PADDY RICE LEAF AREA INDEX (LAI) ESTIMATION WITH MACHINE LEARNING MODELS USING SYNTHETIC APERTURE RADAR AND OPTICAL DATA

SENTETİK AÇIKLIKLI RADAR VE OPTİK VERİLER KULLANILARAK MAKİNE ÖĞRENMESİ MODELLERİYLE ÇELTİK BİTKİSİ YAPRAK ALAN İNDEKS (YAI) TAHMİNİ

GÜLSÜM POSTALLI

PROF. DR. SAYGIN ABDİKAN

Supervisor

Submitted to

Graduate School of Science and Engineering of Hacettepe University

as a Partial Fulfillment to the Requirements

for the Award of the Degree of Master of Science

in Geomatics Engineering

January 2025

ABSTRACT

PADDY RICE LEAF AREA INDEX (LAI) ESTIMATION WITH MACHINE LEARNING MODELS USING SYNTHETIC APERTURE RADAR AND OPTICAL DATA

Gülsüm POSTALLI

Master of Science, Department of Geomatics Engineering

Supervisor: Prof. Dr. Saygın ABDİKAN

January 2025, 66 pages

This thesis aims to estimate the Leaf Area Index (LAI) by examining the relationship between LAI measurements taken from rice fields in the Gönen district of Balıkesir province and images obtained from Sentinel-1 Synthetic Aperture Radar (SAR) and Sentinel-2 optical satellites. In this aim, machine learning models which Linear Regression (LR), Random Forest Regression (RFR), and Support Vector Regression (SVR), were applied and compared. As part of the project, LAI measurements were conducted in-site studies using an LAI meter on four dates in June, July, August, and September 2024. The most up-to-date and suitable (cloud-free) satellite images were selected based on the LAI measurement dates. LAI estimation was performed in two ways: individually for each date and collectively for all dates. For optical data, the lowest RMSE value (0.3236) for a single date was obtained with the RFR model in June 2024, while the lowest %RMSE (19.8574%) was recorded in September. When considering all dates together, the lowest RMSE (0.9797) and %RMSE (24.5200) were found in the LR

model, while the highest correlation was observed in the B8A band (0.7473). For Sentinel-1 data, the lowest RMSE value (0.3197) for a single date was obtained with the LR model in June 2024. When all dates were considered together, the lowest RMSE (1.2907) was also found in the LR model, while the highest correlation (0.6800) was observed in the Radar Vegetation Index band. However, the lowest %RMSE (15.1245) was obtained with the RFR model with September data. When all datasets were considered, the LR model again provided the best results. When Sentinel-1 and Sentinel-2 data were used together, the lowest %RMSE (16.8766) was obtained with the RFR model in September. When all dates were considered collectively, the lowest RMSE (0.9389) and %RMSE (23.5720) were obtained with the RFR model. It was observed that data obtained during certain stages of rice growth yielded higher accuracy. Among the growth stages, the end of the ripening phase provided the best results, and accuracy increased when all dates were combined.

Keywords: Paddy-rice, Sentinel-1, Sentinel-2, Linear Regression, Random Forest Regression, Support Vector Regression, LAI Estimation

ÖZET

SENTETİK AÇIKLIKLI RADAR VE OPTİK VERİLER KULLANILARAK MAKİNE ÖĞRENMESİ MODELLERİYLE ÇELTİK BİTKİSİ YAPRAK ALAN İNDEKS (YAI) TAHMİNİ

Gülsüm POSTALLI

Yüksek Lisans, Geomatik Mühendisliği Bölümü

Tez Danışmanı: Prof. Dr. Saygın ABDİKAN

Ocak 2025, 66 sayfa

Tez kapsamında, Sentinel-1 Sentetik Açıklı Radar (SAR) ve Sentinel-2 optik uydularından alınan görüntüler ile Balıkesir ili Gönen ilçesinde bulunan çeltik bitkisi tarlalarından alınan Yaprak Alan İndeksi (YAİ) ölçümleri arasındaki ilişki incelenerek YAİ kestirimi hedeflenmiştir. Bu kapsamda Lineer Regresyon (LR), Rastgele Orman Regresyon (ROR) ve Destek Vektör Regresyon (DVR) makine öğrenmesi modelleri kullanılarak karşılaştırılmıştır. Proje kapsamında 2024 yılı Haziran, Temmuz, Ağustos ve Eylül aylarında olmak üzere 4 tarihte yerinde YAİ ölçer alet ile ölçümler yapılmıştır. YAİ ölçüm tarihlerine göre en güncel ve en uygun (bulutsuz) uydu görüntüleri tercih edilmiştir. YAİ kestiriminde her tarih ayrı ayrı ve tüm tarihler bir arada olmak üzere iki uygulama yapılmıştır. Optik verilerde tek tarih için Haziran 2024 ROR modelinde en düşük RMSE değeri (0.3236) verirken Eylül ayı %19.8574 ile en düşük %RMSE değerini vermiştir. Tüm tarihlerde ise en düşük RMSE 0.9797 ve %RMSE 24.5200 ile LR modelde ve en yüksek korelasyon B8A bandında 0,7473 olarak bulunmuştur. Sentinel-1 verilerinde tek tarih için Haziran 2024 LR modelinde en düşük RMSE değeri (0.3197) verirken tüm tarihlerde ise en düşük RMSE 1.2907 ile LR modelde ve en yüksek

korelasyon Radar Bitki İndeksi bandında 0.6800 olarak bulunmuştur. Ancak en düşük %RMSE değeri (15.1245) ile Eylül ayında ROR modeli ile elde edilmiştir. Tüm veriler ele alındığında ise yine LR modeli en iyi sonucu sunmuştur. Sentinel-1 ve sentinel-2 verileri beraber kullanıldığında ise en düşük %RMSE değeri=16.8766 ile yine Eylül ayı için ROR modelinden elde edilmiştir. Tüm tarihler ele alındığında ise en düşük hata RMSE=0.9389 ve %RMSE=23.5720 ile ROR modelinden elde edilmiştir. Pirinç bitkisinin gelişiminin bazı aşamalarında elde edilen verilerin daha yüksek doğruluk verdiği görülmüştür. Büyüme evreleri ele alındığında olgunlaşma sonu evresinde en iyi sonucu verirken tüm tarihler birleştirildiğinde doğrulukların arttığı görülmüştür.

Anahtar Kelimeler: Çeltik, Sentinel-1, Sentinel-2, Lineer Regresyon, Rastgele Orman Regresyon, Destek Vektör Regresyon, YAİ Tahmini

ACKNOWLEDGEMENTS

This thesis was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) within the scope of project no 222N088 and titled “Monitoring Rice Fields by Joint Use of Multi-temporal SAR and Optical Data for Yield and Growth Estimation”, 2502 – Research Projects - Bilateral Cooperation Program with the Bulgarian Academy of Sciences (BAS).

I would like to thank Prof. Dr. Saygın Abdikan, the principal investigator of the TÜBİTAK project and my thesis supervisor, who gave me all kinds of support and encouragement within the scope of my thesis.

I would like to thank Dr. Ömer Gökberk Narin, who helped me obtain the data for my thesis and informed me.

I would also like to express my gratitude to my family who supported me throughout my thesis and in every aspect of my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Symbols

dB	decibel backscatter coefficients
kg	kilograms

Abbreviations

LAI	Leaf Area Index
SAR	Synthetic Aperture Radar
RADAR	Radio Detection And Ranging
VIs	Vegetation Indices
NDVI	Normalized Difference Vegetation Index
NDVIRed	Red-Edge Normalized Difference Vegetation Index
LSWI	Land Surface Water Index
CLRE	Red-Edge Chlorophyll Index
EVI	Enhanced Vegetation Index
CLRE	Red-Edge Chlorophyll Index
SAVI	Soil-Adjusted Vegetation Index
EAVI	Enhanced Aquatic Vegetation Index
NDAVI	Normalized Difference Aquatic Vegetation Index
WAVI	Water Adjusted Vegetation Index
RMSE	Root Mean Squared Error

MSE	Mean Squared Error
MAE	Mean Absolute Error
VH	Vertical-Horizontal
VV	Vertical-Vertical
RVI	Radar Vegetation Index
ML	Machine Learning
LR	Linear Regression
RFR	Random Forest Regression
SVR	Support Vector Regression
LİDAR	Light Detection and Ranging
RGB	Red-Green-Blue
ANNR	Artificial Neural Network Regression
TSI	Turkish Statistical Institute
ESA	European Space Agency
GEE	Google Earth Engine
GRD	Ground Range Detected
SRTM	Shuttle Radar Topography Mission
IW	Interferometric Wide
PAR	Photosynthetically Active Radiation
SNAP	Sentinel Application Platform
LUT	Look-up Table

1 INTRODUCTION

Rice is a staple food for over half of the global population, and the most important production and consumption area of rice is the Asian continent. More than 90% of rice production takes place in this continent [1,2]. Rice production generally requires favorable climatic conditions and wetlands. It is mostly produced in floodplains and irrigated plains. In Turkey, the plains where the Meriç River floods are the main production areas [3]. Optimization is required for rice to be produced safely in the face of global demand and climate change conditions. For this reason, detailed monitoring of rice growth and health is essential [4].

Leaf Area Index (LAI), representing the total green leaf area on one side per unit of ground surface, is a key parameter for analyzing vegetation and ecological processes [5]. It is a biological indicator for the energy exchange processes between the atmosphere and the vegetation canopy, photosynthesis and transpiration processes. It is used for monitoring crop growth, plant and vegetation health. Monitoring the LAI in rice is crucial for understanding paddy-rice growth and yield estimation, and optimizing sustainable agricultural practices [6].

There are different methods used for LAI estimation in the literature. Destructive techniques and direct sampling can be given as examples of these traditional methods. These methods require a lot of labor, take a long time and do not provide a fast solution for large-scale applications, therefore they are not practical [7]. Remote sensing technologies offer an effective alternative for LAI estimation due to the ability to monitor the earth's surface at frequent intervals spatially and temporally [8]. Sentinel-1 and Sentinel-2 satellites provided by the European Space Agency are also used for agricultural field observations.

Sentinel-2 is an optical multispectral satellite that provides images in 13 spectral bands at different wavelengths [9]. NDVI and EVI indices derived from Sentinel-2 satellite data are widely used to estimate LAI [10].

Sentinel-1 is a Synthetic Aperture Radar (SAR) satellite, that provides backscatter data in VV and VH polarizations. It operates in C-band. Since it is an active satellite and is not affected by cloudy weather conditions, it is used to provide information about vegetation in agricultural monitoring [11]. The Radar Vegetation Index (RVI) used in vegetation analysis is calculated from the backscatter data of the Sentinel-1 SAR satellite.

Vuolo et al. (2013), presented that the use of Sentinel-2 optical data and Sentinel-1 SAR data together increased accuracy as it combined spectral information in vegetation analysis [12]. Vegetation indices can be used in agricultural analyses by taking different combinations of spectral bands. There are also a variety of indices used for LAI estimation. For LAI estimation in rice fields, Kulkarni and Honda (2020) found that NDVI is promising [13]. After NDVI, indices such as Red-Edge Chlorophyll Index (CLRE) and Land Surface Water Index (LSWI), which provide water stress and chlorophyll content information, can be used for LAI estimation [14,15].

Estimating the LAI in rice crops using remote sensing data from satellites has become a significant focus in precision agriculture. Machine learning algorithms can easily manage complex and nonlinear relations between the input parameters spectral bands and indices and output parameter LAI. While linear regression is a simple and easy-to-implement regression model, it may not adequately capture the complexity needed for accurate LAI estimation. Random Forest Regression (RFR) and Support Vector Regression (SVR) are more advanced machine learning models and have shown superior performance in various studies. They demonstrated their competence in processing high-dimensional data by using RFR and SVR models to map fruit trees and land use types using the Sentinel-2 satellite [16].

SVR is applied to construct a non-linear model by utilizing kernel functions. RFR is an ensemble learning method for machine learning in which multiple decision trees are constructed and outputs are combined. The outcome is the reduction of overfitting and capability to handle high-dimensional data. These characteristics may serve remote sensing applications.

1.1 Aim and Objectives of Thesis

This study aims to estimate LAI using Sentinel-1 SAR satellite data and Sentinel-2 optical satellite data from rice fields determined within the scope of the TÜBİTAK Project (No: 222N088) with machine learning models.

The main objectives of the thesis are as follows:

- To examine the relationship between LAI obtained from in-situ studies and LAI obtained from SAR and optical data
- To examine and compare the relationship between data obtained from different growth stages of rice plants and satellite data
- To evaluate the LAI estimation performance of VV and VH backscatter values, RVI, and the ratio of VH to VV polarizations obtained from Sentinel-1 SAR satellite
- To evaluate the LAI estimation performance of Sentinel-2 spectral bands (B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12) and vegetation index data (NDVI, NDVIred, LSWI, CLRE, EVI, EAVI, NDAVI and WAVI)
- To evaluate the LAI estimation performance of Sentinel-1 SAR data and Sentinel-2 optical data together
- To compare the performances of LR, SVR and RFR machine learning models for LAI estimation
- To use the synergy of SAR and optical data together with machine learning model predictions, selection/determination of the most relevant indices/trait importance for LAI estimation in paddy fields will help to find more accurate and scalable methodologies for crop monitoring and precision agriculture.

1.2 Thesis Structure

This thesis has 6 main chapters. In the first chapter, rice, and LAI, indices are introduced, the aim of the study is mentioned, and it consists of a literature review. The second chapter describes the study area and materials utilized in the research. The third chapter

introduces the methodologies applied. The fourth chapter presents the study's results for optical, SAR, and combined data. The fifth chapter includes a discussion of the findings, while the sixth and final chapter provides the overall conclusion and recommendations for future research.

1.3 Literature Review

LAI represents fifty percent of the total leaf area relative to each unit of surface area and is a critical parameter for understanding vegetation structure, ecological processes, and climate interactions. LAI estimation is an important area of research for understanding plant productivity and ecosystem dynamics [5]. The significance of LAI as a metric for assessing canopy structure and function is well established in the literature, with various methodologies developed to enhance its measurement accuracy and applicability across different vegetation types and environments. Three methods can calculate LAI; direct measurement method, indirect measurement method and remote sensing data inversion method with new technology [6].

Direct LAI estimation method is a destructive method calculated by harvesting the leaves of the plants or a more destructive technique called litter trap, calculated by taking the average of the annual leaf fall rate [17–19]. Because direct methods often require multiple repetitions, they tend to be time-consuming, demanding, and costly [18]. The drawbacks and challenges of direct methods lead to the use of indirect methods that are quick and simple.

With the developing technology, indirect methods started to be used very frequently since the 1990s [20]. Indirect LAI measurement is divided into 2 groups: contact and non-contact methods. In contact methods allometric techniques are used for forests or the method of inclined point is used for low canopies. Non-contact use of optical instruments such as the plant canopy analyzer and hemispherical photography. Optical instruments use radiation transfer theory [19,21].

There are studies in the literature on the use of hemispherical fisheye lenses for LAI estimation with hemispherical photography. There were also studies in the literature on the use of smartphones as an optical analyzer [22]. Zhu et. al. (2023) made an LAI

estimation with a smartphone equipped with a fisheye lens in their study. They compared the results of the study with digital hemispherical photography (DHP) for accuracy. A forest national park in China was selected as the study area. Six sample plots were selected, including three broadleaf forests and three mixed coniferous and broadleaf forests. Otsu threshold segmentation method was used for gap fraction. LAI was estimated by inverting the multi-angle gap fraction linearly. For mixed coniferous forest areas, RMSE values of 0.137, 0.120, and 0.147 were obtained, respectively. For broadleaf forest areas, RMSE values of 0.144, 0.135, and 0.137 were obtained. When compared with the DHP method, it gave compatible results. As a result, it was found that obtaining LAI with a smartphone with a fisheye camera is a fast and accurate method [23].

It is difficult to conduct in situ studies of large areas. For LAI estimation for these areas, aerial photographs taken from drones or airborne surveys measuring above the canopy, satellite imagery and laser scanning are used. LAI estimates made in large areas are promising in terms of accuracy. The techniques used are divided into 2 passive remote sensing and active remote sensing. In studies conducted with both techniques, vegetation indices (VI) were found to be correlated with LAI [24–27].

Yang and Fang (2020) conducted a literature study on the use of LiDAR technology in LAI estimation. LiDAR systems send a laser pulse and measure distance using the difference in time between the laser hitting an object and the reflection reaching the LiDAR system. There are three types of LiDAR systems operating on three platforms: ground-based terrestrial laser scanning, aircraft-based airborne laser scanning, and satellite-based space laser scanning. In this study, LAI measurement methods from these platforms are presented. Terrestrial laser scanning is effective for estimating LAI and vertical foliage profiles within plots, though it is affected by factors like clumping and occlusion. Airborne laser scanning covers larger areas but has limitations in describing within-canopy structures. Spaceborne laser scanning offers global LAI estimates, with accuracy influenced by footprint size and topography [26]

Some studies combine LiDAR technology and satellite images when making LAI estimates. Zixi Shi et al. (2023) used spectral data obtained from the Sentinel-2 satellite and LiDAR waveform together. LAI inversion method based on the GORT model was used. The LAI estimate using a physical model combining the two data obtained a

correlation of 0.81, an R^2 of 0.65, and an RMSE of 1.01. This has shown that it is better to work with combined data rather than working with single data [28].

Zhuo Lu et al. (2022) used UAV-based RGB images for LAI estimation in their study. A wheat field in China was selected as the study area. Canopy height model and canopy coverage methods were used. In-situ studies were also conducted for validation. Then, the performances of these methods were evaluated on RGB and multispectral images. As a result, R^2 = 0.742, RMSE = 0.534 was obtained from RGB images, and R^2 = 0.72, RMSE = 0.556 from multispectral images. This shows that the methods used in the study increase accuracy compared to traditional methods [29].

When using satellites for LAI estimation, we consider sensor features according to the scope of our study. Systems with high (Sentinel 1/2), medium (Landsat, MODIS) or coarse spatial resolution (Advanced Very High-Resolution Radiometer sensors (~1000m)) can be used [30].

Various methods have been used in the literature for LAI estimation from satellite images. Some empirical models use regression equations to relate spectral indices to LAI [31,32], physical models that use radiative transfer theory to simulate canopy reflectance, data assimilation that integrates LAI with crop values for better yield and biomass estimation [33], machine learning models, or hybrid methods [34,35] where these methods are used together.

Recent advancements in machine learning have increased the ability to estimate LAI from satellite imagery, providing more precise and efficient methods. Numerous machine learning algorithms such as RFR, SVR, Artificial Neural Network Regression [36,37], XGBoost and PROSAIL-D Model [35], Bi-LSTM and LSTM Models [38,39] have been utilized in literature to derive LAI from remote sensing data.

Kaplan and Rozenstein (2021) aimed to measure the performance of the Sentinel-2 satellite bands in estimating LAI in cotton, wheat and tomato fields and to determine the best vegetation indices created using these bands. A linear regression model was used to

estimate LAI from the Sentinel-2 satellite by validating with the LAIs obtained from the field. 8 VIs (NDVI8A, NDVI, WDVI, MTCI, EVI, MSAVI, SAVI, DVI) were used and according to the best-performing bands, 2 new VIs were created as WEVI (Water vapor red Edge Vegetation Index) and WNEVI (Water vapor narrow NIR red Edge Vegetation Index). B7, B8, B8A, and B9 gave the best performance in LAI modeling. It has the lowest RMSE value with reNDVI = 0.5. The newly produced WEVI and WNEVI indices outperformed NDVI for all three vegetation areas [40].

Yadav et al. (2019) compared the performances of the different machine learning regression models to estimate LAI using Landsat-8 OLI satellite data. The models used in the study are Artificial neural network regression, SVR, and RFR. In-situ data were collected with an LAI-2200C plant canopy analyzer on 3 different dates at different stages of crops between January and March 2017 and satellite data were used on the same dates. Using NDVI and actual LAI values, the data was split into 60% training and 40% testing, and separate models were trained. The highest accuracy was obtained from RFR with RMSE = 0.404. RMSE = 0.478 was obtained from the SVR model and RMSE = 0.497 from the ANNR model [36].

Obtaining remote sensing LAI from rice fields is more challenging than other crops due to the submerged during the growth stage, which affects the spectral data [41]. Various machine learning techniques have been applied in previous studies to determine LAI from satellite images in rice fields. Among these approaches, SVM and RF have shown strong performance. They effectively use multi-source data. Thus, the prediction accuracy is increased. The data used can be texture features, spectral reflectance and vegetation indices. The RF model has stood out in integrating the data and has made high accuracy LAI predictions [42–44].

In the study conducted by Hosseini et al. (2021), LAI was estimated from rice fields from RADARSAT images. The SVM model was used to estimate LAI. In situ data were collected from the study areas of India using Hemispherical Photos. Images were taken from the RADARSAT-2 satellite using Fine Quad-Polarization Wide (FQW) and

Standard Dual-Polarization (SDP) modes on 4 dates in 2014 and 3 dates in 2018. A high correlation was obtained with RMSE=0.41 [45].

In literature, several machine learning regression models have been evaluated for their effectiveness in estimating LAI from Sentinel-1 satellite and Sentinel-2 satellite data. Models such as SVM, k-Nearest Neighbour, RF, and Gradient Boosting Decision Tree have been tested, with tree-based algorithms like RF and Gradient Boosting Decision (GBDT) Tree showing superior performance in handling data dimensionality and providing accurate LAI estimates [43,46,47] Gaussian Process Regression has been shown to achieve high accuracy in LAI retrieval when combined with radiation transfer models [46,48].

Studies using Sentinel-1 SAR and Sentinel-2 optical data together in machine learning models and deep learning models are effective in LAI estimation with high accuracy [49,50]. The combination of VH and VV polarizations from Sentinel-1, along with optical data from Sentinel-2, has been found to provide more accurate LAI estimates than using either data source alone [51,52].

In the thesis study by Najatishendi (2017), Radar and optical satellite images were used to estimate LAI from 6 paddy-rice fields in the Ipsala district. In-situ LAI measurements were taken with a CI-110 measuring device on 28 August 2015 during the reproductive stage of rice. Sentinel-1 on 27 August 2015 and TerraSAR-X satellite images on 4 September 2015 were used as SAR data. The Water Cloud Model was used for LAI estimation from SAR images. The genetic algorithm was used as the optimization technique. Landsat-8 satellite was used for optical imagery on 28 August 2015. NDVI, Green Normalized Difference Vegetation Index (GNDVI), Modified Normalized Difference Water Index (NDWI), Simple Ratio Index (SR), Gitelson and Merzlyak Index (GMI) and Zygielbaum Water Stress Index (ZWSI) were calculated. Correlation analysis was performed between LAI and the indices. As a result of the study, it was found that LAI values and Landsat-8 satellite data have a higher correlation than Sentinel-1 and TerraSAR-X satellite data [53].

2 STUDY AREA & MATERIALS

This section provides information about the study area and the materials used in the study. This study was carried out with 60 samples taken from the parcels where the rice fields of the Gönen district of Balıkesir province are located. Within the scope of the study, ground observation data were collected from 60 points with in-situ studies on 4 different dates. Sentinel-1 and Sentinel-2 satellite images were downloaded on the closest and most suitable dates for the in-situ studies. VV, VH, ratio and RVI values were obtained from the Sentinel-1 image, and 10 bands with 9 index values (NDVI, NDVIRed, LSWI, CLRE, EVI, SAVI, EAVI, NDAVI, WAVI) were obtained from the Sentinel-2 satellite image. In this section, the study area, rice plant and satellite images are explained in detail.

2.1 Study Area

The study area was chosen from the rice fields of Balıkesir province of the Gönen district. Gönen is located between 40°06' north latitude and 27°38' east longitude, and has a total area of 1152 km². Gönen region is in the northwest of Türkiye, southwest of the Marmara Sea, and is a district on the coast of the Marmara Sea (Figure 2.1). The Gönen has a mild Mediterranean climate.

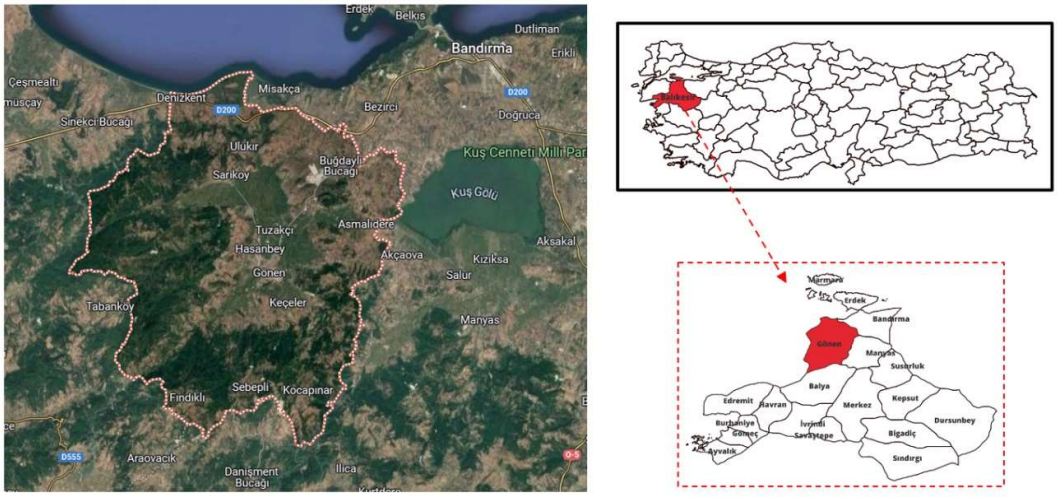


Figure 2.1. Balıkesir Province Gönen District

The region features a diverse range of landscapes, including urban, water areas, agriculture areas and forests. The central and north-eastern parts of Gönen land are covered with plains, and the western and south-eastern parts are covered with hilly and wavy areas. The Gönen Plain is in the central part. The annual average temperature in Gönen is 14°C. August is the hottest month, with temperatures ranging from 19°C to 29°C, while January is the coldest, with temperatures between 1°C and 9°C. The average temperatures of the Gönen district by month are shown in Figure 2.2 [54].

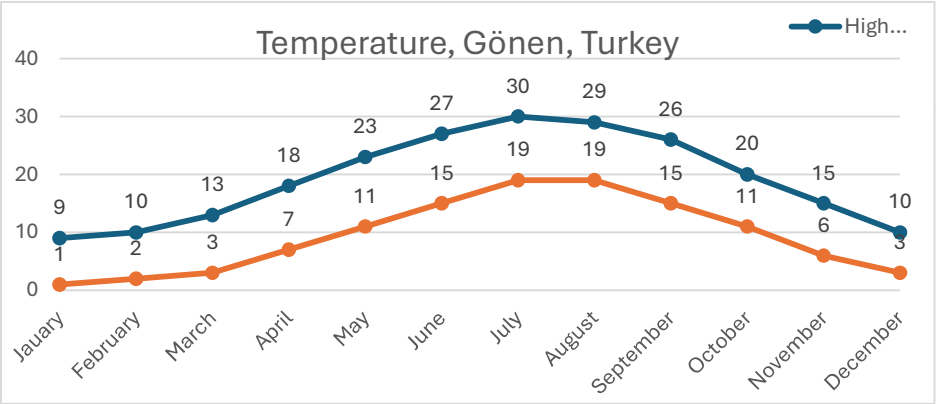


Figure 2.2. Average Temperatures of Gönen District

The life cycle of rice cultivars in Balıkesir-Gönen ranges from 110 to 130 days from germination to maturity. Accordingly, rice sowing time in Gönen is around the end of April and the beginning of May, and the harvest is around the end of September and the beginning of October. The crop growing season in Gönen is shown in Figure 2.3. [54].

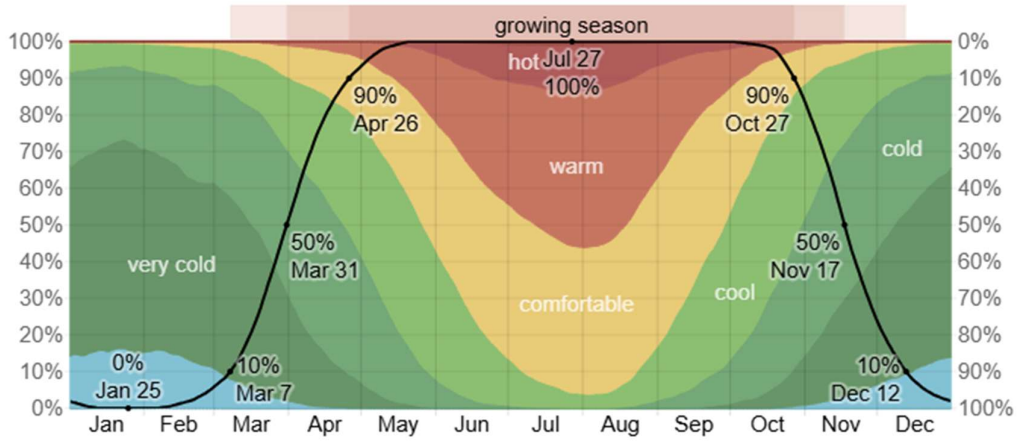


Figure 2.3. Crop Growing Season of Gönen District

Google Earth images of 60 points which were collected from rice fields selected for the study in 4 different regions (red circle) are shown in Figure 2.4.

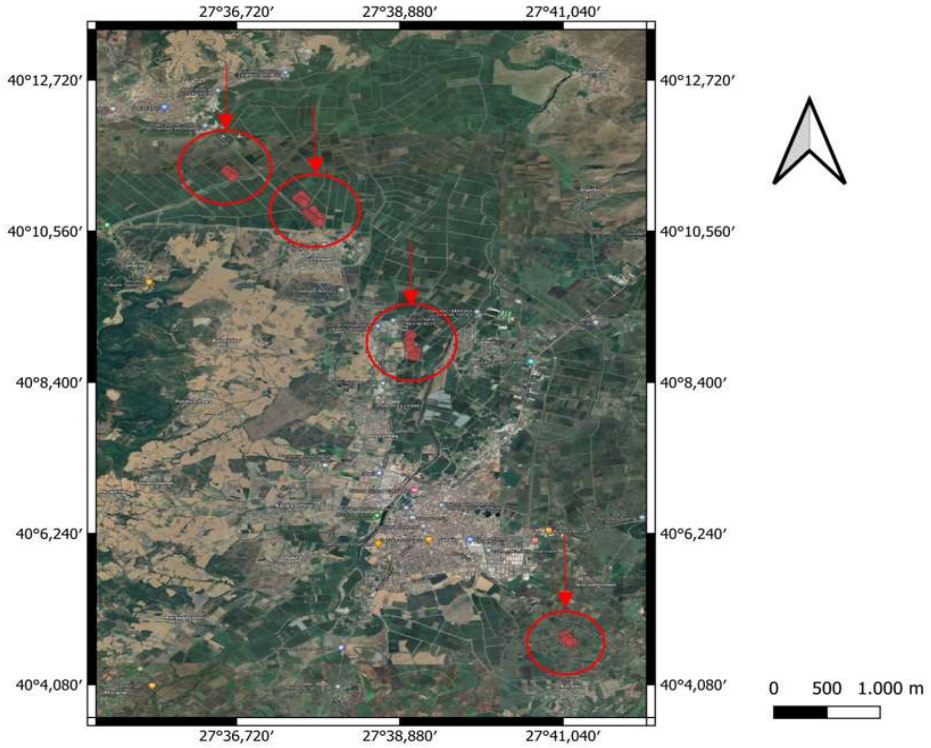


Figure 2.4. Google Earth View of 60 Points of The Paddy Field

2.2 Rice (*Oryza sativa L.*)

Rice (*Oryza sativa L.*), one of the oldest cultivated plants and a member of the Poaceae family, is one of the plants that feed more than 50% of people. Rice is rich in amino acids since it contains 5-10% protein. These amino acids are essential for nutrition. It is one of the basic food sources since it meets most of the nutrient and energy needs of people. It is estimated that rice production in the world will reach 523.9 million tons in 2023/24 [55,56]. It is an annual herbaceous plant and lives in semi-irrigated areas. Since it requires a minimum of 1000 - 1200 mm of rainfall during the growing season, if this amount is not provided by rainfall, technical irrigation methods are used to meet the water need [57].

Most of the rice produced worldwide is grown in tropical and subtropical climate zones. The countries that produce the most rice in the world are Asian countries. China is in first place in paddy production with 144.6 million tons, followed by India with 137.8 million tons and then Bangladesh. The country that exports the most rice is India with 16 million tons. The 10-year production quantities of the top rice-producing countries are given in Table 2.1 [58,59].

Table 2.1. The Most Rice Producing Countries and Annual Production Amounts(tons) in the Last 10 Years

	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
China	146.7	148.5	147.8	148.9	148.5	146.7	148.3	149.0	145.9	144.6
India	105.5	104.4	109.7	112.8	116.5	118.9	124.4	129.5	135.8	137.8
Bangladesh	34.5	34.5	34.6	32.6	34.9	35.9	34.6	35.9	36.4	37.0
Indonesia	35.6	36.2	36.9	37.0	34.2	34.7	34.5	34.4	33.9	33.0
Vietnam	28.2	15.8	27.4	27.7	27.7	27.1	27.4	26.7	27.1	26.3

Türkiye ranks 43rd in the world in paddy rice production in 2023. The highest production was made in the Marmara region with a rate of 65.0286%. Paddy cultivation in Türkiye was done in 28 provinces in 2023. The provinces with the most rice production in 2023

are Edirne, Samsun, Balıkesir, Çanakkale, Çorum and Çankırı [60]. These provinces and their percentage of production quantities are shown in Figure 2.5 and Figure 2.6.

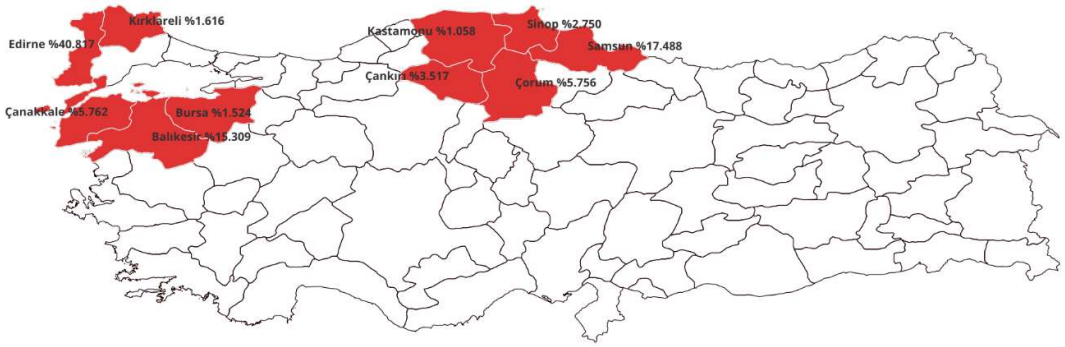


Figure 2.5. Top 10 Provinces Producing Rice in 2023 and Production Percentage

According to the last 5 years of TSI data, the rice area planted and harvested in Gönen and the amount of rice production is given in Table 2.2 [60].

Table 2.2. Gönen Rice Data of the Last 5 Years

Year	Planted area(decare)	Harvested Area(decare)	Production Amount(tons)	Yield Amount (kg/decare)
2019	84,500	84,500	65,010	769
2020	85,000	85,000	67,832	798
2021	88,000	88,000	74,447	846
2022	87,000	87,000	70,400	809
2023	87,000	87,000	75,621	869

In Türkiye, rice cultivation is carried out subject to permission within the framework of the "Paddy Planting Law No. 3039".

Rice consists of 4 main parts; Paddy, root, spear, flower, and grain.

The growth of the rice plant may be divided into three phases:

1) The vegetative stage spans from the moment of germination until the initiation of the panicle.

2)The reproductive phase encompasses the period from panicle initiation to panicle.

3) The stage of development that occurs from the time of flowering until the fruits reach full maturity [1].

Growing Stages of Rice is shown in Figure 2.7.

Variable	35 days	30 days
Vegetative Phase	Reproduction Phase	Ripening Phase

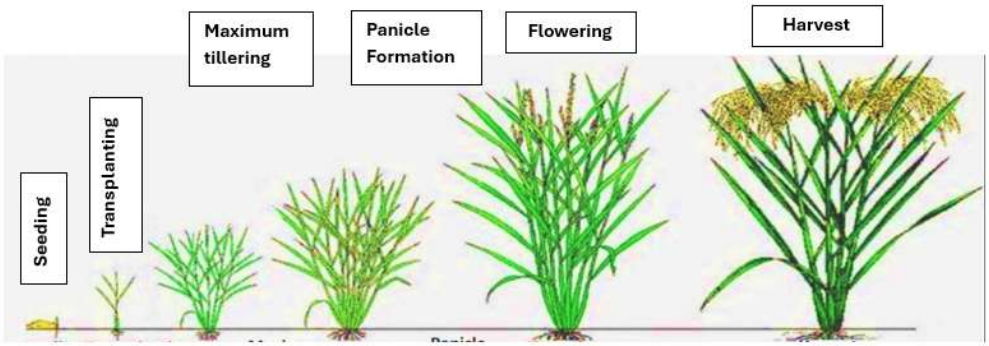


Figure 2.6. Growing Stages of Rice

2.3 Optical Remote Sensing Data for LAI Estimations: SENTINEL-2

The Sentinel-2 satellites provide high-resolution, wide-swath, multi-spectral optical images developed by the European Space Agency (ESA). The Copernicus Sentinel-2 target is two satellites in the same orbit, but phase 180° apart. Sentinel-2A was launched into space in 2015 as part of ESA's Copernicus programme and then Sentinel-2B was launched into space in 2017. The images of these satellites are freely available. The revisit time, which was 10 days over the equator with a single satellite, was reduced to 5 days with the use of two satellites. This contributes to a more detailed observation of the changes on Earth. Optical sensors provide 13 spectral bands with different spatial resolutions. The 4 of these bands provide 10 m, 6 bands provide 20 m, and 3 provide 60

m spatial resolution. The orbital swath width equals 290 km. A new satellite, Sentinel-2C, was launched on September 5, 2024. The band characteristics of Sentinel-2 A and B satellites are given in Table 2.3.

Table 2.3. Spectral Properties of Sentinel 2A and Sentinel 2B Satellites

Band number	S2A			S2B			Description
	Resolution (m)	Central wavelength (nm)	Bandwidth (nm)	Resolution (m)	Central wavelength (nm)	Bandwidth (nm)	
1	60	442.7	27	60	442.2	45	Aerosols
2	10	496.6	98	10	492.1	98	Blue
3	10	560.0	45	10	559	46	Green
4	10	664.5	38	10	665	39	Red
5	20	703.9	19	20	703.8	20	Red Edge 1
6	20	740.2	18	20	739.1	18	Red Edge 2
7	20	782.5	28	20	779.7	28	Red Edge 3
8	10	835.1	145	20	833	133	NIR
8A	20	864.8	33	20	864	32	Vegetation Red Edge
9	60	945.1	26	60	943.2	27	Water vapor
10	60	1373.5	75	60	1376.9	76	SWIR Cirrus
11	20	1613.7	143	20	1610.4	141	SWIR 1
12	20	2202.4	242	20	2185.7	238	SWIR 2

In this study, in-situ data was taken 4 times. The most suitable and most recent 4 Sentinel-2 SR harmonized images for this field study were downloaded and used via Google Earth Engine. Selected optical image dates are shown in Table 2.4.

Table 2.4. Used Sentinel Satellites and Image Acquisition Dates

Satellite name	Image Date
Sentinel-2A	25 Jun 2024
Sentinel-2A	15 July 2024
Sentinel-2B	19 August 2024
Sentinel-2A	23 September 2024

Google Earth Engine (GEE) is a platform that allows us to access the catalogs of several datasets such as Sentinel – MODIS – Landsat satellites data and to perform operations on these images. It can be used for geospatial or image processing algorithms. It can access Sentinel-2 satellite data on GEE from March 28, 2017 to the present. In GEE, Level-2A, Sentinel-2 data is presented after the atmospheric corrections. Bottom-Atmosphere reflectance is made and the atmospheric effects (clouds, water vapor, aerosols, etc.) are removed. It is used for the calculation of Surface Reflectance, which shows the amount of light reflected by the soil surface by remote sensing. No image pre-processing steps are required for these images and were not performed in the study.

Within the scope of this study, 10 bands of Sentinel-2 satellite were used. These bands are B2, B3, B4, B5, B6, B7, B8, B8A, B11 and B12. And then 9 indices were calculated from GEE. These indices and their formulas are shown in Table 2.5.

Table 2.5 Indices and Their Formulas

Index Name	**Sentinel-2 Band Equations**
NDVI (Normalized Difference Vegetation Index)	(B8-B4)/(B8+B4)
NDVIRed (Red-Edge Normalized Difference Vegetation Index)	(B8-B5)/(B8+B5)
LSWI (Land Surface Water Index)	(B8-B10)/(B8+B10)
CLRE (Red-Edge Chlorophyll Index)	(B8/B5)-1
EVI (Enhanced Vegetation Index)	2.5*((B8-B4)/(B8+6*B4-7.5*B2+1))
SAVI (Soil-Adjusted Vegetation Index)	((B8-B4)/(B8+B4+0,5))*(1+0,5)
EAVI (Enhanced Aquatic Vegetation Index)	(B11 + B12 - (B3 – B2)) * ((1610 - 490) / (1610 -560)
NDAVI (Normalized Difference Aquatic Vegetation Index)	(B8-B2)/(B8+B2)
WAVI (Water Adjusted Vegetation Index)	((1 + 0.5) * (B8 – B2)) / (B8 + B2 + 0.5)

In remote sensing, indices are used to analyze environmental parameters derived from the data. NDVI [61] is used to measure plant health. NDVIRed [62] measures plant health more precisely by using the red-edge band. LSWI [63] is used to assess soil moisture and water content on the land surface. CLRE [64] measures the chlorophyll content of plants, allowing for the monitoring of photosynthetic capacity. EVI [65] is used to measure vegetation density and is resistant to atmospheric distortions. SAVI [66] is used to reduce the soil effect in areas with low vegetation density. EAVI [67] is used to monitor the vegetation under water, thus understanding the plant health in wetlands. NDAVI [68] and WAVI [69] are used in areas where the vegetation is composed of water. There is more water than soil in these areas during the first stages of the paddy rice. These indices have a background adjustment factor.

2.4 Radar Remote Sensing Data for LAI Estimations: SENTINEL-1

2.4.1 Synthetic Aperture Radar (SAR)

SAR is an instrument of the Radio Detection And Ranging. It works in the microwave bands of the electromagnetic spectrum and thus high-resolution images can be obtained. SAR uses radar signals, these signals hit objects on the ground, and the reflected signals are collected and processed by SAR antennas. It is an active sensor because it uses radar signals and is independent of sunlight. Since it is an active sensor, it can measure day and night and is not affected by weather conditions.

SAR's fundamental measurements are recorded in specific time bins along the azimuth (sensor's flight direction) and range (perpendicular to the flight path). The key parameters captured include the intensity (amplitude) and phase of the backscattered signal. These measurements are essential for constructing high-resolution radar images and extracting surface information [70]. There are different wavelength bands such as X, C, L, and P [71] (Table 2.6).

Table 2.6. The RADAR Bands and Specifications

Band	Wavelength(cm)	Frequency (GHz)
Ka	1.1–0.8	27–40
K	1.7–1.1	18–27
Ku	2.4–1.7	12–18
X	3.8–2.4	8–12
C	7.5–3.8	4–8
S	15–7.5	2–4
L	30–15	1–2
P	100–30	0.3–1

2.4.2 Sentinel-1 Satellites

The Sentinel-1 satellite constellation was developed under ESA's Copernicus programme. It consists of twin satellites carrying C-band SAR. Sentinel-1A was placed into orbit on 3 April 2014. Sentinel-1B was placed into orbit on 25 April 2016. Since they are 180 degrees in orbit apart, the two satellites have reduced the imaging time from Earth, which is normally 12 days, to 6 days. They have a SAR instrument; therefore they are not affected by weather conditions and can perform imaging regardless of day or night.

Sentinel-1B's mission ended in 2022 due to the SAR sensor shutdown. Sentinel-1C has started its mission on November 4, 2024. Sentinel-1D satellite is also expected to be launched [72]. Sentinel-1 satellite images are used for land, sea, and glacier monitoring, natural disaster mapping and specifically ship detection and tracking. The Sentinel-1 satellite can image in different modes, operate in different bands and produce different types of data. All Sentinel-1A satellite data from 3 October 2014 to the present are freely available on Copernicus Data Space Ecosystem. In this study, the 4 Sentinel-1 images closest to the in-situ studies were used. Selected SAR image dates are shown in Table 2.7.

Table 2.7. Selected SAR Image Dates

Satellite name	Image Date	Orbit
Sentinel-1A	23 Jun 2024	Ascending
Sentinel-1A	17 July 2024	Ascending
Sentinel-1A	22 August 2024	Ascending
Sentinel-1A	15 September 2024	Ascending

Sentinel-1 GRDH images include Ground Range Detected (GRD) data, calibrated, and ortho-corrected products. It can provide different bands, resolutions and instrument modes. Instrument modes are IW (Interferometric Wide Swath), EW (Extra Wide Swath) and SM (Stripmap). Resolutions are 10, 20x22 or 50 meters. Bands are VV, VH, HH, and HV polarizations.

Sentinel-1 images with Interferometric Wide (IW) mode used in the study were obtained via Copernicus and VH and, VV polarizations were used for each date. Interferometric Wide mode has a spatial resolution of 5 x 20 m, swath width of 250 km, and off-nadir angle 29.1° to 46. The Radar Vegetation Index (RVI) and Ratio value (VH/VV) were calculated for each date.

$$RVI = \frac{4\sigma_{VH}^{0}}{\sigma_{VV}^{0} + \sigma_{VH}^{0}} \quad (1)$$

where σ_{VH}^{0}, σ_{VV}^{0} and σ_{VH}^{0} are the polarized backscattering coefficients [73].

RVI is an index through which we provide information about the condition and growth of the plant. We can follow the plant health and growth stages via RVI.

2.5 In-situ LAI Measurements

Within the scope of the study, in-situ studies were carried out on 4 dates during the growing season of rice plants in Gönen where the study area. During these visits, on-site LAI values were obtained using the AccuPAR PAR/LAI Ceptometer Model LP-80 device.

When the AccuPAR LP-80 device is turned on, the first real-time PAR data appears. This device uses PAR and other device parameters to calculate LAI. It calculates LAI based on PAR measurements above and below the canopy along with the zenith angle. Data recorded using the AccuPAR LP-80 device can be downloaded to the computer using Windows Hyperterminal software [74].

Table 2.8. Dates of In-situ Studies

Field Dates	**Condition of Paddy-Rice**
25 Jun 2024	Tillering vegetative phase
17 July 2024	Flowering/grain formation
23 August 2024	Ripening, filling stage
19 September 2024	End of ripening, mature grain

In field studies, the above and below-canopy PAR values of the plant were measured. 10 data at 10 cm intervals in a 1 x 1 meter area were recorded and an average of these data were taken.

3 METHODOLOGY

In this section, the process steps and methods used in the study are explained. The general process steps are shown in Figure 3.1 with the workflow. The software and machine learning models used in the study are also explained respectively.

Figure 3.1 shows the general steps of the methodology used in this study to estimate LAI from optical and SAR images with machine learning models using GEE and Python programming language and to evaluate the performance of the machine learning models.

In the first step, Sentinel-2_SR_Harmonized image collection is imported from the GEE platform. After applying a cloud mask to the images, they were filtered according to the points used in the in-site studies and the closest dates to these studies. The indices were calculated by taking the bands to be used in the study. The image dates and the indices used were introduced in the previous section.

In the second step, Sentinel-1_GRD images were downloaded from the Copernicus Data Space Ecosystem website according to the dates closest to the in-site studies. Pre-processing steps were applied to Sentinel-1 images in the SNAP software to obtain better results within the scope of the thesis. The images were filtered to the points used in the in-site studies. As a result, VV, VH, Ratio (VH/VV) and RVI values were obtained.

Actual LAI values were obtained by going to the field and taking points on 4 dates.

These values are recorded in a csv file and saved in an Excel file together with the LAI values from in-situ studies. Code is written using the Python coding language in the Jupyter notebook working environment. The Excel file created in the previous step is given to the program as input. Linear Regression, Support Vector Regression and Random Forest Regression Machine learning models were run for 3 different data. Firstly, LAI estimation is made with the values obtained only from Sentinel-1 satellite images, secondly, LAI estimation is made with the values obtained only from Sentinel-1 satellite images, and finally, LAI estimation is made by combining the values obtained from Sentinel-1 and Sentinel-2 satellite images. The performance of each machine learning model is evaluated. In this step, RMSE, MSE, MAE, R^2 and percentage RMSE values are calculated. According to the calculated parameters, it is determined which model is better for LAI estimation.

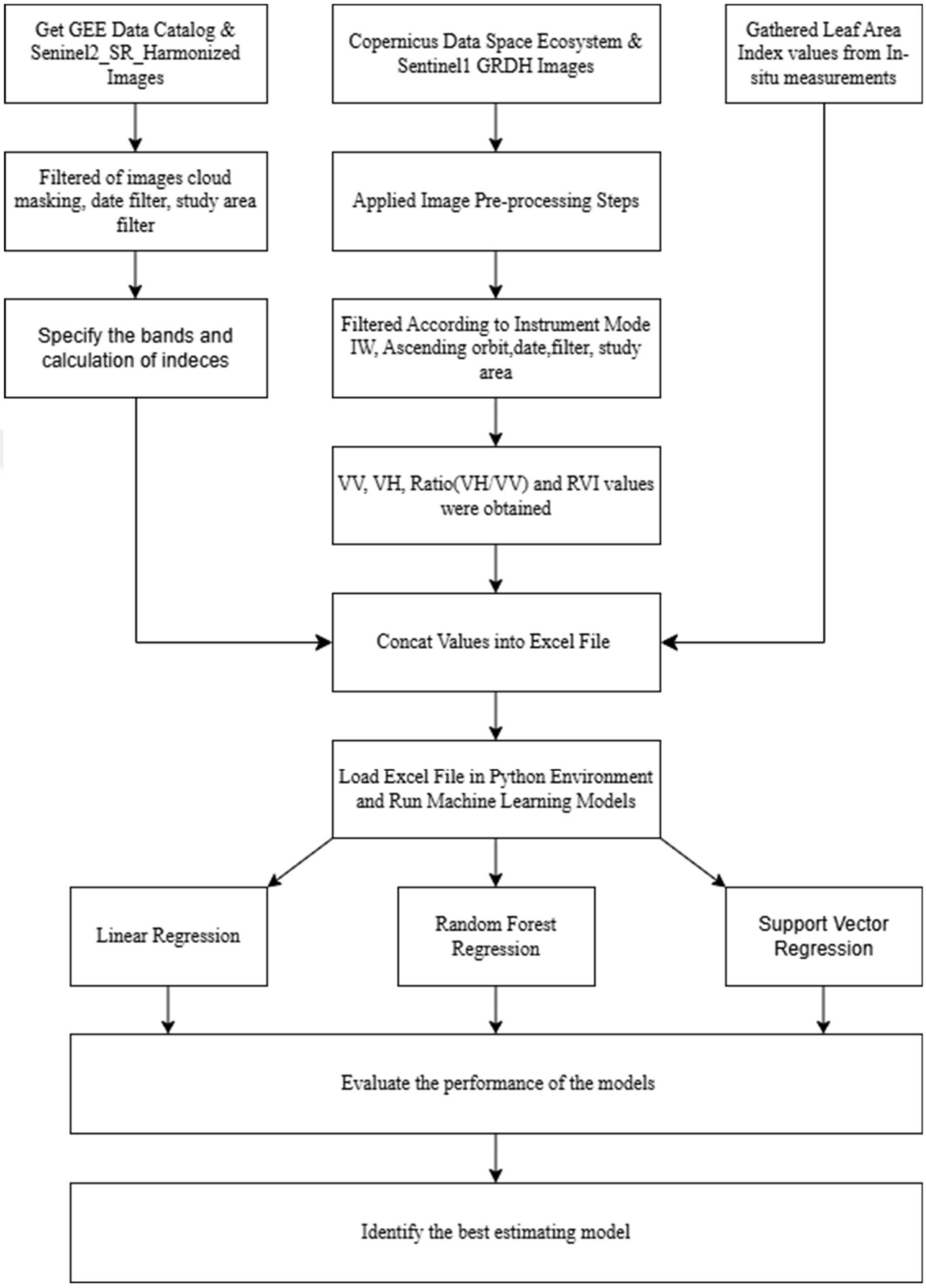


Figure 3.1. Workflow of the Methodology

3.1 Software Used

In this study, GEE was used to download Sentinel-1 and Sentinel-2 satellite images, perform image editing, extract bands, calculate of index and save values as a CSV. It is a technology developed by Google in 2010 to monitor the Earth. Cloud computing combines earth observation data and data from many satellites, allowing large-scale data to be analyzed through algorithms. In this way, high-resolution, accurate information can be obtained. It offers users approximately 50 years of data. In this way, it provides the opportunity to monitor changes on the earth. By using and analyzing the datasets provided by the GEE, studies can be conducted on topics such as global warming, climate change, and sustainable living, which have recently become a hot topic [75,76].

GEE provides several datasets, such as Landsat, Sentinel-1 SAR and Sentinel-2 optical and MODIS image catalogs. It serves land and sea surface temperature, climate, atmospheric and weather data from various sensors. It also contains elevation data [77]. There are machine learning models within GEE. Regression, image classification and accuracy analysis can be done. Thanks to the GEE code editor, we can write and run scripts over the web using the JavaScript-API. Codes can also be written with Python and JavaScript editors using the API [78]. Machine learning models were trained in Jupyter notebook, a web-based idea of the Python programming language, to estimate LAI from data received from Sentinel-1 and Sentinel-2 satellites.

3.1.1 Sentinel Application Platform (SNAP)

The Sentinel Application Platform is free software developed by Sensar, Skywatch, Brockmann Consult and C-S for Earth observation processing and analysis. It also supports sensors other than Sentinel. It is a desktop remote sensing and image processing software. It includes image processing, modeling, and visualization. It can be used for both vector and raster data. It can be used for radar data as well as optical data [79].

3.1.1.1 Sentinel-1 GRDH Pre-processing Steps in SNAP

In this study, pre-processing steps were applied to Sentinel-1 satellite images before using them for analysis. Sentinel-1 GRDH data were downloaded from the Copernicus Data Space Ecosystem site and image preprocessing steps were applied on SNAP Software. The workflow is designed to perform essential corrections, including precise orbit files, thermal noise removal, border noise removal, radiometric calibration, and Range Doppler and terrain corrections. The preprocessing steps are shown in Figure 3.2.

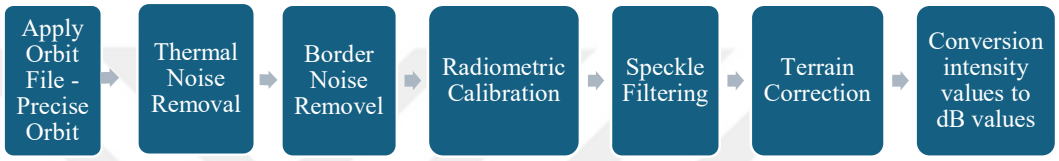


Figure 3.2. Pre-processing Steps of Sentinel-1 Data

- Apply Orbit File -Precise Orbit

Sentinel-1 may contain small errors as it uses real-time orbital information during imaging. These errors are corrected by applying correction files containing precise orbital information published by ESA later. To obtain the precise orbit, the SNAP software enables the orbital state vectors to be automatically downloaded and updated for each SAR scene in the product metadata.

- Thermal Noise Removal

Thermal noise may occur in low reflectance areas in Sentinel-1 GRD images. In order to prevent this noise from causing incorrect results in the analysis, thermal noise is eliminated by correcting the brightness changes in the image. Removal process Radar → Radiometric → S1 thermal noise removal on SNAP Software.

- Radiometric Calibration

Sentinel-1 radiometric calibration is applied to convert pixel values in Sentinel-1 SAR data into physically meaningful radar backscatter coefficients (sigma naught, gamma naught, beta naught). Sentinel-1 Level-1 data does not contain radiometric calibration,

so radiometric calibration is required. This calibration ensures comparable and consistent measurements between different images.

- Speckle Filtering

This step is optional. In SAR systems like Sentinel-1, RADAR images contain noise called speckles. Filters such as Lee Sigma, Refined Lee, and Forest are used to reduce noise and make the image clear. These filters should be selected carefully to avoid loss of detail.

The Lee filter, an adaptive filtering technique, is commonly utilized in the preprocessing of Sentinel-1 SAR GRD data to minimize speckle noise. It effectively maintains linear features, point targets, edges, and texture details. In this study, the lee filter was preferred.

- Terrain Correction

There are distortions related to the side-looking geometry in the SAR images because the images are taken at angles greater than 0 degrees. The reason for terrain correction is to want the image to be geometrically close to the Earth.

In this study, range-doppler terrain correction has been applied. In the terrain correction operation step, the 30-meter SRTM digital elevation model was used.

The Doppler Terrain model uses a digital elevation model to correct the position of each pixel.

- Conversion intensity values to dB values

The results of Terrain Correction images have intensity values. The back scattering coefficients' digital numbers from the result of terrain correction are unitless. Convert these to decibel backscatter coefficients (dB) using the formula below. This formula was applied to both VH polarization and VV polarization using band math in SNAP software [80].

$$\beta^{\circ}_{dB} = 10 \times \log_{10}(\beta^{\circ}) \quad (2)$$

3.2 Machine Learning

In recent years, with the increasing technology, the use of Artificial Intelligence has increased. One of the branches of artificial intelligence is machine learning. In recent years, machine learning has been extensively applied across various fields and analyses due to its capability to process and interpret large datasets [81]. The main types of machine learning are shown in Figure 3.3 [82].

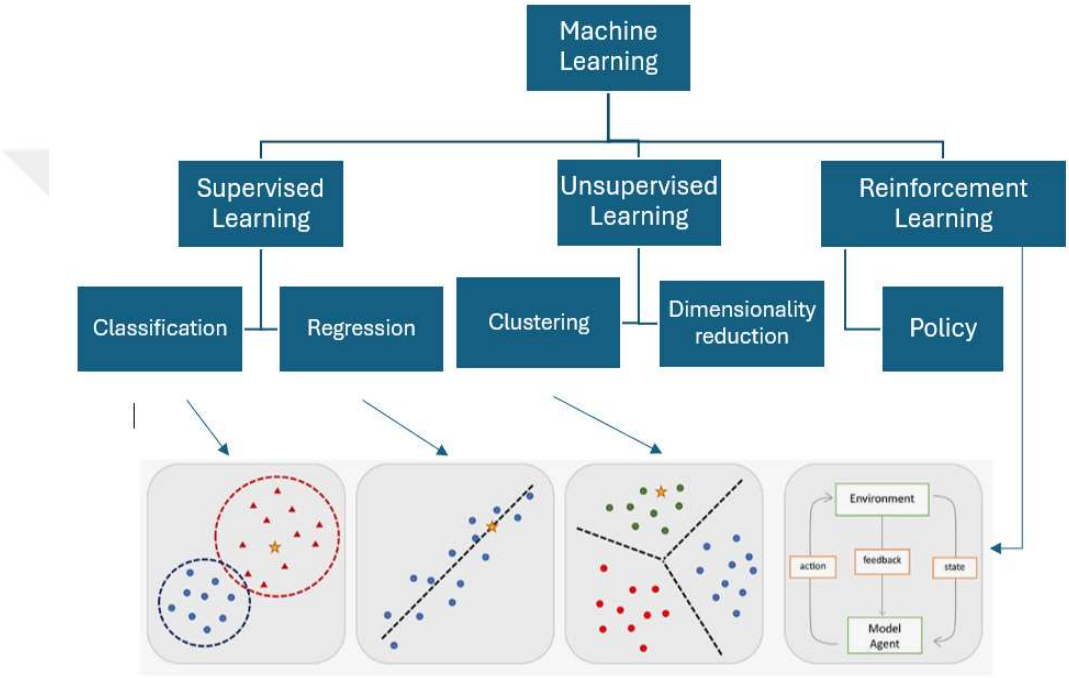


Figure 3.3. Diagram of Machine Learning Algorithms

In supervised learning, the model learns from labeled data. In unsupervised learning, it is trained using unlabeled data. In reinforcement learning, the model interacts with its environment, takes actions, and receives feedback along with status updates.

Machine learning is used in every area of our lives, such as health, economy, transportation, finance and banking, shopping, security, and agriculture. Machine learning models are also used in the analysis of remote sensing and satellite images. High accuracy results are obtained in various fields such as object detection, land cover/use classification, change detection, earth observation, environmental monitoring and disaster management using remote sensing images [83,84].

3.2.1 Supervised Classification

In supervised learning, the model learns to classify and predict outputs from a set of labeled inputs. It learns the relationship between input and output and then uses this for the input data [85]. Training and prediction are the steps of supervised learning. As an example of supervised learning, when classification is desired from remote sensing images, the algorithm is given a training set labeled as forest, city, water, vegetation area, etc. The model learns the relationship between the labeled data and the outputs and classifies the outputs into these classes. Using this relationship, the model makes predictions for new unseen data to be given in the future.

3.2.1.1 Regression

Regression, a statistical method and a form of supervised learning is utilized to analyze the relationship between a dependent variable and one or more independent variables [86]. It is used to predict continuous data. Regression models used in machine learning: Simple and Multiple Linear Regression, Polynomial Regression, Decision Trees Regression, Support Vector Regression and Ensemble Method. In this study, Linear Regression, Support Vector Regression and Random Forest Regression models were used. After completing the model study, the performance of regression algorithms is evaluated. In this evaluation, metrics such as MSE, RMSE, MAE and $R2$ can be used [87].

3.2.1.2 Linear Regression Model

One type of supervised learning is regression. The purpose of the linear regression model is to find the linear line that provides the best fit between the variables which are independent and dependent [88].

Linear Regression is divided into 2 groups simple linear regression or multiple regression. The simple linear regression equation is given below, where y is the predicted value of the dependent variable, x is the independent variable, a_0 is the point where the Y-axis intercepts, a_1 is the slope of the x variable, and e is the error term [86].

$$y = a_0 + a_1 x + e \quad (3)$$

3.2.1.3 Random Forest Regression

The Random Forest algorithm was introduced by Breiman(2001) and as the name suggests, it is a tree-based learning algorithm. It can be used for classification and regression in supervised algorithms as well as in unsupervised algorithms. Random forest creates multiple decision trees. Random forest combines the trees it creates into a forest to make better predictions [89]. It is a collection of decision trees created by adding randomness to the bagging method [90].

Random Forest divides each node into branches. It does this separation by selection. Random variables are selected from each node and then the branching process is performed by using the best among the variables. New data sets are generated by replacing the original data set [91] and the random feature is used when developing trees [92]. Figure 3.4 shows the Working Process of Random Forest Regression Algorithm [93].

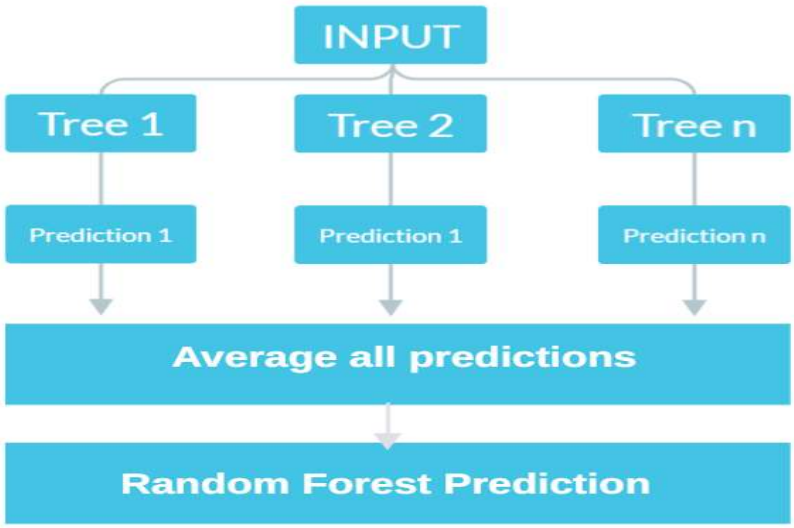


Figure 3.4. The Working Process of Random Forest Regression Algorithm

In random forest regression, two parameters are entered by the user: the number of variables to be used in the node (m) and the number of decision trees (N). The number of decision trees is developed in the regression infrastructure.

3.2.1.4 Support Vector Machine Regression

One of the supervised learning models is the Support Vector Machine [94]. It can be used for both regression and classification. The goal of SVR is to obtain a function close to the training data. While doing this, it tries to minimize the prediction errors and reduce the risk of reaching local values. For this, the smoothness of the established function is also tried to be maximized. The reason why the SVR Model is powerful for regression models with complex relationships is that it uses a kernel. It maps data to a higher dimensional space using a kernel. Thus, it manages non-linear relationships between input and target variables [95].

3.2.2 Removing Outliers

Removing outliers is done to obtain good results in machine learning models. The Z-score was calculated to clean outliers. Z-score is a statistical method that shows how much standard deviation a value is from the mean of the data set. The Z-score is calculated by subtracting the mean value of the data set from the data value and dividing it by the standard deviation. In statistics, it is accepted that 99.7% of the data is 3 standard deviations away. In the thesis, values with $|Z| > 3$ were accepted as outliers. Before giving it as input to machine learning models, outliers in In-situ LAI, Sentinel-1 and Sentinel-2 data were cleaned.

$$Z = \frac{(x - u)}{\sigma} \tag{4}$$

x=observation value, u= mean value of dataset, σ= standard deviation

3.2.3 Evaluation of Machine Learning Models

In this study, MAE, MSE, RMSE, R^2 and %RMSE, which are metrics used in machine learning regression problems, are used to evaluate the performance of the models. The formulas of the metrics used are given in Table 3.1.

Table 3.1. Machine Learning Evaluation Metrics

Metrics	Formula	
Mean Absolute Error (MAE)	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	(5)
Mean Squared Error (MSE)	$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	(6)
Root Mean Squared Error (RMSE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	(7)
%RMSE	$\left(\frac{RMSE}{\bar{y}}\right) \times 100$	(8)
Determination Coefficient (R^2)	$1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$	(9)

Where y_i=actual value, $\hat{y}_i$= estimated value from the model, n= number of values and $\bar{y}$ = mean of the actual values

MAE is the average of the absolute differences between the actual values and the values predicted by the model. MSE is the mean squared error, which is the average of the squared differences between the actual and predicted values. Squaring the differences gives more weight to larger errors. Although it's difficult to interpret because the unit is squared, a smaller MSE value indicates a better model performance. RMSE is the root mean squared error, which is the square root of the MSE. It is based on MSE, making it more sensitive to metric errors. %RMSE is the normalized RMSE, which is obtained by dividing the RMSE by the mean of the actual values. The error is expressed as a percentage. R^2 is the proportion of the variance in the dependent variable that is predictable from the independent variable(s). An R^2 of 1 indicates that the model perfectly fits the data. R^2 is used to summarize the overall performance of the model, as it measures the effect of the independent variables on the dependent variable.

4 RESULTS

4.1 In-situ LAI Values

Within the scope of this study, in-situ LAI values were obtained on 4 dates. In June, LAI values were obtained from 39 points in the Tillering vegetative stage of the rice plant because their height was short. Since the leaves were not completely closed at this stage, low values were obtained between 0.43 and 2.33. In July, LAI values were obtained from 60 points in the flowering/grain formation stage of the rice plant. These values were between 1.48 and 8.21. In August, LAI values were obtained from 60 points in the ripening, filling stage. These values were between 2.17 and 8.10. In September, LAI values were obtained from 52 points at the end of ripening and mature grain stage because the harvest period began. These values were between 2.78 and 7.45.

The max, min, standard deviation and mean values of the LAI values gathered from points for each date are given in Table 4.1. LAI distributions on 4 dates are shown as scatter plots in Figure 4.1.

Table 4.1. Mean of the In-situ LAI

	25.06.2024	17.07.2024	23.08.2024	19.09.2024
Max	2,33	8,21	8,10	7,45
Min	0,43	1,48	2,17	2,78
Mean	1,00	3,60	5,32	4,51
Std. Dev	0,44	1,58	1,28	1,07

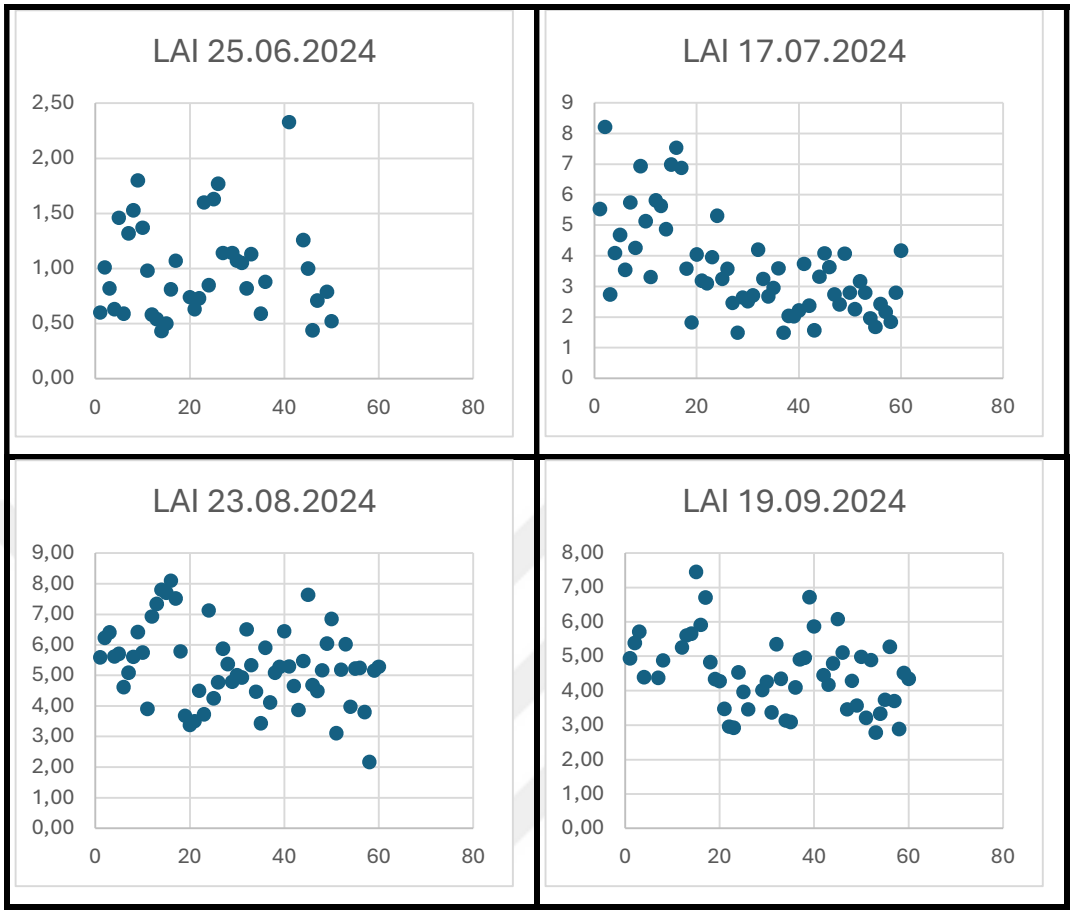


Figure 4.1. LAI Distributions on 4 Dates

4.2 Satellite Data Based Estimation Analysis

The results section with satellite data consists of 3 parts. LR, RFR, and SVR machine learning models are used in all sections.

- The first part includes the LAI estimation results obtained only from the Sentinel-2 satellite,
- The second part includes the LAI estimation obtained from the Sentinel-1 satellite,
- The third part includes the LAI estimation results determined by using both optical and SAR data together.
- In all models, 80% of the data was reserved for testing and 20% for training.

4.2.1 Results of Sentinel-2 Optical Data

The average of the reflectance values of the vegetation indices derived from Sentinel-2 data at points on 4 dates examined in relation to LAI used in the study is shown in Figure 4.2.

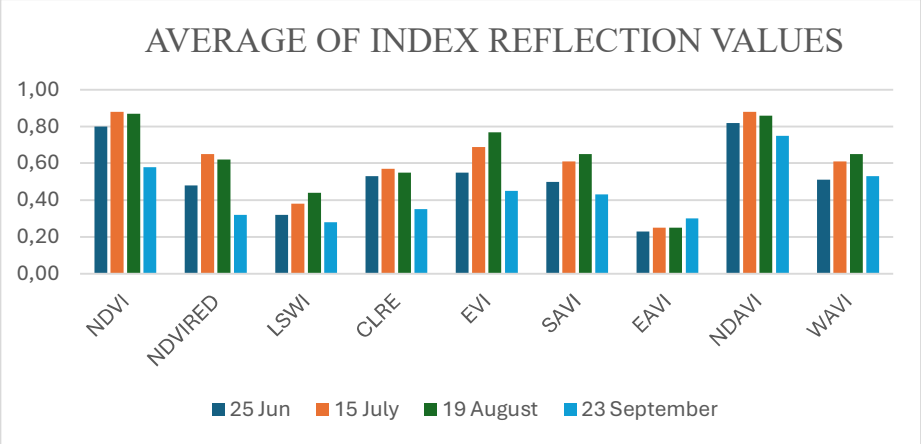


Figure 4.2. Average Vegetation Index Values

In general, the reflectance values of the indices started to increase after June and started to decrease after August. This is due to the growth stages of the paddy-rice plant in the region.

When estimating LAI with Sentinel-2 data, the optical data obtained on 4 dates were given as input to the machine learning models separately and the LAI was estimated for the phenological analysis. Then, the values of all dates were given as input to the machine learning models together and the LAI was estimated.

4.2.1.1 Results of four Sentinel-2 dates

Optical values obtained from the images were entered into machine learning models separately for each date and LAI was estimated. Outliers in in-situ LAI, and Sentinel-2 data were cleaned before feeding them as input to the machine learning models. MAE, MSE, RMSE, %RMSE, R^2 values were calculated to test the performance of the models. Model Results are given in Table 4.2.

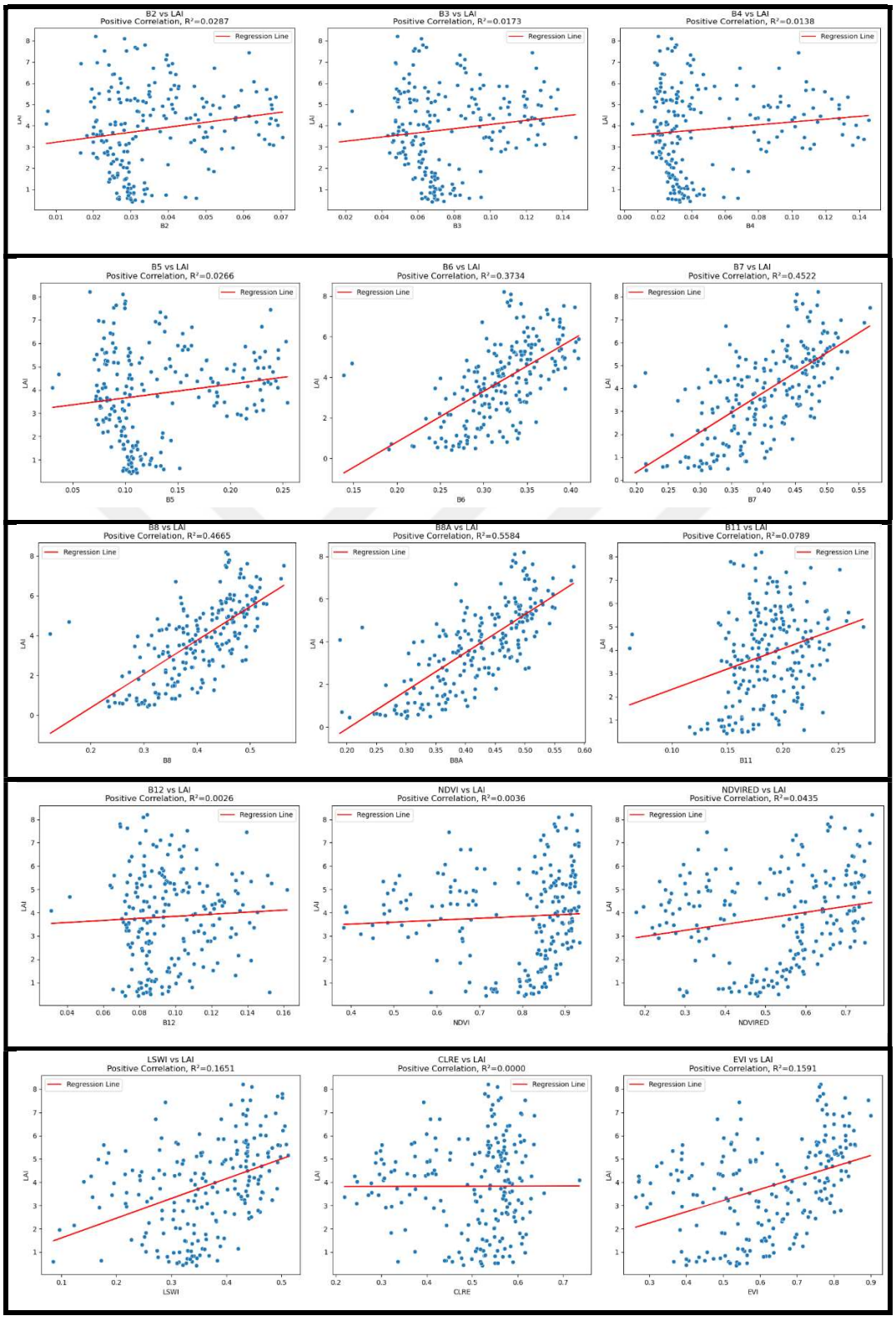
Table 4.2. Model Results of Sentinel-2 for 4 Dates

		25.06.2024	**15.07.2024**	**19.08.2024**	**23.09.2024**
LR	MAE	0.2980	1.2336	1.0243	0.7280
	MSE	0.1363	2.1900	1.3415	0.7696
	RMSE	0.3692	1.4798	1.1582	0.8773
	%RMSE	38.2076	36.7793	22.9283	21.7886
	R^2	0.2656	0.0405	0.2422	0.1856
RFR	MAE	0.2413	0.7658	1.3507	0.6903
	MSE	0.1047	0.7353	2.2694	0.7321
	RMSE	**0.3236**	0.8575	**1.5064**	0.8556
	%RMSE	33.3149	22.2361	26.1807	19.8574
	R^2	0.4936	0.7265	0.1520	0.3438
SVR	MAE	0.3079	0.5182	1.0694	0.7235
	MSE	0.1826	0.4350	1.7752	0.7303
	RMSE	0.4273	0.6595	1.3323	0.8545
	%RMSE	39.8468	22.7445	23.7607	18.1054
	R^2	0.4318	0.2943	0.0007	0.1503

It was revealed that satellite images taken at different growth stages and on different dates of paddy-rice crop gave different results in LAI estimation accuracy. The lowest RMSE value was obtained from the RFR model on June 25, while the highest RMSE value was observed in the RFR model on August 19. The lowest %RMSE value was obtained from the SVR model on September 23, while the highest %RMSE value was observed in the SVR model on June 25. All machine learning models returned low RMSE results on August 19. According to all models, during the vegetative growth period of the crop in June showed consistently low RMSE values and highest R^2 values in all machine learning models.

4.2.1.2 Results of All Sentinel-2 Dates

The correlation values between the measured LAI and the optical data obtained from all dates Sentinel-2 are given in the table. The relation graph between LAI and the data is shown in Figure 4.3. The highest correlation value with LAI is R=0.7473 in the B8A band and the lowest correlation value is R=0.0028 in the CLRE index.


B2 vs LAI
Positive Correlation, R²=0.0287
B3 vs LAI
Positive Correlation, R²=0.0173
B4 vs LAI
Positive Correlation, R²=0.0138
B5 vs LAI
Positive Correlation, R²=0.0266
B6 vs LAI
Positive Correlation, R²=0.3734
B7 vs LAI
Positive Correlation, R²=0.4522
B8 vs LAI
Positive Correlation, R²=0.4665
B8A vs LAI
Positive Correlation, R²=0.5584
B11 vs LAI
Positive Correlation, R²=0.0789
B12 vs LAI
Positive Correlation, R²=0.0026
NDVI vs LAI
Positive Correlation, R²=0.0036
NDVIRED vs LAI
Positive Correlation, R²=0.0435
LSWI vs LAI
Positive Correlation, R²=0.1651
CLRE vs LAI
Positive Correlation, R²=0.0000
EVI vs LAI
Positive Correlation, R²=0.1591
Regression Line
LAI
B2
B3
B4
B5
B6
B7
B8
B8A
B11
B12
NDVI
NDVIRED
LSWI
CLRE
EVI

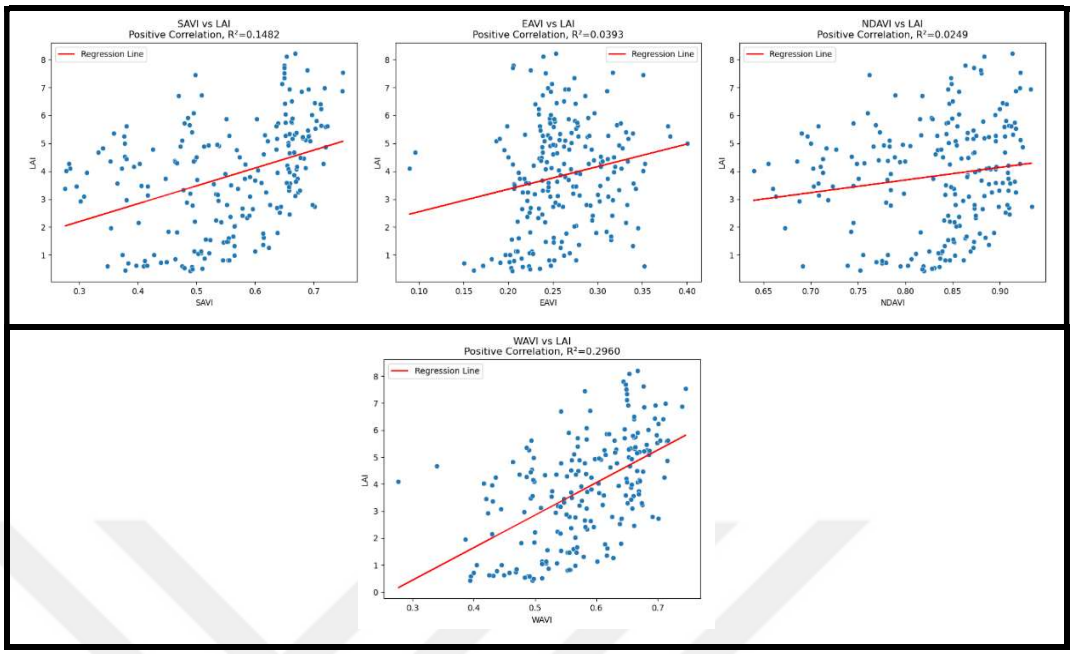


Figure 4.3. Graphics of Relation Between In-situ LAI and Optical Data

The machine learning model results are given in Table 4.3. Among the models trained using only Sentinel-2 data, the best RMSE value was obtained from the LR model. The worst result was obtained from SVR. Figure 4.4 shows the Graphics of LAI Prediction Performance of Models from Sentinel-2 data.

Table 4.3. Model Results of Sentinel-2 for All Dates

	RMSE	MSE	MAE	R^2	%RMSE
Linear Regression	0.9767	0.9540	0.7557	0.7412	24.5200
Random Forest Regression	1.0505	1.1037	0.8364	0.7006	26.3736
Support Vector Regression	1.1231	1.2612	0.9371	0.6579	28.1929

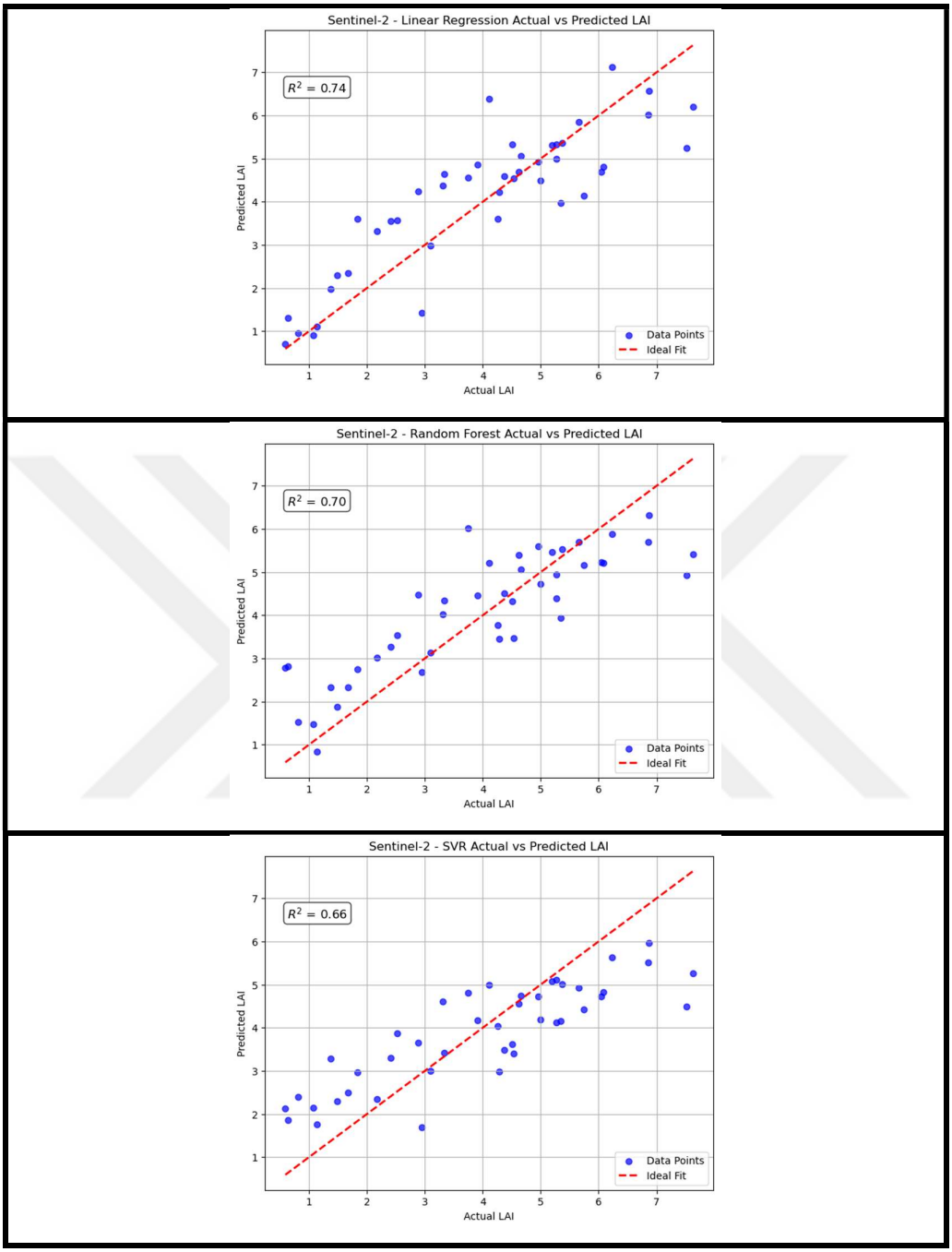


Figure 4.4. Graphics of LAI Prediction Performance of Models from Sentinel-2 Data

4.2.2 Results of Sentinel-1 SAR Data

The average of the backscatter values in VH and VV polarizations, Ratio (VH/VV) and RVI values produced from Sentinel-1 SAR images at points on 4 dates whose relationship with the LAI used in the study is examined are shown in Figure 4.5 and Figure 4.6.

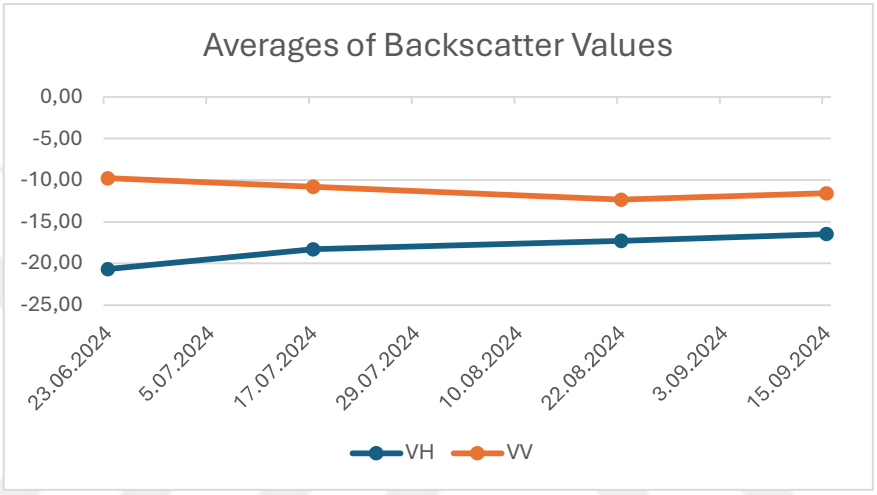


Figure 4.5. Averages of Backscatter Values

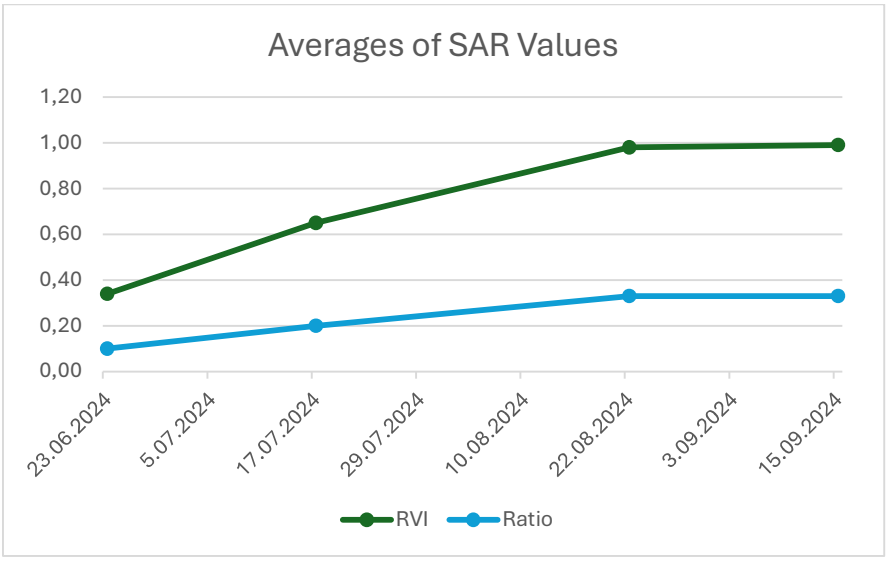


Figure 4.6. Averages of SAR Values

When estimating LAI with Sentinel-1 data, the SAR data obtained on 4 dates were given as input to the machine learning models separately and the LAI was estimated. Then, the values of all dates were given as input to the machine learning models together and the LAI was estimated.

4.2.2.1 Results of four Sentinel-1 four data

SAR values obtained from the images were entered into machine learning models separately for each date and LAI was estimated. Outliers in in-situ LAI and Sentinel-1 data were cleaned before feeding them as input to the machine learning models. MAE, MSE, RMSE, %RMSE, R^2 values are given in Table 4.4.

Table 4.4. Model Results of Sentinel-1 for 4 Dates

		23.06.2024	**17.07.2024**	**22.08.2024**	**15.09.2024**
LR	MAE	0.2603	0.7193	0.9113	1.1461
	MSE	0.1022	0.8509	1.7470	1.851
	RMSE	**0.3197**	0.9224	1.321	1.3606
	%RMSE	31.1601	28.5759	24.2154	27.6244
	R^2	0.1099	0.2079	0.0039	0.1459
RFR	MAE	0.3854	0.9035	0.7808	0.5678
	MSE	0.3227	1.1302	1.1647	0.4432
	RMSE	0.5681	1.0631	1.0792	0.6657
	%RMSE	52.9737	35.4368	19.4223	15.1245
	R^2	0.0176	0.2673	0.2274	0.1631
SVR	MAE	0.3628	1.0126	1.1883	0.8234
	MSE	0.2993	1.990	2.1434	1.0348
	RMSE	0.5471	1.4107	**1.4640**	1.0172
	%RMSE	52.1660	38.0259	26.9335	23.0582
	R^2	0.1008	0.0573	0.0130	0.1262

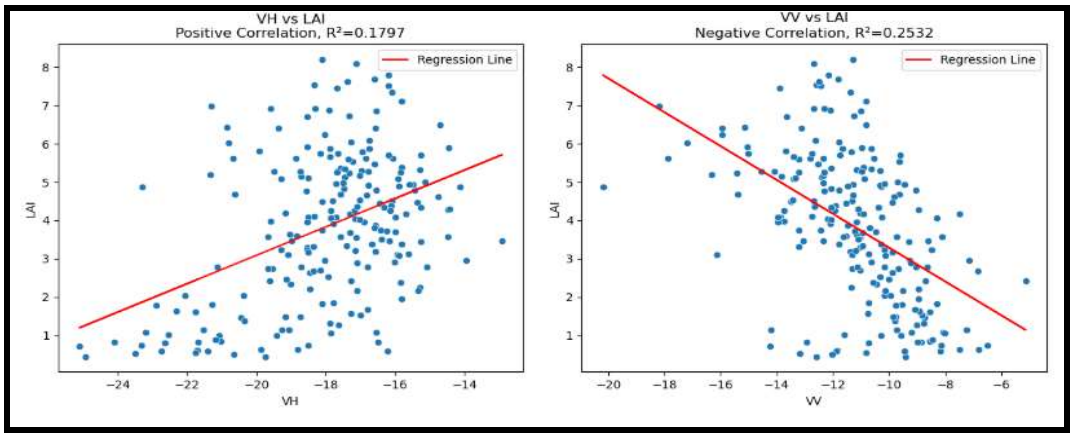
The lowest RMSE value was obtained from the LR model on June 23, while the highest RMSE value was observed in the SVR model on August 22. On September 15, the RFR model showed the lowest error rate with %RMSE = 15.1245. Since the RMSE and %RMSE values were the lowest, it gave the best result. On August 22, the SVR model gave the worst result with RMSE = 1.4640, since the RMSE and %RMSE values were the highest. On August 23, during the plant's ripening, filling phase, all machine learning models returned low-accuracy results. This shows that the SAR data obtained on August 22 is not consistent with the actual rice field data.

4.2.2.2 Results of all Sentinel-1 data

The correlation values between the measured LAI and the SAR data obtained from all dates Sentinel-1 are given in Table 4.5 and the relation graph between LAI and the SAR data is shown in Figure 4.7. While there is a positive correlation between VH, RVI, Ratio values and LAI, there is a negative correlation with the VV value. The value with the highest correlation with LAI is RVI with 0.6800. Among the SAR data, the RVI value is the most sensitive to the paddy rice field.

Table 4.5. Correlation (R value) Between In-situ LAI and SAR Data

VH	VV	RVI	Ratio
0.4239	-0.5032	0.6800	0.6481



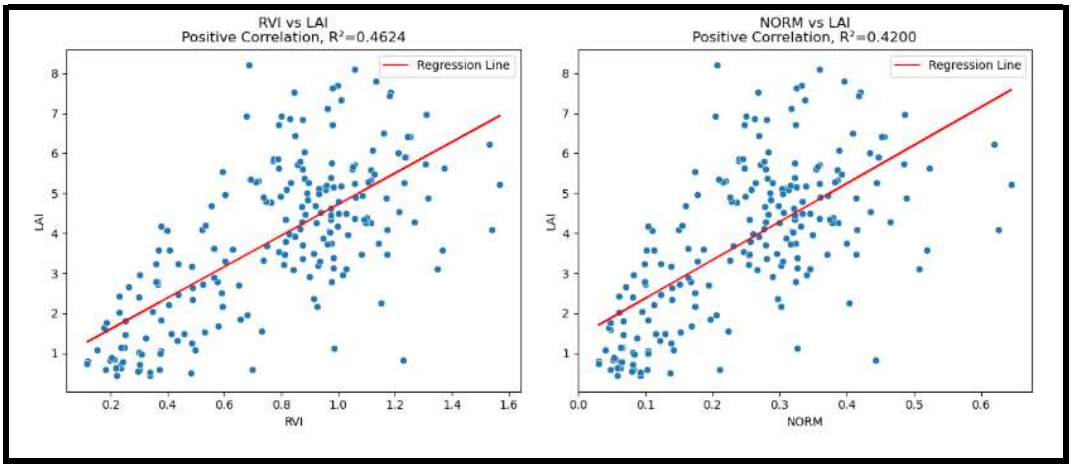


Figure 4.7. Graphics of Relation Between In-situ LAI and SAR Data

The machine learning model and model performance evaluation metric results are given in Table 4.6.

Table 4.6. Model Results of Sentinel-1 for All Dates

	RMSE	MSE	MAE	R^2	%RMSE
LR	1.2907	1.665	1.0224	0.5482	32.4019
RFR	1.4930	2.2291	1.2016	0.3954	37.4806
SVR	1.3465	1.8131	1.070	0.5082	33.8027

Among the models trained using only Sentinel-1 data, the best RMSE value was obtained from the LR model. The worst result was obtained from Random Forest regression. Worse results were obtained from the results obtained using Sentinel-1 satellite images. Figure 4.8. shows the graphics of LAI Prediction Performance of Models from Sentinel-1 data.

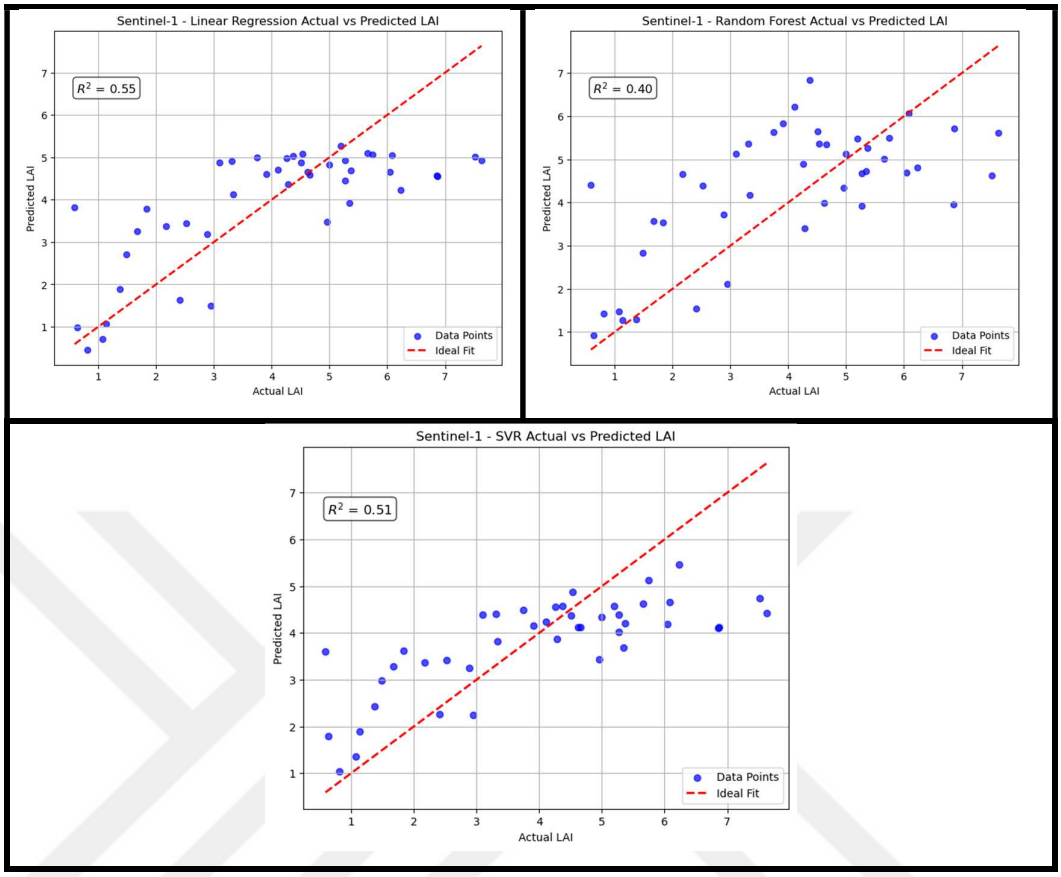


Figure 4.8. Graphics of LAI Prediction Performance of Models from Sentinel-1 Data

4.2.3 Result of Optical and SAR Data

4.2.3.1 Result of Four Dates

Optical and SAR data from close dates are considered together. (Tables 2.4 and 2.7). Each month's data was fed into the machine learning models separately and LAI was estimated. Outliers in in-situ LAI, Sentinel-1 and Sentinel-2 data were cleaned before feeding them as input to the machine learning models. MAE, MSE, RMSE, %RMSE, and R^2 values of each month and each machine learning model is given in Table 4.7.

Table 4.7. Model Results of Optical and SAR Data for 4 Dates

		June	July	August	September
LR	MAE	0.3778	0.7347	0.9343	0.8474
	MSE	0.1725	0.6287	1.3537	0.9081
	RMSE	0.4154	0.7929	1.1634	0.9529
	%RMSE	33.9464	24.9142	23.8176	21.5295
	R^2	0.4441	0.6482	0.1369	0.1588
RFR	MAE	0.2430	0.7901	0.9966	0.6572
	MSE	0.0742	0.7636	1.4172	0.5936
	RMSE	**0.2725**	0.8738	1.1904	0.7704
	%RMSE	24.1161	21.8420	20.5016	16.8766
	R^2	0.3448	0.4698	0.1320	0.1421
SVR	MAE	0.2591	0.9298	1.0722	0.6894
	MSE	0.1061	1.7158	1.8039	0.9327
	RMSE	0.3257	1.3098	**1.3431**	0.9658
	%RMSE	34.5600	36.3582	24.3318	21.2008
	R^2	0.1193	0.0574	0.0058	0.1098

The lowest RMSE value was obtained from the RFR model in June, while the highest RMSE value was observed in the SVR model in August. The highest %RMSE value was obtained from the SVR model in July, while the lowest %RMSE value was observed in the RFR model in September. All machine learning models returned highest RMSE results in August.

4.2.3.2 Results of All Dates

All optical and SAR data obtained from 4 dates were given as input to machine learning models to estimate LAI. The results of the machine learning models where all values are given as independent variables are given in Table 4.8.

Table 4.8. Model Results of Optical and SAR data for All Dates

	RMSE	MSE	MAE	R^2	%RMSE
LR	0.9423	0.8879	0.7471	0.7591	23.6560
RFR	0.9389	0.8817	0.7346	0.7608	23.5720
SVR	1.3869	1.9237	1.1003	0.4782	34.8186

The best RMSE value was obtained from the RFR model with 0.9423. The worst RMSE value was obtained from the SVR model with 1.3869. Accordingly, it is more appropriate to use the RFR model to obtain more accurate results in cases where all data are used together.

The heat map of the correlation between all values is given in Figure 4.9. The heat map shows the correlation between the values used in the study and the in-situ LAI and the LAI values estimated from the models. The highest correlation is shown in red and the lowest correlation is shown in blue. The highest correlation with in-situ LAI is the LAI value obtained from the RFR model with 0.88. This is followed by the LAI value obtained from the LR model, B8A band and the LAI value obtained from the SVR model, respectively. The lowest correlation with in-situ LAI is found in the B12 band with -0.10.

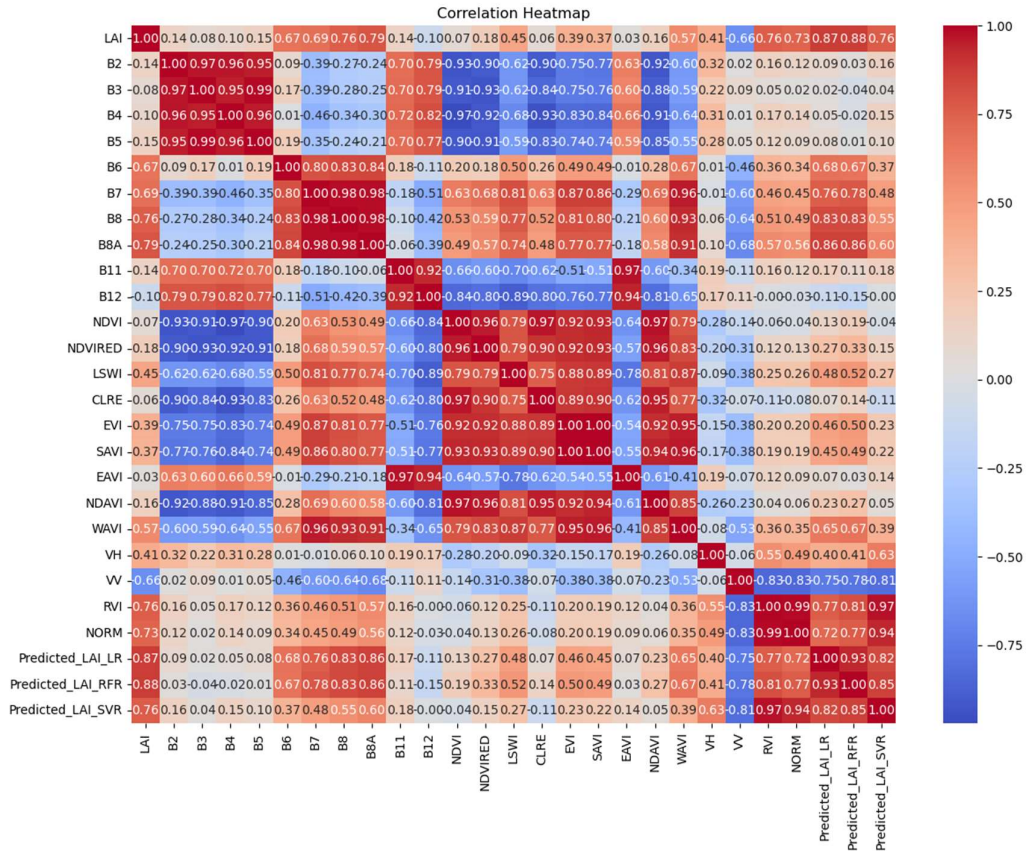
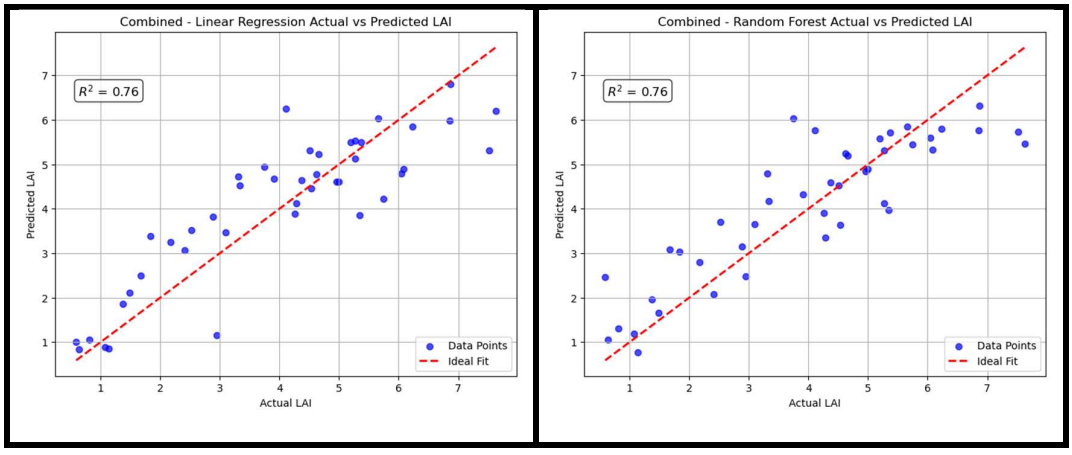


Figure 4.9. Correlation Heatmap of All Values

Graphics of LAI Prediction Performance of Models from All Values in Figure 4.10.



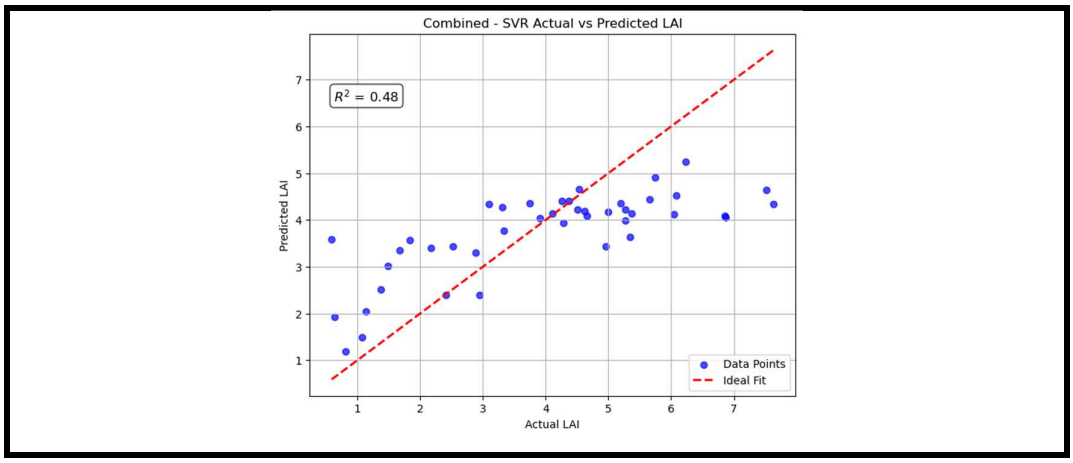


Figure 4.10. Graphics of LAI Prediction Performance of Models from All Values

It was observed that the results obtained from SAR data were lower than the models using optical data. Low RMSE values were obtained in all machine learning models estimated from Sentinel-2 data with machine learning models.

5 DISCUSSIONS

The discussion section includes key findings and model performances for LAI estimation. The study provided insights into estimating LAI in paddy rice fields using machine learning models from Sentinel-1 SAR and Sentinel-2 optical satellite data.

In this study, LAI was collected from paddy rice fields by in-situ study. In-situ studies were carried out on 4 dates corresponding to different growth stages of rice plants. Depending on the growth and harvesting conditions of the LAI plant, LAI could not be collected from the same number of points on each date. LAI was obtained from fewer points in June due to short plant height and in September due to harvest. LAI values obtained from in-situ studies conducted on different dates showed that they were variable depending on the growth stages of the rice plant. Since the leaves of the plant were completely closed in July and August, higher LAI values were obtained. The average LAI value increased between June and August, while it decreased between August and September. These results obtained from in-studies, as in studies examining the trends in the phenological development stages of rice plants, indicate that LAI values continuously increase throughout the tillering, flowering/grain formation and ripening stages, and that LAI values decrease after this stage [96,97]. In July, the LAI values of the rice plant in the flowering/grain formation stage were between 1.48 and 8.21. The reasons for obtaining very different LAI values from the rice plant on the same date from different field which may have different farmer applications. Different planting, irrigation and maintenance factors have been affected. Some studies in the literature [53,98] do not consider the uncertainties in different growth stages of rice because they measure LAI for a single date. In this study, LAI values taken at different dates were examined and LAI changes at different growth stages of rice plants were also examined.

In the first part of the study, data were obtained from the Sentinel-2 satellite. The highest correlation value with LAI is R=0.7473 in the B8A band and the lowest correlation value is R=0.0028 in the CLRE index. We can conclude that the B8A band is more sensitive to rice vegetation and there is no correlation between the CLRE index and rice vegetation. Vegetation indices derived from Sentinel-2 satellite imagery bands showed varying

sensitivity to LAI. The red-edge bands of the Sentinel-2 satellite and the VIs obtained from these bands are very sensitive to vegetation growth [99]. Among these indices, those including the B8A band were particularly effective. It also showed similarity with studies showing that near infrared bands are critical for monitoring vegetation (Clevers & Gitelson, 2013) [100]. Indices such as CLRE showed low sensitivity to paddy-rice vegetation. The results obtained in the study showed that the selection of the vegetation index to be used according to the plant structure and growth stage is important for the accuracy of the study.

It was revealed that satellite images taken at different growth stages and on different dates of paddy-rice crop gave different results in LAI estimation accuracy. When the machine learning models were compared, the RFR model reached the lowest RMSE values when trained on individual dates and the best performance was obtained on June 25. Sentinel-2 satellite images taken during the tillering stage of the paddy in June showed consistently low RMSE values in all machine learning models. Considering only the plant growth phases we used in the study, low LAI values were also obtained in the tillering phase in the study conducted by Prabhakar et. al., (2024) [101].

Among the machine learning models trained on the combined Sentinel-2 data from all dates, the LR model gave the highest performance. The worst result was obtained from SVR. Panigrahi and Das (2021) compared regression models in LAI estimation. In this study, the RFR ($R^2 \approx 0.73$) model gave better results than SVR ($R^2 \approx 0.71$) [102].

In the analysis of Sentinel-1 data in the study, it was concluded that the RVI value showed the highest correlation (R=0.6800) with in-situ LAI. In-situ LAI has positive correlations with backscatter in VH polarization and Ratio values. On the other hand, it has a negative correlation with backscatter values in VV polarization. Najatishendi found that VV polarization backscatter values have a positive correlation with LAI, while VH polarization backscatter values have a negative correlation with LAI. This shows the opposite result [53]. During the rice growth period, canopy density increases with time. This increases the randomness in SAR values and VH variation becomes large. As in the study, low RVI values were obtained in June, and the rice plant reached the highest RVI

value before harvest [103]. As in the study by Chen et. al. (2009), the VV/HH backscattering ratio value obtained from the C-band can be used to estimate LAI at different growth stages of rice plants [104].

When Sentinel-1 data is entered into machine learning models one by one for each date, the lowest RMSE value was obtained from the LR model on June 23, while the highest RMSE value was observed in the SVR model on August 22. On September 15, the RFR model showed the lowest error rate with %RMSE = 15.1245. Since the RMSE and %RMSE values were the lowest, it gave the best result. On August 22, the SVR model gave the worst result with RMSE = 1.4640, since the RMSE and %RMSE values were the highest. On August 22, during the plant's ripening phase, all machine learning models returned low-accuracy results. This shows that the SAR data obtained on August 22 is not consistent with the actual rice field data.

Among the machine learning models trained on the combined Sentinel-1 data from all dates, the LR model gave the highest performance. The worst result was obtained from RFR. In the study conducted by Salma et. al. (2024), the lowest RMSE (0.25) value was obtained with RFR. In this study, on the contrary, the highest RMSE value was obtained from the RFR model [105].

Similar to Najathishendi's study, it was observed that LAI results obtained from SAR data were less accurate than models using optical data [53]. Low RMSE values were obtained in all machine learning models estimated from Sentinel-2 data with machine learning models. It has a higher correlation with in-situ LAI. As stated in the study conducted by Clevers & Gitelson (2013) [100], optical sensors show high sensitivity in detecting canopy structure, vegetation chlorophyll content and leaf biochemical properties. Veloso et al. (2017) stated in their studies that optical data showed high performance [106].

Optical and SAR data from close dates are considered together. (Tables 2.4 and 2.7). Each month's data was fed into the machine learning models separately and LAI was estimated. The lowest RMSE value was obtained from the RFR model in June, while the highest RMSE value was observed in the SVR model in August. The highest %RMSE value was

obtained from the SVR model in July, while the lowest %RMSE value was observed in the RFR model in September. All machine learning models returned highest RMSE results in August. According to all models, the date on which the LAI was obtained from optical data is most consistent with the LAI obtained from in-situ studies in June. This showed that it had a higher correlation with in-situ LAI values. This was due to the canopy biomass and active chlorophyll content obtained well by both optical and SAR data in the vegetative stage in June, which is similar to Yeshanbele et al. (2005). Satellite data obtained during the reproductive stage of the plant in August showed low accuracy in machine learning models. This is due to the complexity of the canopy structure of the plant in the reproductive stage. It can be said that the differences in backscatter values in SAR data and the saturation effect for optical data are caused by the change in plant-water interaction [107]. According to the LAI values obtained by in-situ studies and LAI estimates obtained by machine learning models using satellite data taken at different stages, it was shown that temporal monitoring at different dates is important to capture the crop dynamics of the rice crop.

The combined use of all SAR and optical data improved LAI estimation in all machine learning models. The best RMSE value was obtained from the RFR model with 0.9389 and %RMSE of 23.5720. The worst RMSE value was obtained from the SVR model with 1.3869. Accordingly, it is more appropriate to use the RFR model to obtain more accurate results in cases where all data are used together. The robustness of SAR data to atmospheric conditions and the sensitivity of optical data to chlorophyll and vegetation increased the estimation reliability of machine learning models when combined. These two data types can be interpreted as complementary to each other. Previous studies have also shown that the combined use of these two data types has the potential for crop parameter estimation [106]. Our study is similar to the study of Campos-Taberner et. al. (2017), who concluded that using optical and C-band SAR data together is suitable for LAI estimations [49].

Machine learning models used in LAI estimation have shown different performances depending on the model architecture, feature selection process, and data type. LR is easy to implement but gives low accuracy results in complex cases. RFR has given good and

robust results on different datasets for estimating LAI [47,51]. SVR has potential when combined with other ML models for estimating LAI. However, it has not given as good results as RFR in studies [108]. SVR with SAR data on individual dates showed good performance and demonstrated its ability to effectively model nonlinear relationships [109]. The performance of the SVR model decreased when optical data and SAR data were used together. This is due to the increased size and overfitting.

When estimating LAI, the data is taken from which growth stage of rice, which region, which satellite is used, the bands and indices used and which algorithm is used for estimation affect the estimation accuracy. It is important to select the most appropriate data according to the study region and time. In this thesis, since the in-situ LAI dates and some satellite data are not the same date, it might affected the result of the study.

6 CONCLUSIONS AND FUTURE WORK

This thesis investigates the estimation of LAI using the data collected within paddy rice fields located in the Gönen district of Balıkesir province. The study investigates how different combinations of data derived from Sentinel-2 optical satellite and Sentinel-1 SAR satellite images influence LAI estimation. Data was collected over four distinct months, enabling an assessment of satellite image effectiveness at various growth stages of the rice crop. For this study, Sentinel-2 optical satellite and Sentinel-1 SAR satellite images were acquired on 4 dates. For LAI estimation, B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12 bands and NDVI, NDVIred, LSWI, CLRE, EVI, EAVI, NDAVI and WAVI indices obtained from these bands, backscatter values of VV and VH polarizations, RVI and Ratio value were used. LR, RFR, and SVM machine learning models were employed for LAI estimation. To evaluate model performance, RMSE, MSE, MAE, R^2, and %RMSE metrics were computed, and predicted LAI values were compared against reference LAI data obtained from in-situ measurements. The findings are detailed below.

Firstly, the LAI estimation results obtained using Sentinel-2 optical satellite images are mentioned.

- When LAI calculations were made separately for 4 dates from machine learning models, the lowest RMSE value was obtained from the RFR model as 0.3236 on June 25 2024 during tillering vegetative stage. On the other hand, the highest RMSE value was obtained from the RFR model as 1.5064 on August 19, 2024 in the ripening stage. However, in the end of ripening stage lowest %RMSE values were obtained.
- When the correlation between optical data and in-situ LAI values on all dates is examined, it is seen that the highest correlation is in the B8A band (R=0.7473).
- Optical data from all dates were entered as independent variables into machine learning models for LAI estimation. As a result, the best accuracy was obtained from the Linear regression model (RMSE=0.9767 and R^2=0.7412).

Secondly, the LAI estimation results obtained using Sentinel-1 SAR satellite images are mentioned.

- When LAI calculations were made separately for 4 dates from machine learning models, the lowest RMSE value was obtained from the LR model as 0.3197 on June 23, 2024. On the other hand, the highest RMSE value was obtained from the SVR model as 1.4640 on August 22, 2024. On September 15, the RFR model showed the lowest error rate with %RMSE = 15.1245. Since the RMSE and %RMSE values were the lowest, it gave the best result. On August 22, the SVR model gave the worst result with RMSE = 1.4640, since the RMSE and %RMSE values were the highest. The best result was obtained in the end of ripening stage and the worst result was obtained in the tillering stage.
- High RMSE (greater than 1) values were obtained from all models on August 22, 2024. It corresponds to the ripening stages of the plant. However, high %RMSE values obtained in the first stage of the phenology.
- When the correlation between SAR data and real LAI values on all dates is examined, it is seen that the highest correlation is in RVI (R=0.6800). The correlation between VV value and real LAI values is negative.
- SAR data from all dates were entered as independent variables into machine learning models for LAI estimation. As a result, the best accuracy was obtained from the Linear regression model (RMSE=1.2907, %RMSE=32.4019, and R^2=0.5482).

Thirdly, the LAI estimation results obtained using Combined values (Sentinel-2 Optical + Sentinel-1 SAR) are mentioned.

- When LAI calculations were made separately for 4 dates from machine learning models, the lowest RMSE value was obtained from the RFR model as 0.2725 on June 2024.
- On the other hand, the highest RMSE value was obtained from the SVR model as 1.3431 on August 2024. The lowest RMSE value was obtained in the tillering vegetative stage and the highest RMSE value was obtained in the ripening stage.
- The highest %RMSE value was obtained from the SVR model in July, while the lowest %RMSE value was observed in the RFR model in September in the end of ripening stage.

- When the correlation between the combined data and the real LAI values in all dates is examined, the highest correlation is seen in the B8A band. It is followed by the B8 band and the RVI value (0.79, 0.76 and 0.76 respectively).
- First, all of the combined data were given as independent variables to machine learning models and LAI was estimated. Among the machine learning models where all values are given as independent variables, the highest accuracy was obtained from the RFR model (RMSE=0.9389 and $R^2 = 0.7608$).

It is more appropriate to prefer optical data instead of SAR data for LAI estimation from the study area. Estimates made using only SAR data will have low accuracy. The LAI estimation accuracy increased by using optical and SAR data together from LR and RFR machine learning models. The LAI estimation accuracy decreased by using optical and SAR data together from the SVR machine learning model.

Recommendations for future studies,

- Different indices compatible with rice plants can be calculated for optical data. In addition to Sentinel-2 satellite, data can be obtained from different optical satellites.
- Different values can be calculated, and different SAR vegetation indices can be calculated to estimate LAI from SAR data. In addition to Sentinel-1 satellite, data can be obtained from different radar satellites.
- Other machine learning models such as XGBoost and LightGBM can also be tested for LAI estimation. Deep learning models can also be tested.
- Instead of taking data from a single date in the months used, all satellite data observing the study area in that month can be taken and averaged.
- This study only covers the year 2024. Long-term data can be examined by taking data from previous years.

7 REFERENCES

[1] IRRI (International Rice Research Institute), Rice facts and figures, Los Baños, Philippines, **2020**.

[2] FAO (Food and Agriculture Organization), Rice Market Monitor, Volume XXII - Issue No. 1, Rome, Italy, **2019**.

[3] Bakiye Kılıç Onuk, Coğrafi İşaretli Tarım Ürünlerimizden: Pirinç, https://www.turktarim.gov.tr/Haber/1034/cografi-isaretli-tarim-urunlerimizden-pirinc (accessed: **December 26, 2024**).

[4] J. S. J. Wijesingha, N. L. Deshapriya, and L. Samarakoon. Rice crop monitoring and yield assessment with modis 250m gridded vegetation products: A case study of sa kaeo province, thailand. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-7/W3:121–127, **2015**.

[5] Gregory Asner, Jonathan Scurlock, and Jeffrey Hicke. Global synthesis of leaf area index observations: implications for ecological and remote sensing studies: Global leaf area index. Global Ecology and Biogeography, 12:191–205, **2003**.

[6] Hongliang Fang, Baret Frederic, Stephen Plummer, and Gabriela Schaepman-Strub. An overview of global leaf area index (LAI): Methods, products, validation, and applications. Reviews of Geophysics, 57, **2019**.

[7] Jing M. Chen and Todd A. Black. Defining leaf area index for non-flat leaves. Plant Cell and Environment, 15:421–429, **1992**.

[8] Zhe Zhu, Michael Wulder, David Roy, Curtis Woodcock, Matthew Hansen, Volker Radeloff, Sean Healey, Crystal Schaaf, Patrick Hostert, Peter Strobl, J.-F Pekel, Leo Lymburner, Nima Pahlevan, and Theodore Scambos. Benefits of the free and open landsat data policy. Remote Sensing of Environment, 224:382–385, **2019**.

[9] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. Remote Sensing of Environment, 120:25–36, **2012**.

[10] Jesus Delegido, Jochem Verrelst, C.M. Meza, Juan Rivera Caicedo, Luis Alonso, and Jose Moreno. A red-edge spectral index for remote sensing estimation of green lai over agroecosystems. European Journal of Agronomy, 46:42–52, **2013**.

[11] S. Kobayashi, H. Ide, Rice Crop Monitoring Using Sentinel-1 SAR Data: A Case Study in Saku, Japan, Remote Sensing, 14 (**2022)**.

[12] F. Vuolo, N. Neugebauer, S. Bolognesi, C. Atzberger, G. D'Urso, Estimation of Leaf Area Index Using DEIMOS-1 Data: Application and Transferability of a Semi-Empirical Relationship between two Agricultural Areas, Remote Sensing, 5 (**2013**) 1274–1291.

[13] R. Kulkarni, K. Honda, Estimating LAI of Rice Using NDVI Derived from MODIS Surface Reflectance, Advances in Science, Technology and Engineering Systems Journal 5 (**2020**), 1047–1053.

[14] A.A. Gitelson, Y.J. Kaufman, R. Stark, D. Rundquist, Novel algorithms for remote estimation of vegetation fraction, Remote Sensing of Environment, 80 (**2002**) 76–87.

[15] X. Xiao, S. Boles, S. Frolking, C. Li, J.Y. Babu, W. Salas, B. Moore, Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images, Remote Sensing of Environment, 100 (**2006**) 95–113.

[16] Y. Chabalala, E. Adam, K.A. Ali, Machine Learning Classification of Fused Sentinel-1 and Sentinel-2 Image Data towards Mapping Fruit Plantations in Highly Heterogenous Landscapes, Remote Sensing, 14 (**2022**) 2621.

[17] S.T. Gower, C.J. Kucharik, J.M. Norman, Direct and Indirect Estimation of Leaf Area Index, fAPAR, and Net Primary Production of Terrestrial Ecosystems, Remote Sensing of Environment, 70 (**1999**) 29–51.

[18] J. Černý, P. Haninec, R. Pokorný, Leaf area index estimated by direct, semi-direct, and indirect methods in European beech and sycamore maple stands, Journal of Forestry Research, 31 (**2020**) 827–836.

[19] I. Jonckheere, S. Fleck, K. Nackaerts, B. Muys, P. Coppin, M. Weiss, F. Baret, Review of methods for in situ leaf area index determination, Agricultural and Forest Meteorology, 121 (**2004**) 19–35.

[20] G. Yan, R. Hu, J. Luo, M. Weiss, H. Jiang, X. Mu, D. Xie, W. Zhang, Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives, Agricultural and Forest Meteorology, 265 (**2019**) 390–411.

[21] N.J.J. Breda, Ground-based measurements of leaf area index: a review of methods, instruments and current controversies, Journal of Experimental Botany, 54 (**2003**) 2403–2417.

[22] Y. Qu, Z. Wang, J. Shang, J. Liu, J. Zou, Estimation of leaf area index using inclined smartphone camera, Computers and Electronics in Agriculture, 191 (**2021**) 106514.

[23] L. Zhu, P. Guan, W. Liu, Y. Zheng, Optimization of Lai Estimation Method Based on Smartphones with Fisheye Lens, International Journal of Circuits, Systems and Signal Processing, 17 (**2023**) 112–122.

[24] T. Alexandridis, D. Stavridou, S. Strati, S. Monachou, N. Silleos, LAI measurement with hemispherical photographs at variable conditions for assessment of remotely sensed estimations, **2013**.

[25] M. Goude, U. Nilsson, E. Holmström, Comparing direct and indirect leaf area measurements for Scots pine and Norway spruce plantations in Sweden, European Journal of Forest Research, 138 (**2019**).

[26] Y. Wang, H. Fang, Estimation of LAI with the LiDAR Technology: A Review, Remote Sensing, 12 (**2020**) 3457.

[27] J.M. Chen, J. Cihlar, Retrieving leaf area index of boreal conifer forests using Landsat TM images, Remote Sensing of Environment, 55 (**1996**) 153–162.

[28] Z. Shi, S. Shi, W. Gong, L. Xu, B. Wang, J. Sun, B. Chen, Q. Xu, LAI estimation based on physical model combining airborne LiDAR waveform and Sentinel-2 imagery, Frontiers in Plant Science, 14 (**2023**).

[29] Z. Lu, L. Deng, H. Lu, An Improved LAI Estimation Method Incorporating with Growth Characteristics of Field-Grown Wheat, Remote Sensing , 14 (**2022**) 4013.

[30] M. Campos-Taberner, F. García-Haro, L. Busetto, L. Ranghetti, B. Martínez, M.A. Gilabert, G. Camps-Valls, F. Camacho, M. Boschetti, A Critical Comparison of Remote Sensing Leaf Area Index Estimates over Rice-Cultivated Areas: From Sentinel-2 and Landsat-7/8 to MODIS, GEOV1 and EUMETSAT Polar System, Remote Sensing, 10 (**2018**) 763.

[31] T. Mannschatz, B. Pflug, E. Borg, K.-H. Feger, P. Dietrich, Uncertainties of LAI estimation from satellite imaging due to atmospheric correction, Remote Sensing of Environment, 153 (**2014**) 24–39.

[32] P. Cañete-Salinas, F. Zamudio, M. Yáñez, J. Gajardo, H. Valdés, C. Espinosa, J. Venegas, L. Retamal, S. Ortega-Farias, C. Acevedo-Opazo, Evaluation of models to determine LAI on poplar stands using spectral indices from Sentinel-2 satellite images, Ecological Modelling, 428 (**2020**) 109058.

[33] X. Zhan, Z. Xiao, J. Jiang, H. Shi, A Data Assimilation Method for Simultaneously Estimating the Multiscale Leaf Area Index From Time-Series Multi-Resolution Satellite Observations, IEEE Transactions on Geoscience and Remote Sensing, 57 (**2019**) 9344–9361.

[34] C. Wei, J. Huang, L. Mansaray, Z. Li, W. Liu, J. Han, Estimation and Mapping of Winter Oilseed Rape LAI from High Spatial Resolution Satellite Data Based on a Hybrid Method, Remote Sensing, 9 (**2017**) 488.

[35] D. Zhao, J. Zhen, Y. Zhang, J. Miao, Z. Shen, X. Jiang, J. Wang, J. Jiang, Y. Tang, G. Wu, Mapping mangrove leaf area index (LAI) by combining remote sensing

images with PROSAIL-D and XGBoost methods, Remote Sensing in Ecology and Conservation, 9 (**2023**) 370–389.

[36] V.P. Yadav, R. Prasad, R. Bala, A.K. Vishwakarma, S.A. Yadav, S.K. Singh, A Comparison Of Machine-Learning Regression Algorithms For The Estimation Of Lai Using Landsat - 8 Satellite Data, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 42 (**2019**) 679–683.

[37] Y. Jiang, Z. Zhang, H. He, X. Zhang, F. Feng, C. Xu, M. Zhang, R. Lafortezza, Research on Leaf Area Index Inversion Based on LESS 3D Radiative Transfer Model and Machine Learning Algorithms, Remote Sensing,16 (**2024**) 3627.

[38] T. Liu, H. Jin, X. Xie, H. Fang, D. Wei, A. Li, Bi-LSTM Model for Time Series Leaf Area Index Estimation Using Multiple Satellite Products, IEEE Geoscience and Remote Sensing Letters, 19 (**2022**) 1–5.

[39] T. Liu, H. Jin, A. Li, H. Fang, D. Wei, X. Xie, X. Nan, Estimation of Vegetation Leaf-Area-Index Dynamics from Multiple Satellite Products through Deep-Learning Method, Remote Sensing,14 (**2022**) 4733.

[40] G. Kaplan, O. Rozenstein, Spaceborne Estimation of Leaf Area Index in Cotton, Tomato, and Wheat Using Sentinel-2, Land, 10 (**2021**) 505.

[41] M. Campos-Taberner, F.J. García-Haro, G. Camps-Valls, G. Grau-Muedra, F. Nutini, A. Crema, M. Boschetti, Multitemporal and multiresolution leaf area index retrieval for operational local rice crop monitoring, Remote Sensing of Environment, 187 (**2016**) 102–118.

[42] X. Du, L. Zheng, J. Zhu, Y. He, Enhanced Leaf Area Index Estimation in Rice by Integrating UAV-Based Multi-Source Data, Remote Sensing, 16 (**2024**) 1138.

[43] L.R. Mansaray, A.S. Kanu, L. Yang, J. Huang, Dynamic modelling of rice leaf area index with quad-source optical imagery and machine learning regression models, Geocarto International, 37 (**2022**) 828–840.

[44] S. Li, F. Yuan, S.T. Ata-UI-Karim, H. Zheng, T. Cheng, X. Liu, Y. Tian, Y. Zhu, W. Cao, Q. Cao, Combining Color Indices and Textures of UAV-Based Digital Imagery for Rice LAI Estimation, Remote Sensing, 11 (**2019**) 1763.

[45] M. Hosseini, H. McNairn, S. Mitchell, L.D. Robertson, A. Davidson, N. Ahmadian, A. Bhattacharya, E. Borg, C. Conrad, K. Dabrowska-Zielinska, D. de Abelleyra, R. Gurdak, V. Kumar, N. Kussul, D. Mandal, Y.S. Rao, N. Saliendra, A. Shelestov, D. Spengler, S.R. Verón, S. Homayouni, I. Becker-Reshef, A Comparison between Support Vector Machine and Water Cloud Model for Estimating Crop Leaf Area Index, Remote Sensing (Basel), 13 (**2021**) 1348.

[46] M. Rossi, G. Candiani, F. Nutini, M. Gianinetto, M. Boschetti, Sentinel-2 estimation of CNC and LAI in rice cropping system through hybrid approach modelling, European Journal of Remote Sensing, 56 (**2023**).

[47] L. Wang, Q. Chang, J. Yang, X. Zhang, F. Li, Estimation of paddy rice leaf area index using machine learning methods based on hyperspectral data from multi-year experiments, PLoS One, 13 (**2018**).

[48] J. Verrelst, J.P. Rivera, F. Veroustraete, J. Muñoz-Marí, J.G.P.W. Clevers, G. Camps-Valls, J. Moreno, Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods – A comparison, ISPRS Journal of Photogrammetry and Remote Sensing, 108 (**2015**) 260–272.

[49] M. Campos-Taberner, F. García-Haro, G. Camps-Valls, G. Grau-Muedra, F. Nutini, L. Busetto, D. Katsantonis, D. Stavrakoudis, C. Minakou, L. Gatti, M. Barbieri, F. Holecz, D. Stroppiana, M. Boschetti, Exploitation of SAR and Optical Sentinel Data to Detect Rice Crop and Estimate Seasonal Dynamics of Leaf Area Index, Remote Sensing (Basel), 9 (**2017**) 248.

[50] S. Kumar Singh, R. Prasad, S.A. Yadav, P.K. Srivastava, G. Singh, H. Shanker Srivastava, Fusion of Optical and SAR Data Using Three Approaches for the Estimation of LAI With Modified Integral Equation Model, IEEE Geoscience and Remote Sensing Letters, 21 (**2024**) 1–5.

[51] L.R. Mansaray, F. Wang, A.S. Kanu, L. Yang, Evaluating Sentinel-1A datasets for rice leaf area index estimation based on machine learning regression models, Geocarto International, 37 (**2022**) 1225–1236.

[52] C. Gilardelli, T. Stella, R. Confalonieri, L. Ranghetti, M. Campos-Taberner, F.J. García-Haro, M. Boschetti, Downscaling rice yield simulation at sub-field scale using remotely sensed LAI data, European Journal of Agronomy, 103 (**2019**) 108–116.

[53] E. Najatishendi, Radar Ve Optik Görüntüler Kullanarak Çeltik Bitkisi Yaprak Alan İndeks Kestirimi, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Bilişim Enstitüsü, İstanbul, **2017**.

[54] Anonim, Climate and Average Weather Year Round in Gönen, https://weatherspark.com/y/94384/Average-Weather-in-G%C3%B6nen-Turkey-Year-Round (accessed: **June 12, 2024**).

[55] Bienvenido O. Juliano, Rice in Human Nutrition, 1st ed., Food and Agriculture Organization of the United Nations, International Rice Research Institute, Rome, **1993**.

[56] Shirley Mustafa, Food Outlook – Biannual report on global food markets, FAO, **2024**.

[57] N. Taşlıgil, G. Şahin, Türkiye'de Çeltik (Oryza sativa L.) Yetiştiriciliği Ve Coğrafi Dağılımı, Adiyaman University Journal of Social Sciences, (**2013**) 182–182.

[58] S. Arslan, Çeltik Tarım Ürünleri Piyasaları Raporu, **2024**.

[59] Anonim, Rice | USDA Foreign Agricultural Service, https://fas.usda.gov/data/production/commodity/0422110 (accessed: **December 10, 2024**).

[60] Tahıllar Ve Diğer Bitkisel Ürünler Raporu, https://biruni.tuik.gov.tr/medas/ (accessed: **December 5, 2024**).

[61] J.W. Rouse, R.H. Haas, J.A. Schell, D.W. Deering, Monitoring vegetation systems in the great plains with ERTS, **1973**.

[62] A. Gitelson, M.N. Merzlyak, Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves, Journal of Photochemistry and Photobiology B: Biology, 22 (**1994**) 247–252.

[63] C. Jurgens, The modified normalized difference vegetation index (mNDVI) a new index to determine frost damages in agriculture based on Landsat TM data, International Journal of Remote Sensing, 18 (**1997**) 3583–3594.

[64] A.A. Gitelson, Y. Gritz, M.N. Merzlyak, Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves, Journal of Plant Physiology, 160 (**2003**) 271–282.

[65] A.R. Huete, H.Q. Liu, K. Batchily, W. Van Leeuwen, A comparison of vegetation indices over a global set of TM images for EOS-MODIS, Remote Sensing of Environment, 59 (**1997**) 440–451.

[66] A.R. Huete, A soil-adjusted vegetation index (SAVI), Remote Sensing of Environment, 25 (**1988**) 295–309.

[67] H. Wang, Y. Li, S. Zeng, X. Cai, S. Bi, H. Liu, M. Mu, X. Dong, J. Li, J. Xu, H. Lyu, Y. Zhu, Y. Zhang, Recognition of aquatic vegetation above water using shortwave infrared baseline and phenological features, Ecological Indicators, 136 (**2022**) 108607.

[68] P. Villa, A. Laini, M. Bresciani, R. Bolpagni, A remote sensing approach to monitor the conservation status of lacustrine Phragmites australis beds, Wetlands Ecology and Management, 21 (**2013**) 399–416.

[69] P. Villa, A. Mousivand, M. Bresciani, Aquatic vegetation indices assessment through radiative transfer modeling and linear mixture simulation, International Journal of Applied Earth Observation and Geoinformation, 30 (**2014**) 113–127.

[70] SAR Basics: Google Earth Engine Community Tutorial, https://developers.google.com/earth-engine/tutorials/community/sar-basics (accessed: **December 10, 2024**).

[71] What is Synthetic Aperture Radar? , https://www.earthdata.nasa.gov/learn/earth-observation-data-basics/sar (accessed: **December 10, 2024**).

[72] Sentinel-1, https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1 (accessed: **October 12, 2024**).

[73] R. Nasirzadehdizaji, F. Balik Sanli, S. Abdikan, Z. Cakir, A. Sekertekin, M. Ustuner, Sensitivity Analysis of Multi-Temporal Sentinel-1 SAR Parameters to Crop Height and Canopy Coverage, Applied Sciences, 9 (**2019**) 655.

[74] AccuPAR PAR/LAI Ceptometer Model LP-80, www.decagon.com, **2013**.

[75] L. Kumar, O. Mutanga, Google Earth Engine Applications Since Inception: Usage, Trends, and Potential, Remote Sensing (Basel), 10 (**2018**) 1509.

[76] R. Moore, Introducing Earth Engine for governments and businesses, https://blog.google/products/earth/introducing-earth-engine-for-governments-and-businesses/ (accessed: **December 10, 2024**).

[77] Anonim, Earth Engine Data Catalog, (n.d.). https://developers.google.com/earth-engine/datasets (accessed: December 10, 2024).

[78] Anonim, Google Earth Engine Platform, https://earthengine.google.com/platform/ (accessed: **December 10, 2024**).

[79] Anonim, SNAP, https://earth.esa.int/eogateway/tools/snap (accessed: **December 10, 2024**).

[80] F. Filipponi, Sentinel-1 GRD Preprocessing Workflow, in: F. Filipponi (Ed.), 3rd International Electronic Conference on Remote Sensing, MDPI, Basel Switzerland, **2019**, p. 11.

[81] R.M. Río-Belver, G. Garechana, I. Bildosola, E. Zarrabeitia, Evolution and scientific visualization of Machine learning field, in: Proceedings of the 2nd International Conference on Advanced Research Methods and Analytics (CARMA 2018), Universitat Politècnica València, Valencia, **2018**.

[82] J. Peng, E.C. Jury, P. Dönnes, C. Ciurtin, Machine Learning Techniques for Personalised Medicine Approaches in Immune-Mediated Chronic Inflammatory Diseases: Applications and Challenges, Front Pharmacol 12 (**2021**).

[83] M. Čurlin, S. Drobnjak, I. Potić, D. Đorđević, S. Bakrač, Application of machine learning methods in the classification of satellite images, Tehnika, 79 (**2024**) 152–158.

[84] W.R. Moskolaï, W. Abdou, A. Dipanda, Kolyang, Application of Deep Learning Architectures for Satellite Image Time Series Prediction: A Review, Remote Sensing (Basel), 13 (**2021**) 4822.

[85] M. Praveena, V. Jaiganesh, A Literature Review on Supervised Machine Learning Algorithms and Boosting Process, International Journal of Computer Applications, 169 (**2017**) 32–35.

[86] S. Rong, Z. Bao-wen, The research of regression model in machine learning field, MATEC Web of Conferences, 176 (**2018**) 01033.

[87] Sabita, Everything you need to know about Machine Learning, https://www.analyticsvidhya.com/blog/2021/03/everything-you-need-to-know-about-machine-learning/ (accessed: **December 12, 2024**).

[88] Deepanshi, All you need to know about your first Machine Learning model – Linear Regression, https://www.analyticsvidhya.com/blog/2021/05/all-you-need-to-know-about-your-first-machine-learning-model-linear-regression/ (accessed: **December 12, 2024**).

[89] M. Babaoğlu, A. Coşkunçay, T. Aydın, Modeling Automobile Sales in Turkiye with Regression-Based Machine Learning Algorithms, Journal of Data Applications, 0 (**2023**) 19–33.

[90] L. Breiman, Bagging predictors, Machine Learning, 24 (**1996**) 123–140.

[91] Ö. Akar, O. Güngör, Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması, Journal of Geodesy and Geoinformation, 1 (**2012**) 139–146.

[92] K.J. Archer, R. V. Kimes, Empirical characterization of random forest variable importance measures, Computational Statistics & Data Analysis, 52 (**2008**) 2249–2260.

[93] J. Wood, Machine Learning Regression Models, https://github.com/jeremywood-ai/ML_Regression (accessed: **December 10, 2024**).

[94] C. Cortes, V. Vapnik, Support-vector Networks, Machine Learning, 20 (**1995**) 273–297.

[95] H. Durgun, E. Yılmaz İnce, M. İnce, H.O. Çoban, M. Eker, Evaluation of Tree Diameter and Height Measurements in UAV Data by Integrating Remote Sensing and Machine Learning Methods, Gazi Journal of Engineering Sciences, 9 (**2023**) 113–125.

[96] N.H. Badrul Hisham, N. Hashim, N.M. Saraf, N. Talib, Monitoring of Rice Growth Phases Using Multi-Temporal Sentinel-2 Satellite Image, IOP Conference Series: Earth and Environmental Science, **2022**.

[97] T. Wu, Z. Zhang, Q. Wang, W. Jin, K. Meng, C. Wang, G. Yin, B. Xu, Z. Shi, Estimating rice leaf area index at multiple growth stages with Sentinel-2 data: An evaluation of different retrieval algorithms, European Journal of Agronomy, 161 (**2024**) 127362.

[98] X.Q. Xu, J.S. Lu, N. Zhang, T.C. Yang, J.Y. He, X. Yao, T. Cheng, Y. Zhu, W.X. Cao, Y.C. Tian, Inversion of rice canopy chlorophyll content and leaf area index based on coupling of radiative transfer and Bayesian network models, ISPRS Journal of Photogrammetry and Remote Sensing ,150 (**2019**) 185–196.

[99] Q. Hu, H. Yin, M.A. Friedl, L. You, Z. Li, H. Tang, W. Wu, Integrating coarse-resolution images and agricultural statistics to generate sub-pixel crop type maps and reconciled area estimates, Remote Sensing of Environment, 258 (**2021**) 112365.

[100] J.G.P.W. Clevers, A.A. Gitelson, Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3, International Journal of Applied Earth Observation and Geoinformation, 23 (**2013**) 344–351.

[101] M. Prabhakar, K.A. Gopinath, N. Ravi Kumar, M. Thirupathi, U. Sai Sravan, G. Srasvan Kumar, G. Samba Siva, P. Chandana, V.K. Singh, Mapping Leaf Area Index at Various Rice Growth Stages in Southern India Using Airborne Hyperspectral Remote Sensing, Remote Sensing (Basel), 16 (**2024**) 954.

[102] N. Panigrahi, B.S. Das, Evaluation of regression algorithms for estimating leaf area index and canopy water content from water stressed rice canopy reflectance, Information Processing in Agriculture, 8 (**2021**) 284–298.

[103] S. Dey, Radar Vegetation Index Code for Dual Polarimetric Sentinel-1 Data in EO Browser, https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-1/radar_vegetation_index_code_dual_polarimetric/supplementary_material.pdf (accessed: **December 25, 2024**).

[104] J. Chen, H. Lin, C. Huang, C. Fang, The relationship between the leaf area index (LAI) of rice and the C-band SAR vertical/horizontal (VV/HH) polarization ratio, International Journal of Remote Sensing, 30 (**2009**) 2149–2154.

[105] S. Salma, S.K. Ket, B.M. Dodamani, Analysis of RVI for rice crops in small-scale agricultural fields using Sentinel-1 SAR data: case study on LAI retrieval using regression algorithms, Paddy and Water Environment (**2024**).

[106] A. Veloso, S. Mermoz, A. Bouvet, T. Le Toan, M. Planells, J.-F. Dejoux, E. Ceschia, Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications, Remote Sensing of Environment, 199 (**2017**) 415–426.

[107] Y. Alebele, X. Zhang, W. Wang, G. Yang, X. Yao, H. Zheng, Y. Zhu, W. Cao, T. Cheng, Estimation of Canopy Biomass Components in Paddy Rice from Combined Optical and SAR Data Using Multi-Target Gaussian Regressor Stacking, Remote Sensing (Basel), 12 (**2020**) 2564.

[108] N. Panigrahi, B.S. Das, Evaluation of regression algorithms for estimating leaf area index and canopy water content from water stressed rice canopy reflectance, Information Processing in Agriculture, 8 (**2021**) 284–298.

[109] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Statistics and Computing, 14 (**2004**) 199–222