference when the input was VH and DBH.

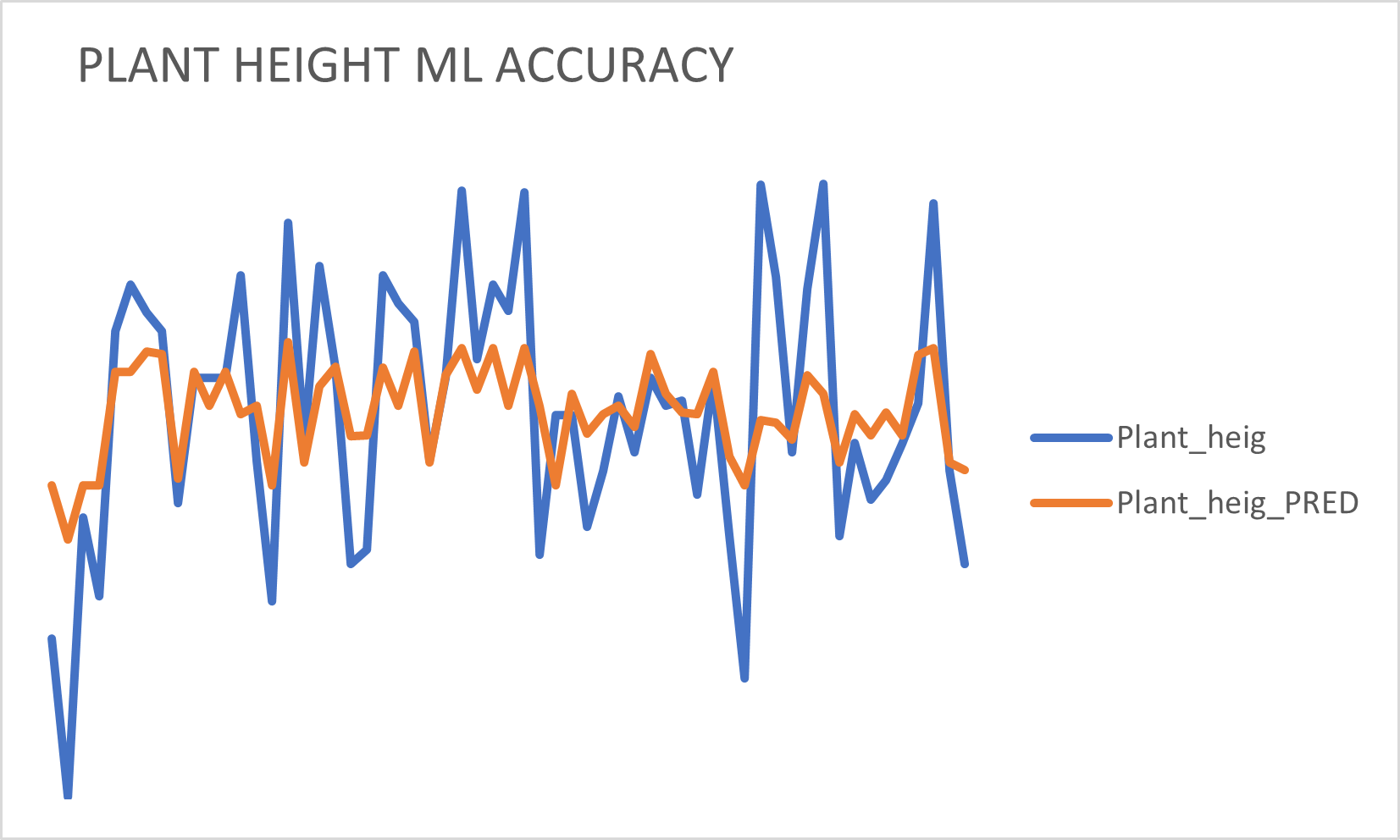
Following are the Accuracy plots for different Parameters:

****

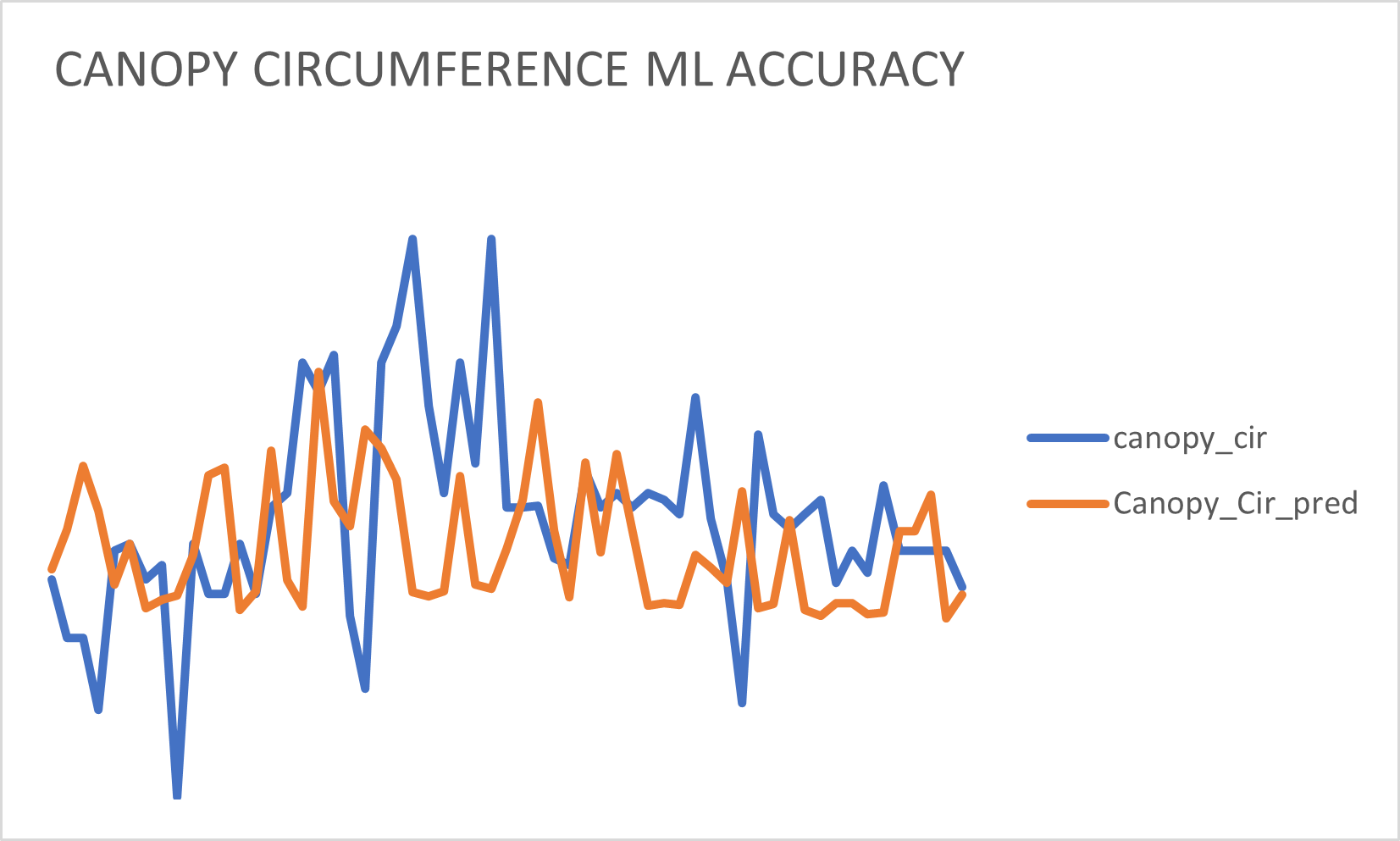
***Fig 15.* LAI Actual vs Predicted Values**

****

***Fig 16.* DBH Actual vs Predicted Values**

****

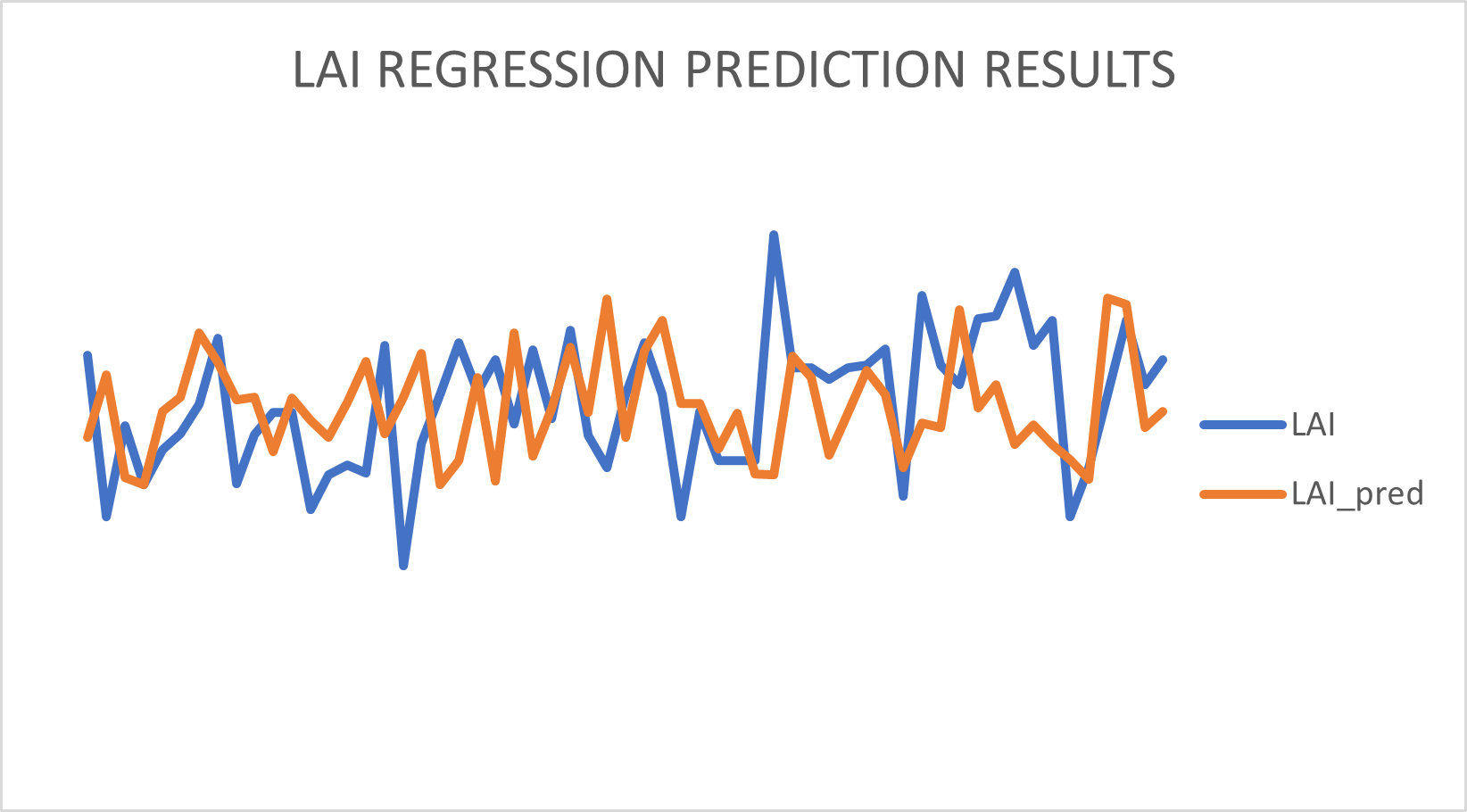
***Fig 17.* Orchard height Actual vs Predicted Values**

****

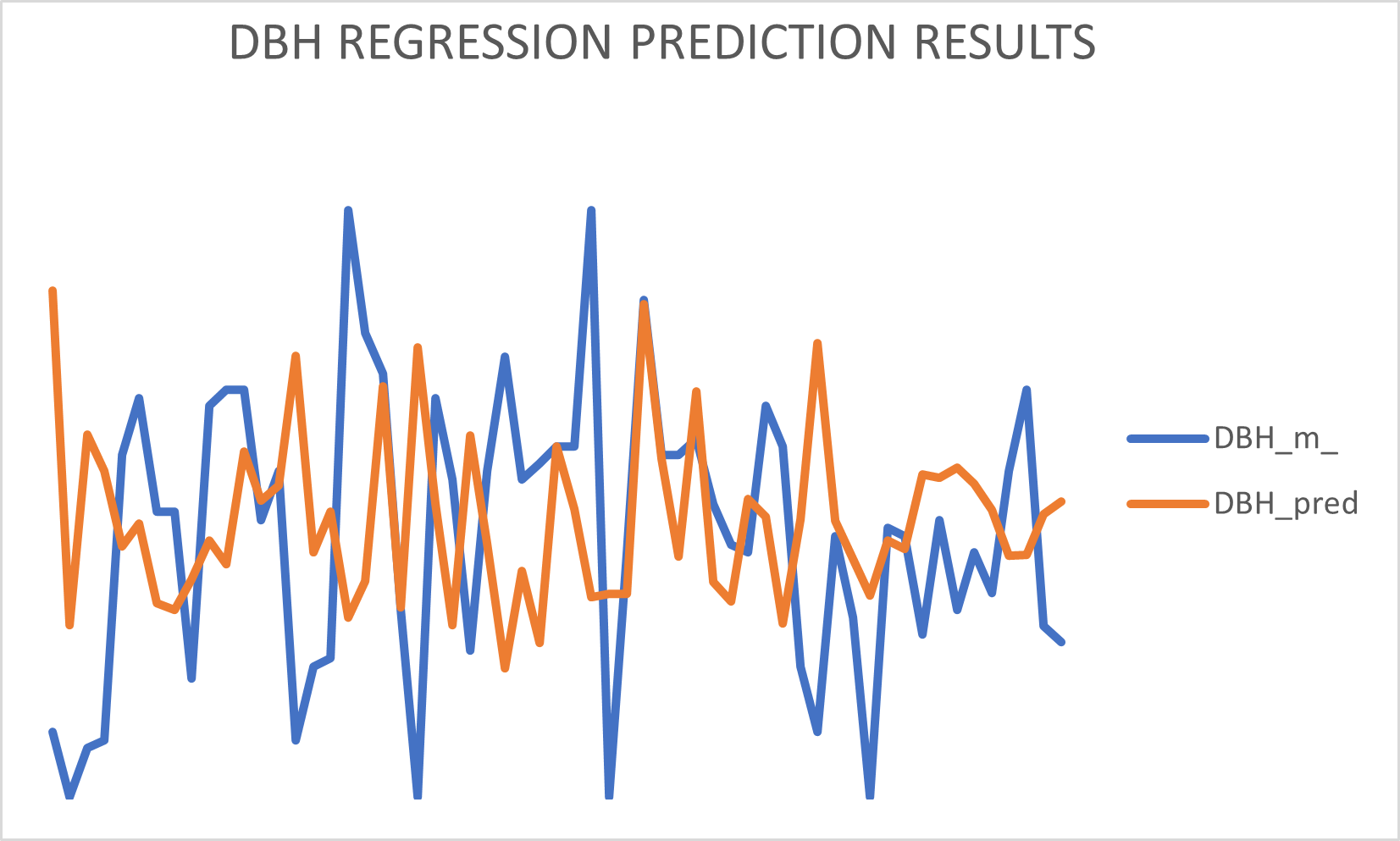
***Fig 18.* Canopy Circumference Actual vs Predicted Values**

**9.2. Regression Analysis:**

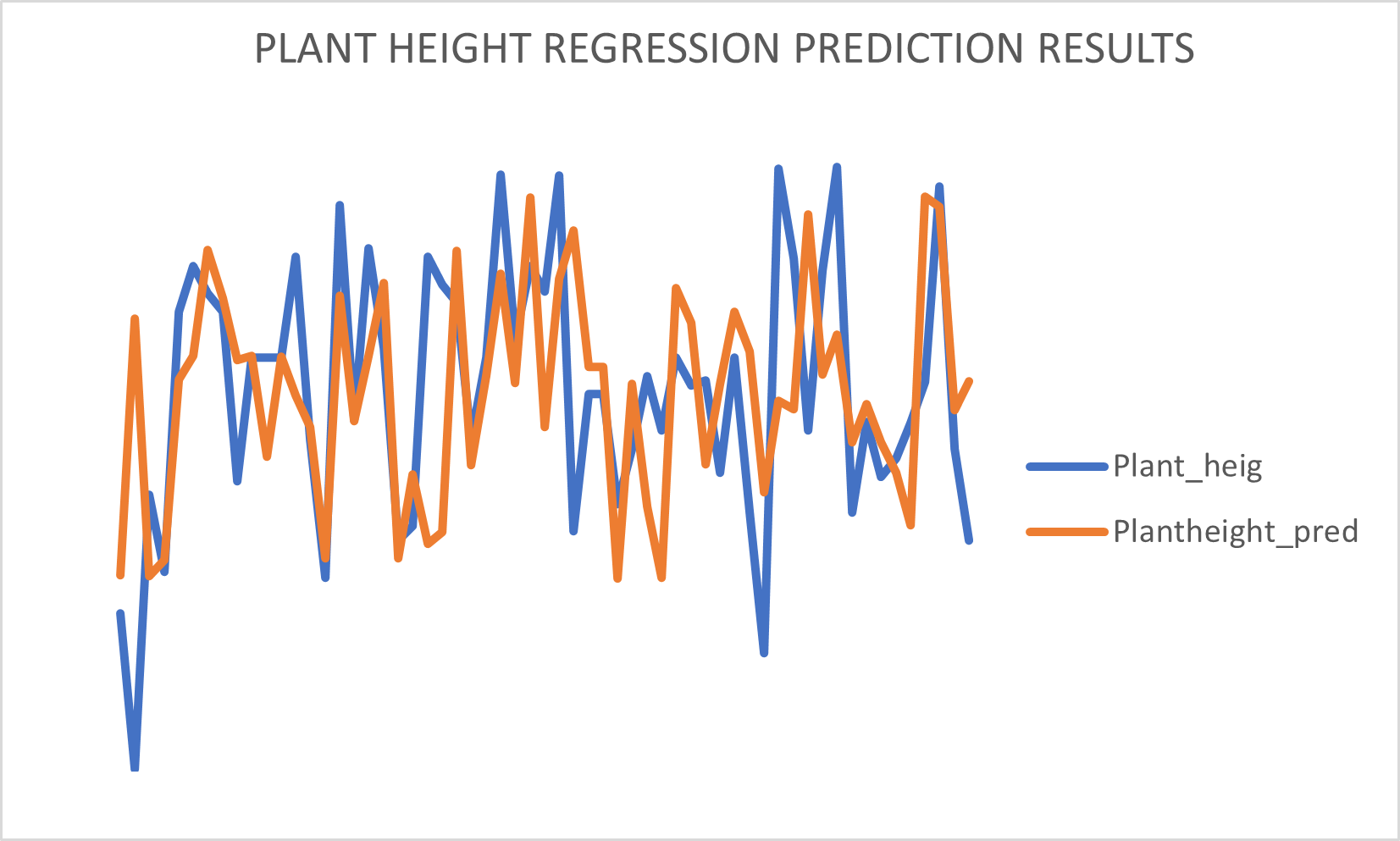
The values of all the Biophysical parameters were calculated for the year 2024 using the equations that were obtained from the scatter plots. Following are the plots that signify how far the predicted values are from the target values.



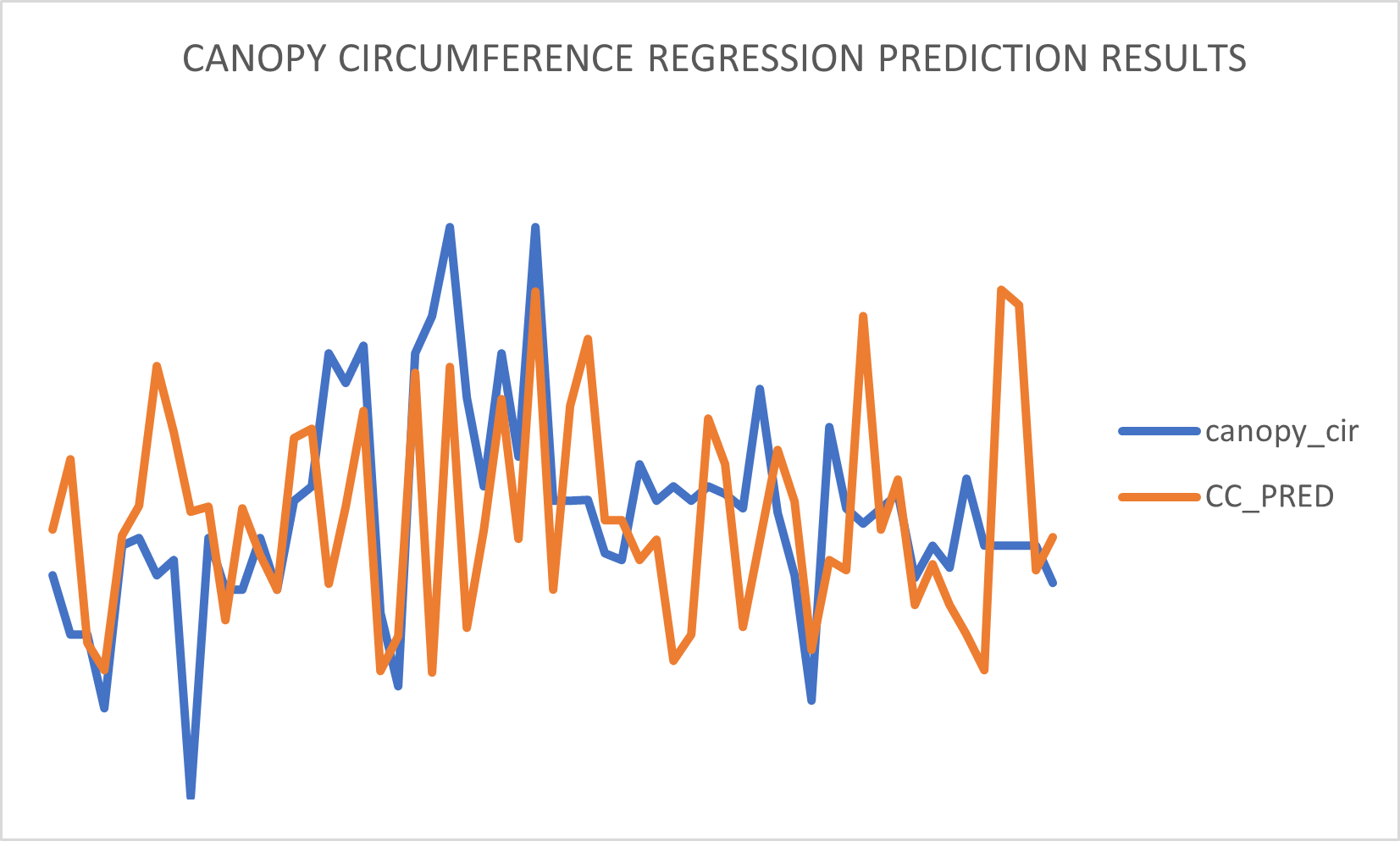
***Fig 19.* LAI Actual vs Predicted**



***Fig 20.* DBH Actual vs Predicted**

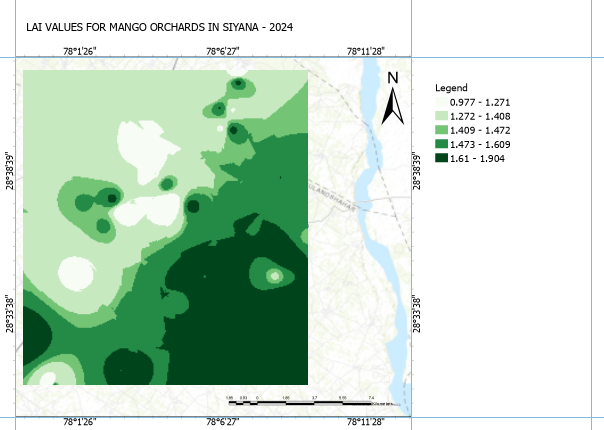


***Fig 21.* Orchard Height Actual vs Predicted**

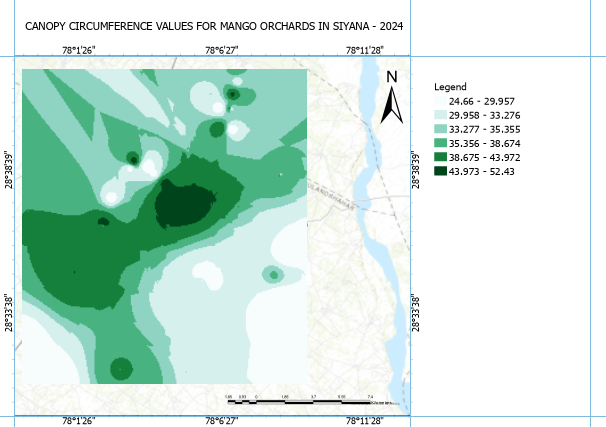


***Fig 22.* Canopy Circumference Actual vs Predicted**

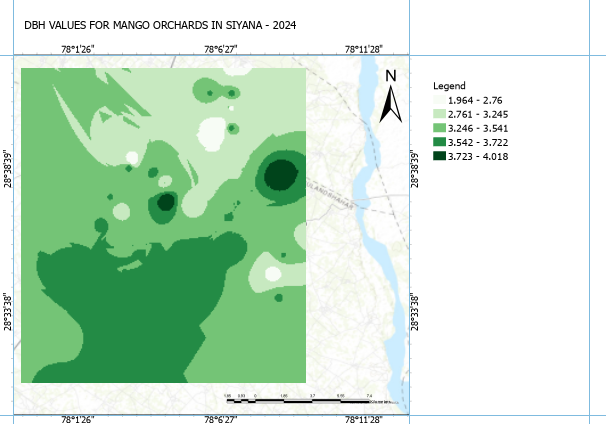
**9.3. FINAL MAP**

****

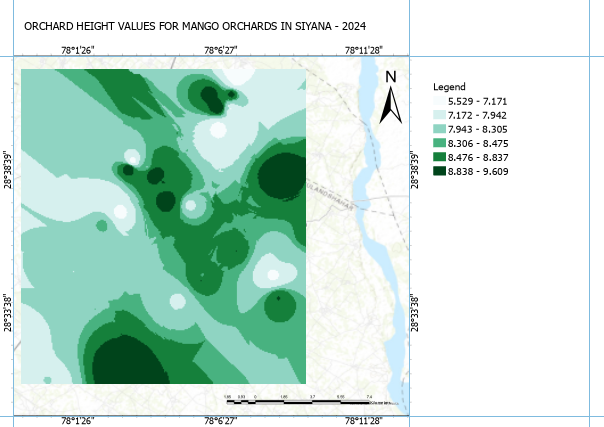
***Fig 23.* LAI Predicted Map**



***Fig 24.* Predict Map of Canopy Circumference**



***Fig 25.* Predicted Map of DBH**



***Fig 26.* Predicted Map of Orchard Height**

**10. Conclusion**

The amalgamation of Synthetic Aperture Radar (SAR) data analysis with advanced Machine Learning (ML) techniques surpasses traditional regression models in estimating biophysical parameters crucial for agricultural monitoring and land management. ML algorithms exhibit superior predictive capabilities, leveraging the complex relationships within SAR data to deliver more accurate and nuanced estimations of parameters like Leaf Area Index (LAI), Plant Height, and Soil Moisture. This advancement signifies the transformative potential of ML in revolutionizing agricultural assessments, offering unparalleled precision and insight for sustainable land management practices.

**11. References**

1. Sahu, H., Haldar, D., Danodia, A., & Kumar, S. (2020). Time series potential assessment for biophysical characterization of orchards and crops in a mixed scenario with Sentinel-1A SAR data. *Geocarto International*, *35*(14), 1627–1639.

<https://doi.org/10.1080/10106049.2019.1583776>

2. Lu, B., & He, Y. (2019). Leaf Area Index Estimation in a Heterogeneous Grassland Using Optical, SAR, and DEM Data. *Canadian Journal of Remote Sensing*, *45*(5), 618–633.<https://doi.org/10.1080/07038992.2019.1641401>

3. A. Allies et al., "Evaluation of Multiorbital SAR and Multisensor Optical Data for Empirical Estimation of Rapeseed Biophysical Parameters," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 7268-7283, 2021, [doi: 10.1109/JSTARS.2021.3095537](https://www.researchgate.net/publication/353107349_Evaluation_of_Multiorbital_SAR_and_Multisensor_Optical_Data_for_Empirical_Estimation_of_Rapeseed_Biophysical_Parameters).

4. Bahrami, H.; Homayouni, S.; Safari, A.; Mirzaei, S.; Mahdianpari, M.; Reisi-Gahrouei, O. Deep Learning-Based Estimation of Crop Biophysical Parameters Using Multi-Source and Multi-Temporal Remote Sensing Observations. *Agronomy* **2021**, *11*, 1363. <https://doi.org/10.3390/agronomy11071363>

5. Rehman, A.U.; Zhang, L.; Sajjad, M.M.; Raziq, A. Multi-Temporal Sentinel-1 and Sentinel-2 Data for Orchards Discrimination in Khairpur District, Pakistan Using Spectral Separability Analysis and Machine Learning Classification. *Remote Sens.* **2024**, *16*, 686. <https://doi.org/10.3390/rs16040686>

6. Pandey, Arvind & Ram, Hebbar & Palni, Sarita & Rawat, Jiwan & Raj, Uday. (2022). Inventory and Phenological Assessment of Apple Orchards Using Various Remote Sensing Techniques for Shopian District of Jammu and Kashmir State, India. [10.1007/978-981-16-7731-1\_16](https://www.researchgate.net/publication/360165745_Inventory_and_Phenological_Assessment_of_Apple_Orchards_Using_Various_Remote_Sensing_Techniques_for_Shopian_District_of_Jammu_and_Kashmir_State_India).

7. Jesus, Janisson & Kuplich, Tatiana. (2020). Applications of SAR data to the estimate of forest biophysical variables in Brazil. Cerne. 26. 88. [10.1590/01047760202026012656.](https://www.scielo.br/j/cerne/a/fxLL8MrKSF8H9Mh45BCtkdm/?lang=en)

8. Santos, J. R., Gama, F. F., & Bispo, P. C. (Year). Estimating forest biomass by remote sensing radar data in Brazil. [Journal Title], [Volume](Issue), [Pages].